

Marek Nahotko

**KOMUNIKACJA NAUKOWA
W ŚRODOWISKU CYFROWYM**

WYDAWNICTWO

SBP



NAUKA - DYDAKTYKA - PRAKTYKA



Marek Nahotko, doktor nauk humanistycznych w zakresie bibliologii (2003). Absolwent Bibliotekoznawstwa i Informatyki Naukowej na Uniwersytecie Jagiellońskim (1984). W latach 1992-2000 dyrektor Biblioteki Głównej Politechniki Krakowskiej. Adiunkt w Instytucie Informatyki Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego od 2002 r., specjalizuje się w zagadnieniach opracowania dokumentów elektronicznych, publikowania elektronicznego i wykorzystania Internetu w działalności informacyjnej. Jest autorem książek *Naukowe czasopisma elektroniczne* (2007); *Opis dokumentów elektronicznych : teoretyczny model i możliwości jego aplikacji* (2006); *Metadane : sposób na uporządkowanie Internetu* (2004)

i współautorem książki *Zastosowania komputera w bibliotece* (1996) oraz licznych artykułów publikowanych m.in. na łamach „Przeglądu Bibliotecznego”, „Zagadnień Informatyki Naukowej” oraz „Biuletynu EBIB”. Członek Rady Redakcyjnej „Przeglądu Bibliotecznego”.

Komunikacja naukowa w środowisku cyfrowym

Polish Librarians Association
SCIENCE-DIDACTICS-PRACTICE

Marek Nahotko

**Scholarly communication
in digital environment.**

Global digital library in e-Research



Warsaw 2010

Stowarzyszenie Bibliotekarzy Polskich

NAUKA-DYDAKTYKA-PRAKTYKA

Marek Nahotko

**Komunikacja naukowa
w środowisku cyfrowym**

**Globalna biblioteka cyfrowa
w informatycznej infrastrukturze nauki**

**WYDAWNICTWO
SBP**



Warszawa 2010

Komitet Redakcyjny serii wydawniczej
«NAUKA – DYDAKTYKA – PRAKTYKA»

Marcin Drzewiecki (przewodniczący), Stanisław CZAJKA, Artur JAZDON,
Barbara SOSIŃSKA-KALATA, Danuta KONIECZNA, Dariusz Kuźmina,
Krzysztof MIGOŃ, Mieczysław MURASZKIEWICZ, Janusz NOWICKI (sekretarz),
Joanna Papużyńska-Beksiak, Wanda PINDŁOWA, Maria Próchnicka,
Jadwiga Sadowska, Barbara STEFANIAK, Elżbieta Stefańczyk,
Hanna TADEUSIEWICZ

**Publikacja dofinansowana przez Instytut Informacji
Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego**

Recenzent
dr hab. prof. UW Barbara SOSIŃSKA-KALATA

Redakcja techniczna i korekta
Anna LIS

© Copyright by Stowarzyszenie Bibliotekarzy Polskich

ISBN 978-83-61464-29-7

CIP - Biblioteka Narodowa

Nahotko, Marek
Komunikacja naukowa w środowisku cyfrowym :
globalna biblioteka cyfrowa w informatycznej
infrastrukturze nauki / Marek Nahotko ;
Stowarzyszenie Bibliotekarzy Polskich. - Warszawa :
Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich,
2010, - (Nauka, Dydaktyka, Praktyka ; 112)

Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich
00-335 Warszawa, ul. Konopczyńskiego 5/7, tel. (22) 827-52-96

Warszawa 2010. Wyd. I. Ark. wyd. 26,8. Ark. druk. 17,75.

Łamanie: Anna Lis

Druk i oprawa: Zakład Poligraficzny PRIMUM S.C.,
Kozerki, ul. Marsa 20, 05-825 Grodzisk Mazowiecki, tel. (22) 724-18-76, zp.primum@intelgraf.pl

SPIS TREŚCI

OD AUTORA.....	9
WPROWADZENIE: BIBLIOTEKI CYFROWE JAKO PRZEDMIOT BADAŃ NAUKOWYCH.....	11
1. KOMUNIKACJA NAUKOWA W SIECI	26
1.1. Rozwój komunikacji naukowej.....	26
1.2. Współczesna nauka i komunikacja naukowa.....	34
1.3. Organizacja nauki w środowisku cyfrowym: informatyczna infrastruktura nauki.....	41
1.4. Dane, informacja i wiedza w IIN.....	48
1.5. Globalna biblioteka cyfrowa jako część IIN.....	55
2. STRUKTURY.....	64
2.1. Kodowanie znaków.....	65
2.2. Formaty danych.....	67
2.3. Języki kodowania.....	73
2.4. Protokoły.....	77
2.5. Obiekty cyfrowe.....	80
2.6. Publikacje elektroniczne.....	85
2.7. Organizacja informacji i wiedzy.....	100
2.8. Metadane.....	104
3. PRZESTRZENIE.....	117
3.1. Internet jako przestrzeń publiczna.....	118
3.2. Globalna biblioteka cyfrowa jako miejsce.....	120
3.3. Kryteria i miary oceny.....	125
3.4. Przestrzenie nazw.....	132
3.5. Identyfikatory obiektów cyfrowych.....	134
3.6. Interfejsy użytkownika.....	136
4. PROCESY.....	141
4.1. Publikowanie.....	142
4.2. Tworzenie metadanych.....	148
4.3. Wyszukiwanie informacji.....	152
4.4. Zapewnienie integralności i autentyczności obiektów cyfrowych.....	181
4.5. Dostępność informacji i autoryzacja dostępu.....	184
4.6. Współdziałanie.....	187
4.7. Archiwizacja (długotrwałe przechowywanie).....	189
4.8. Ocena (kontrola jakości).....	193
4.9. Semantyczny Web oraz Web 2.0 i Nauka 2.0.....	200
5. SPOŁECZNOŚCI.....	206
5.1. Osoby i instytucje.....	207
5.2. Zagadnienia ekonomiczne.....	221
5.3. Zagadnienia prawne.....	224
6. MODEL – METAFORA KOMUNIKACJI NAUKOWEJ W GBC.....	230
ZAKOŃCZENIE: KIERUNKI ROZWOJU KOMUNIKACJI NAUKOWEJ.....	241
PUBLIKACJE WYKORZYSTANE.....	249
INDEKS OSOBOWO-PRZEDMIOTOWY.....	271
SUMMARY.....	282

CONTENTS

FROM THE AUTHOR	9
INTRODUCTION: DIGITAL LIBRARIES AS AN OBJECT OF SCHOLARLY INVESTIGATIONS	11
1. SCHOLARLY COMMUNICATION IN THE NET	26
1.1. Development of scholarly communication.....	26
1.2. Contemporary science and scholarly communication.....	34
1.3. Science organization in digital environment: e-Research infrastructure.....	41
1.4. Data, information and knowledge in e-Research infrastructure.....	48
1.5. Global digital library as a part of e-Research infrastructure.....	55
2. STRUCTURES	64
2.1. Representing characters.....	65
2.2. Data formats.....	67
2.3. Markup languages.....	73
2.4. Protocols.....	77
2.5. Digital objects.....	80
2.6. Electronic publications.....	85
2.7. Information and knowledge organization.....	100
2.8. Metadata.....	104
3. SPACES	117
3.1. Internet as a public space.....	118
3.2. Global digital library as a place.....	120
3.3. Evaluation criteria and measures.....	125
3.4. Namespaces.....	132
3.5. Digital object identifiers.....	134
3.6. User interfaces.....	136
4. PROCESSES	141
4.1. Scholarly publishing.....	142
4.2. Metadata development.....	148
4.3. Information retrieval.....	152
4.4. Integrity and authenticity of digital objects.....	181
4.5. Information accessibility and access authorization.....	184
4.6. Interoperability.....	187
4.7. Archiving (long-term preservation).....	189
4.8. Evaluation (quality control).....	193
4.9. Semantic Web, Web 2.0 and Science 2.0.....	200
5. SOCIETIES	206
5.1. Persons and institutions.....	207
5.2. Economics problems.....	221
5.3. Intellectual property.....	224
6. MODEL – SCHOLARLY COMMUNICATION METAPHOR	230
CONCLUSIONS: DEVELOPMENT DIRECTIONS OF SCHOLARLY COMMUNICATION	241
REFERENCES	249
INDEX	271
SUMMARY	282

WYKAZ TABEL I RYSUNKÓW

Tab. 1. Elementy i funkcje biblioteki tradycyjnej, bibliotek cyfrowych i GBC.....	63
---	----

*

Rys. 1. Interdyscyplinarne badania nad bibliotekami cyfrowymi (wg [Chen, Kiernan 2002]).....	20
Rys. 2. Metamodel biblioteki cyfrowej wg zasad 5S (wg [Fox 2005]).....	23
Rys. 3. Porównanie wpływu wzrostu cen czasopism na prenumeratę (lewa strona) [Schlimgen, Kronenfeld 2004, s. 310] oraz wzrostu liczby czytanych artykułów (prawa strona) [Tenopir, King 2008].....	39
Rys. 4. Usługi IIN tworzące środowisko wiedzy dla badań i edukacji.....	45
Rys. 5. Złożony obiekt cyfrowy.....	83
Rys. 6. Przejście od metadanych do danych.....	105
Rys. 7. Rodzaje wyszukiwania wspólnego.....	155
Rys. 8. Interfejs systemu Medline – zastosowanie języka naturalnego i logiki rozmytej.....	160
Rys. 9. Integracja wielu zasobów informacyjnych.....	178
Rys. 10. Serwisy GBC w obrębie kolejnych sfer infrastruktury informacji.....	178
Rys. 11. Obieg informacji naukowej w systemie tradycyjnym (publikacji drukowanych).....	232
Rys. 12. Dystrybucja obiektów elektronicznych w Sieci.....	234
Rys. 13. Kontinuum publikacji elektronicznych.....	235
Rys. 14. Nieformalne powiązania pomiędzy uczonymi.....	237
Rys. 15. Hybrydowy model komunikacji naukowej.....	237
Rys. 16. Model komunikacji naukowej w GBC.....	239

Jestem ciekaw, czy oglądając, drogi czytelniku, jakiś film sprzed kilkunastu lat, przedstawiający ówczesną rzeczywistość, a więc czasy dobrze Ci znane, bo osobiście przeżyte, nie czujesz lekkiego zaskoczenia, przechodzącego w zażenowanie, gdy główny bohater filmu wpada w tarapaty, gdyż nie było sposobu ostrzec go na czas o czyhającym niebezpieczeństwie, chociaż jego otoczeniu było ono znane nawet godzinę wcześniej? Zawsze wówczas myślę sobie: gdyby to było kilka lat później, wystarczyłoby zadzwonić na komórkę... Jak widać postęp w technologiach informacyjnych ma znaczny wpływ nawet na sposoby budowania napięcia i fabuły przez scenarzystów filmowych.

Nie jest to jednak jedyne oddziaływanie nowych technologii komunikacyjnych na nasze życie. Prawie niezauważalnie już staliśmy się połączeniem *homo computerus* [Morbitzer 2000], czy jak kto woli *homo informaticus* [Golka 2008, s. 87] z *homo communicans* [Filipiak 2004], a to tylko początek tej drogi. Producenci sprzętu elektronicznego i dostawcy treści zdążyli już bowiem zauważyć, że telefon komórkowy służyć może nie tylko do rozmów i wysyłania SMS-ów. Już wkrótce za pomocą tego typu urządzeń wszyscy będziemy przyłączeni, przez 24 godziny na dobę, do naszych domów, w których zdalnie sterować będziemy urządzeniami AGD, do miejsc pracy, z których na bieżąco napływać będą dokumenty i informacje, a także do członków rodziny i znajomych. Innymi słowy za pomocą swojego przenośnego komputera (bo tak naprawdę można nazwać współczesne telefony komórkowe) połączeni zostaniemy z wszystkimi innymi komputerami (może prawie wszystkimi), umieszczonymi w domu, miejscu pracy, bibliotece, banku, samochodzie, parkomacie i innych urządzeniach sterowanych. Takie multifunkcjonalne urządzenie (będące zegarkiem, telefonem, faksem, skanerem, aparatem fotograficznym, odtwarzaczem muzyki oraz komputerem biurowym w jednym – wymieniałem tylko najbardziej oczywiste funkcje) spowoduje, że podział czasu na spędzany w pracy bądź w domu lub w drodze do domu, a więc rozróżnienie czasu pracy i odpoczynku, zmienią sens.

Jak z tego wynika, nowe sposoby komunikowania zmieniają nie tylko codzienne życie w naszym postindustrialnym i postmodernistycznym świecie, ale także sposób prowadzenia działalności zawodowej przez każdego z nas. W bardzo szczególnie sposób dotyczy to pracowników nauki, których funkcjonowanie zawsze zależało od właściwie zorganizowanej komunikacji, zarówno tej formalnej (publikacje takie jak książki i artykuły w czasopiśmie, obecnie elektroniczne), jak i nieformalnej (dyskusje prowadzone na drodze korespondencji pocztowej lub w kulturalach konferencyjnych, obecnie za pomocą poczty elektronicznej, czatów i blogów). Nie dziwi to tym bardziej, że Internet, oprócz zastosowań militarnych, od początku miał ułatwić właśnie komunikację między uczonymi.

Moja książka traktuje o zmianach, jakie nastąpiły i nadal następują w wyniku zastosowania technologii informacyjnej w komunikacji naukowej. Słowo „zmiany” powtarza się w niej bardzo często. Wynika to z tego, że jak pisze Lawrence Lessig, Internet stał się początkiem końca XX-wiecznych gałęzi przemysłu, tworzących i dystrybuujących treści, a przynajmniej początkiem istotnych zmian [Lessig 2005, s. 35]. W nauce zmiany te są tak rozległe, że wpływają na sposób realizacji samego procesu badawczego, którego częścią jest przecież z jednej strony rozpowszechnienie wyników badań, a z drugiej – poznanie wyników badań prowadzonych przez innych badaczy. Natężenie zmian powoduje nawet wrażenie chaosu i jest źródłem dyskomfortu wynikającego z płynności sytuacji,

ale mimo wszystko wydaje się, że prowadzą one do sensownych rozwiązań [Bard, Söderqvist 2006, s. 44]. Podobnie jak na użytek nas wszystkich powstaje jednolite środowisko informacyjne, coraz bardziej oparte na elektronicznych urządzeniach mobilnych, również w zakresie komunikacji naukowej powstaje takie środowisko, które metaforycznie nazywałem globalną biblioteką cyfrową. Moim zadaniem podczas pisania tej książki było opisanie tego środowiska w aktualnym, dynamicznym stanie oraz znalezienie zależności, umożliwiających przewidywanie choćby najbliższej przyszłości.

Jednocześnie zdaję sobie sprawę, że opisywanie tego typu zjawisk jest jak gonitwa za uciekającym horyzontem – to, co opisane zostało jako rozwiązania najbardziej aktualne, w momencie oddania książki do rąk Czytelników będzie już przeszłością. Życie szybko (coraz szybciej) dopisze do niej ciąg dalszy.

Na koniec pragnę bardzo podziękować Pani prof. Barbarze Sosińskiej-Kalacie z IINiSB UW i Panu prof. Jackowi Wojciechowskiemu z IINiB UJ, którym ta książka w poważnym stopniu zawdzięcza swój kształt, gdyż przysłużyli się jej swoimi cennymi uwagami, radami i chętnie uwzględnianymi propozycjami poprawek.

Dziękuję także Pani dr hab. Marii Próchnickiej, wspierającej to przedsięwzięcie na każdym etapie.

WPROWADZENIE: Biblioteki cyfrowe jako przedmiot badań naukowych

Dzięki temu, że przy szczupłych rozmiarach mieścić może zawartość intelektualną i artystyczną wielkiej wagi, dzięki temu, że z łatwością przechodzić może z ręk do ręk, że może być do woli kopiowana i powielana – stanowi ona najprostsze narzędzie zdolne wyzwolić, przy pomocy określonego bodźca, całe niezliczone bogactwo marzeń, obrazów, uczuć, idei, informacji otwierając wrota czasu i przestrzeni; następnie [...] pozwala na uporządkowanie rozproszonych danych, które przekazuje do mnóstwa innych punktów rozsypanych na przestrzeni wieków i kontynentów; tworząc z nich niezliczoną ilość kombinacji całkowicie od siebie odmiennych

[Escarpit 1969, s. 15-16].

Cytat na początku tekstu pochodzi z 1965 r., kiedy w ten sposób o książce drukowanej pisał Robert Escarpit. Czy takiej apoteozy doczekają się w przyszłości dokumenty cyfrowe, krążące w sieciach rozległych? Wszak nowe technologie dają niemal nieograniczone możliwości: nie tylko tak, jak druk umożliwiając masowe rozpowszechnianie wszelkich (również naukowych) treści, ale i demokratyzując sam proces twórczy. Obecnie każdy może nie tylko korzystać z materiałów wyzwalających marzenia, obrazy, uczucia, idee i informacje (w tym naukowe), ale także sam je tworzyć i bezpośrednio rozpowszechniać. Negatywnym efektem jest to, że obok treści istotnych, o wysokiej jakości, egzystują dzieła wtórne lub bezwartościowe. Komunikację cyfrową, niemal od czasu jej powstania, krytykuje się za niestałość, łatwość mnożenia tekstów sposobem „kopiuj i wklej”, częsty brak autorstwa w tradycyjnym sensie i niszczenie indywidualnej kreatywności, funkcjonowanie treści ważnych obok zupełnie bezwartościowych [Burk, Kerr, Pope 2000; Harter 1997; Keen 2007; Waters 2009, s. 29]. Wydaje się jednak, że problemy dotyczące internetowego środowiska informacyjnego nie są związane z jego nowymi cechami, ale raczej z intensywnością (powszechnością) występowania cech (zarówno pozytywnych, jak i negatywnych), które odnaleźć można w każdej organizacji komunikacji naukowej – tradycyjnej lub elektronicznej.

Być może szczególna nieufność do artefaktów umieszczanych w Sieci wynika z tego, że aby mogły się tam znaleźć, przedmioty fizyczne będące nośnikami treści, muszą przejść znaczącą transformację, która według Nicholasa Negroponte, polega na zmianie środków przekazu z materialnych atomów na niematerialne bity [Negroponte 1996, s. 43], czyli na cyfryzacji. Uważał on, że wraz ze zmniejszaniem się liczby ludzi pracujących przy produkcji dóbr materialnych i wzrostem grupy osób świadczących usługi lub udzielających informacji¹, następuje przejście od produkcji i dystrybucji przedmiotów (czyli skupisk atomów) do produkcji i rozpowszechniania artefaktów mniej materialnych, nienamacal-

¹ Obecnie (2004) w USA w usługach pracuje 76% zatrudnionych, w przemyśle – 23% i w rolnictwie ok. 0,7%. W Polsce struktura ta wygląda następująco: usługi – 54,5%, przemysł – 30,7%, rolnictwo – 14,8% (dane wg Wikipedii). Natomiast już w latach siedemdziesiątych XX w. w USA „działania informacyjne” dawały 46% PKB [Barney 2008, s. 16]

nych – usług i informacji (określanej za pomocą abstrakcyjnych bitów). Przewidywał jednocześnie szybkie przejście od starych, pasywnych mediów do nowych, interaktywnych. Sytuacja ta ma oczywiście silny wpływ na całą kulturę, w tym również na naukę. Obecnie odchodzi się od materialnych przedmiotów, takich jak naukowe książki i czasopisma drukowane, zamieniając je na obcowanie z bitami informacji pozbawionej tradycyjnej (atomowej) formy nośnika. Informacja ta może być udostępniona wszędzie: na ekranie laptopa, komórki czy iPoda, na ścianie, a nawet, jeśli ktoś woli, na drzwiach lodówki. Nowe technologie umożliwiły transmisję różnymi kanałami tych samych treści, przyjmujących wiele odmiennych form w punkcie odbioru, co nazywane jest konwergencją mediów². Technologie stare i nowe (cyfrowe) wciąż egzystują obok siebie, powodując konwergencję informacji i mediów informacyjnych, przy okazji zacierając granice pomiędzy komercyjnym profesjonalizmem i amatorstwem oraz między pracą i zabawą. Taka konwergencja jest efektem okresu przejściowego, w którym coraz więcej usług i produktów informacyjnych tworzonych jest w postaci cyfrowej, co ułatwia mieszanie zastosowań.

Nowe narzędzia zmieniają tych, którzy się nimi posługują, a jak twierdził Marshall McLuhan, środek przekazem sam jest przekazem, co oznacza, że nowe środki przekazu (narzędzia) powodują zmiany w skali, tempie i/lub wzorcach ludzkich zachowań i wzajemnego oddziaływania na siebie [McLuhan 2004, s. 460]. Podobnie jest w komunikacji naukowej, której istotnym czynnikiem rozwoju od kilkunastu lat są technologie informacyjne. Zastosowania nowych technologii w zakresie komunikacji poprzez sieci komputerowe (scalone w jedną Sieć – Internet), wykorzystanie komputerów do tworzenia, przechowywania i wyszukiwania informacji naukowej oraz migracja od nośnika drukowanego do elektronicznego w publikacjach naukowych – to przykłady osiągnięć wywierających znaczny wpływ na sposób wymiany informacji w środowisku naukowym. Coraz częściej można odnieść wrażenie, że to, czego nie ma w Internecie (czyli jest trudne do wyszukania), nie istnieje w ogóle [Kisiłowska 2009, s. 88]. Jest to nie tylko wrażenie; jak piszą Paul Delany i George Landow, teksty, a nawet całe systemy językowe, które nie przechodzą do nowego, elektronicznego medium, ulegają marginalizacji, stając się dosłownie niewidzialnymi [Delany, Landow 2008, s. 87].

Pojawiają się też głosy mówiące, że konwergencja mediów nie wyjaśnia wystarczająco dobrze obserwowanych zjawisk; lepszym sposobem jest przyjęcie paradygmatu sieciowego, w którym zwraca się uwagę na sieciową formę organizacji mediów [Cardoso 2008, s. 589]. Od przenikania się mediów ważniejsze są ich wzajemne powiązania. Tu również technologia odgrywa istotną rolę, gdyż umożliwia na przykład współpracę różnych urzędzeń, mobilnych i stacjonarnych, tworzących sieć narzędzi, treści i użytkowników [Dempsey 2009]. W tej teorii poszczególne media nie są więc (lub są nie tylko) konwergowane, lecz przede wszystkim łączone w sieć.

We współczesnym społeczeństwie sieciowym³ zasadnicza rola nowych technologii in-

² Henry Jenkins jako konwergencję rozumie przepływ treści pomiędzy różnymi platformami medialnymi, współpracę różnych przemysłów medialnych oraz migracyjne zachowania odbiorców mediów. Konwergencja to pojęcie opisujące zmiany technologiczne, przemysłowe, kulturowe i społeczne możliwe dzięki nowym technologiom [Jenkins 2007, s. 9]. Jest to zjawisko lub proces, w którym można obserwować zmieniające się i wzajemnie przenikające zależności pomiędzy treściami medialnymi, kulturowymi i komercyjnymi oraz ich twórcami i odbiorcami [Jaskowska 2008].

³ Według Mieczysława Muraskiewicza, rozpowszechnienie paradygmatu sieciowego powoduje zastąpienie terminu „społeczeństwo informacyjne” określeniem „społeczeństwo sieciowe” [Muraskiewicz 2005, s. 85]. Społeczeństwo informacyjne zdefiniować można jako takie, w którym jakość życia oraz perspektywy rozwoju społecznego i gospodarczego zależą coraz bardziej od informacji i jej wykorzystania [Ołędzki 2001, s. 18]. Produkcja, gromadzenie i obieg informacji są niezbędne do jego funkcjonowania [Golka 2008, s. 80]. W społeczeństwie sieciowym kładzie się większy nacisk na sposób komunikacji, a nie jej treść. Charakterystyczna dla niego jest obecność wyrafinowanych – głównie cyfrowych – technologii komunikacji sieciowej oraz stosowanie sieci jako podstawowej formy organizacji społecznej [Barney 2008, s. 37].

formacyjnych, zwłaszcza w zakresie digitalizacji i cyfryzacji⁴ zasobów informacyjnych nie podlega dyskusji. Świat publikacji naukowych wykorzystuje na wielką skalę nowe możliwości technologiczne (zob. p. 1.6). Wystarczy powiedzieć, że ponad 90% czasopism naukowych wydawanych w krajach wysoko rozwiniętych i powyżej 60% czasopism w skali ogólnoswiatowej posiada swoje wersje elektroniczne, co oznacza istnienie ok. 20 tys. tytułów czasopism elektronicznych na świecie, podczas gdy jeszcze w 1997 r. było ich 3400 [Nahotko 2007b, s. 43]. Podobna skala wzrostu jest widoczna w zakresie elektronicznych czasopism recenzowanych: w 1996 r. było ich 131, a w 2003 r. ich liczba doszła do 14 338 tytułów, co daje 62% z 23 187 czasopism recenzowanych⁵, wydawanych na świecie [Morrow i in. 2008, s. 8]. Zmianom ilościowym towarzyszą także przemiany jakościowe w sposobie organizacji komunikacji naukowej, widoczne np. w szybkim rozpowszechnianiu się udostępniania materiałów naukowych w trybie Open Access⁶. W publikacjach elektronicznych, początkowo kopiujących treść i formę publikacji drukowanych, obecnie zaczyna się coraz pełniej wykorzystywać możliwości, jakie daje wyłącznie medium elektroniczne. Przykładem jest Web 2.0, w którym szeroko zastosowano m.in. interaktywność nowych mediów i inne narzędzia służące komunikowaniu się społeczności, w tym uczonych. W efekcie coraz powszechniejszego stosowania tych narzędzi zmieniają się role wszystkich uczestników procesu komunikowania naukowego: pracowników nauki (autorów, recenzentów, redaktorów), wydawców, dystrybutorów, bibliotekarzy. Zmiany te są bardzo dynamiczne i zachodzą właśnie teraz, współcześnie, co powoduje, że ich badanie jest trudne, ale niezbędne. Dlatego też czas przyszły, często używany w tej książce, nie powinien mylić: już w trakcie opisywania przyszłość staje się teraźniejszością, potem szybko przeszłością, albo nawet (czasami) nieprawdą.

Rewolucyjne zmiany technologiczne powodują, że fizyczna lokalizacja traci znaczenie, a technologie informacyjne stają się wszechobecne podczas codziennej pracy, nauki, prowadzenia przedsiębiorstwa czy komunikacji międzyludzkiej wszelkiego rodzaju (w szczególności naukowej). Bezpośredni dostęp do zasobów informacyjnych, bez względu na miejsce i czas, powoduje coraz głębsze zmiany w sposobach komunikowania. Instytucje, takie jak biblioteki, archiwa, muzea, wydawnictwa, księgarnie, uczelnie, aby przetrwać muszą ulec radykalnej transformacji, co oznacza także radykalną zmianę stosunków i współzależności między różnymi grupami; autorzy uniezależniają się od wydawców, użytkownicy informacji naukowej od bibliotek⁷, instytucje naukowe znajdują nowe modele publikowania i oceny działalności naukowej [Borgman 2003a, s. 2].

Rozwój technologii stanowi pewne continuum; Internet może być traktowany jako kontynuacja wcześniejszych technologii komunikacyjnych, na przykład telefonu, telegra-

⁴ Cyfryzacją nazywam zastosowanie technologii cyfrowych w komunikacji naukowej do tworzenia materiałów cyfrowych jako formy pierwotnej (oryginalnej, *digital*, *born digital*), termin digitalizacja (*digitized*) rezerwując na określenie działań służących zamianom formy tradycyjnej (analogowej) na cyfrową, np. w procesie skanowania. W procesie digitalizacji otrzymuje się obiekty cyfrowe wtórne [Płoszajski, red. 2008, s. 15]. Takie rozróżnienie wprowadzono ze względu na to, że w bibliotekach powszechnie realizuje się procesy polegające na przenoszeniu treści druków do środowiska cyfrowego, nazywając je digitalizacją (lub dygitalizacją). Cyfryzacja jest często domeną autorów dokumentów. Cyfryzacja obecnie staje się coraz powszechniejsza, gdyż nawet dokumenty przeznaczone do rozpowszechniania w formie drukowanej pierwotnie tworzone są w formie cyfrowej, począwszy od „maszynopisu” tworzonych w edytorze tekstu, a kończąc na profesjonalnym składzie elektronicznym: skład ten następnie służy realizacji druku.

⁵ Ich liczba wzrasta przeciętnie o 3,25% rocznie, co oznacza podwajanie się ich ilości co 20 lat [Crow 2006].

⁶ Należy stwierdzić, że ruch Open Access rozwija się w sposób zróżnicowany, zależny na przykład od dziedziny, polityki naukowej danego kraju i jego instytucji naukowych. Odsetek tytułów Open Access w stosunku do ogólnej liczby czasopism elektronicznych sięga 10%, a ponad 19% z 1 346 000 artykułów publikowanych rocznie w czasopismach recenzowanych dostępnych jest w OA [Björk, Roos, Lauri 2008, s. 184].

⁷ Przykładem zmian w relacjach bibliotek z ich użytkownikami jest konsorcjum SCOAP (<http://scoop3.org/>), którego celem jest przeniesienie wszystkich czasopism z dyscypliny fizyki wysokich energii do OA.

fu lub telewizji, również mających za podstawę rozbudowaną infrastrukturę sieciową. Za gwałtownymi zmianami technologicznymi, w tym przypadku – w technologiach informacyjnych i komunikacyjnych (zwanych po angielsku ICT – Information and Communications Technology⁸) z trudem nadążają ewolucyjne w swoim charakterze zmiany społeczne, na przykład powstawanie i uświadomienie potrzeb informacyjnych. Rozwój nowych technologii informacyjnych, a w szczególności ich rozpowszechnienie, zależy od potrzeb społecznych, które zmieniają się dużo wolniej niż technika. W ten sposób działa prawo opóźnienia kulturowego (*social change*) Williama Ogburna: za zmianami w sferze technosfery nie nadążają zmiany logosfery (wartości, wzorów, znaczeń, instytucji) [Krzysztofek 2007, s. 18]. Siła inercji, nie tylko w stosunku do wynalazków, ale także nowych idei, czyli tzw. dysonans poznawczy⁹, jest znacząca [Krzysztofek 2005, s. 199; Bard, Söderqvist 2006, s. 55]. Innowacje technologiczne zazwyczaj znajdują uznanie małej grupy zapaleńców (2,5% użytkowników według Everetta Rogersa [Rogers 2003, s. 282]), zainteresowanych wszelkimi nowinkami, którym nie przeszkadza brak komercyjnej „ogłady” i niezawodności nowych produktów. Etap ten może trwać na tyle długo, że będą się pojawiać opinie o ograniczonym zastosowaniu nowości technologicznych¹⁰; zdarza się nawet, że nowa technologia nie znajduje rzeczywicie szerszego odbioru i zostaje zapomniana, albo wraca w nieco zmienionej postaci, po czasie niezbędnym do wykształcenia się nowej potrzeby u użytkowników; bywa też adaptowana tylko częściowo.

Dla pełnego wykorzystania nowych technologii i całkowitej zmiany stosunków społecznych wynikających z ich stosowania dochodzi więc w dłuższym okresie, niezbędnym do ich przyswojenia. Wcześniej powstają rozwiązania hybrydowe, łączące w sobie cechy starej i nowej technologii¹¹. Na tym właśnie etapie znajdują się obecnie biblioteki naukowe, zmuszone obsługiwać jednocześnie publikacje tradycyjne i cyfrowe do realizacji potrzeb informacyjnych swoich użytkowników [Pindłowa 2008, s. 152]. Równoległe zarówno biblioteki, jak i wszystkie inne instytucje związane z komunikacją naukową, przechodzą wspomniane zmiany, związane z pełnym wykorzystaniem nowych i ciągle pojawiających się (a więc nawet trudnych do przewidzenia) możliwości technologicznych.

Jak twierdzi Marcin Sieńko, Internet ma szansę stać się totalnym, globalnym archiwum ludzkiej wiedzy [Sieńko 2002, s. 85]. Pomimo trudności, związanych z długotrwałą archiwizacją informacji cyfrowej, o czym będzie jeszcze mowa (zob. p. 4.7), rozmiar tego archiwum oraz efekty skali, jakie generuje, często przerastają wyobraźnię obserwatorów. Pojawiają się wspomniane już głosy krytyczne, traktujące Internet jako chaotyczny zasób,

⁸ ICT jest ogólnym terminem, pod którym rozumie się wszystkie technologie służące komunikacji informacji. Obejmuje on technologie służące zapisowi informacji oraz jej rozprzestrzenianiu. Istnieje także termin Information Technology (IT), który dotyczy badania, projektowania, rozwoju, stosowania i zarządzania komputerowymi systemami informacyjnymi, rozumianymi w szczególności jako połączenie sprzętu i oprogramowania. IT często rozumiana jest jako zastosowanie sprzętu i oprogramowania komputerowego w sieciach rozległych do konwersji, zapisu, ochrony, archiwizacji, przetwarzania, transmitowania i bezpiecznego wyszukiwania danych i informacji.

⁹ Teoria dysonansu poznawczego głosi, że docieranie do człowieka sprzecznych elementów poznawczych (np. informacji, ocen, twierdzeń) powoduje nieprzyjemne napięcie psychiczne. W związku z tym dochodzi do wypierania części z nich – najczęściej tych nowych, do których człowiek się jeszcze nie przyzwyczaił, nie „połubił” ich.

¹⁰ Znane są opinie specjalistów, którzy po pojawieniu się Internetu uważali, że będzie miał on wyłącznie ograniczone zastosowanie militarno-naukowe, do czego wystarczy wykorzystanie kilku tysięcy komputerów na całym świecie; według ich przewidywań przemysł komputerowy miał przed sobą kiepskie perspektywy.

¹¹ Stosowany przeze mnie termin „hybrydyzacja” oznacza łączenie różnych produktów i związany jest z okresem przejściowym stosowania technologii, gdy wytwory starej i nowej technologii współistnieją w tym samym czasie. W bibliotekarstwie oznacza on możliwość funkcjonowania w dwóch środowiskach: elektronicznym i drukowanym. Jak pisze Lidia Szczygłowska, hybrydyzacja jest typowym zjawiskiem występującym we wszystkich okresach przejściowych, również w przyrodzie (np. w trakcie ewolucji) [Szczygłowska 2006].

pełen wszelkiego rodzaju informacji, mądrych i głupich. Człowiek, skupiwszy się na tych szcążkowych odpryskach wiedzy może zatracić intuicję i zdolność wykrzesania z siebie przeblisku geniuszu, niezbędnego w pracy naukowej. Jednak, jak zauważył amerykański ekonomista Peter Ferdinand Drucker, zwolennik postprzemysłowej gospodarki usług, mądrość i wiedza nie znajduje się ani w książkach, ani w Internecie [Morbitzer 2002]. Tam są jedynie dane i informacje. Mądrość i wiedza zawsze są atrybutami istoty ludzkiej, są zdobywane przez uczącą się osobę i przez nią wykorzystywane (zob. p. 1.4).

Badanie zachodzących zmian i uzyskanie poprawnych wyników, może mieć wpływ nie tylko na rozwój poszczególnych dyscyplin, ale nauki w ogóle. Wśród nauk badających zjawiska związane z komunikacją naukową, istnieje również miejsce dla nauki o informacji i bibliotekoznawstwa. Potrzeby badawcze w tym zakresie są dostrzegane przez decydentów, o czym świadczy tworzenie odpowiednich programów, na przykład ePolska [MNIł 2003] lub eEurope¹². Omawiane w tej książce zagadnienia wpisują się także w pejzaż informatycznej infrastruktury nauki (IIN)¹³, której budowa jest celem takich projektów, jak *Cyberinfrastructure* w USA oraz *e-Science* w Wlk. Brytanii i *Cyber-Science Infrastructure* w Japonii (zob. p. 1.3).

W książce przedstawiam rezultaty własnych badań w zakresie:

- publikowania elektronicznego i jego nowych form, w tym głównie bibliotek cyfrowych, repozytoriów, czasopism elektronicznych, autoarchiwizacji, blogów naukowych i szarej literatury;
- roli metadanych w opisie dokumentów elektronicznych dla ich efektywniejszego wyszukiwania i zarządzania;
- zmian ról uczestników komunikacji naukowej, w tym głównie bibliotekarzy oraz modelowania ontologicznego procesów służących komunikacji naukowej.

Może ona przyczynić się do wyjaśnienia sposobów transformacji komunikacji naukowej, wynikającej z przenoszenia jej form do środowiska elektronicznego. Jest też próbą wyjaśnienia wpływu zmian, dokonywanych w tej komunikacji, na sposób prowadzenia samych badań i prezentacji ich wyników. Efektem powinno być określenie tempa i kierunków zmian, ze szczególnym uwzględnieniem roli bibliotek naukowych.

Badania, przedstawione w kolejnych rozdziałach niniejszej książki, są kontynuacją moich dotychczasowych analiz; mają one dwa komplementarne cele:

1. Poznanie aktualnych problemów i tendencji w elektronicznej komunikacji naukowej, formalnej i nieformalnej¹⁴.

2. Wypracowanie modelu komunikacji naukowej, nazwanego przeze mnie globalną biblioteką cyfrową (GBC), w globalnej infrastrukturze nauki, uwzględniającego najnowsze tendencje i zjawiska, wynikające z implementacji ICT.

Pozna moim zainteresowaniem pozostała komunikacja nienaukowa, chociaż zdaję sobie sprawę ze znaczenia procesów cyfryzacji i digitalizacji w organizacji odbioru literatury pięknej, materiałów dydaktycznych na wszystkich etapach kształcenia lub literatury

¹² http://ec.europa.eu/information_society/europe/2005/index_cn.htm.

¹³ Wyras „informatyczna” w terminie „Informatyczna Infrastruktura Nauki” może mylić czytelnika co do zakresu tej książki; jak to wyjaśniam w p. 1.3. opisywać w niej będę jedynie tę część IIN, która tworzy infrastrukturę informacyjną, obsługującą e-Naukę. Odpowiada to w dużym zakresie koncepcji IIN Christine Borgman, przedstawionej w dalszej części Wprowadzenia.

¹⁴ Na podstawie przeglądu literatury przedmiotu Christina Pikas uważa, że formalna komunikacja naukowa odbywa się za pośrednictwem materiałów publikowanych, recenzowanych, podlegających obróbce redakcyjnej i możliwych do wyszukania w systemach informacyjnych. Komunikacja nieformalna to taka, która nie spełnia któregoś z wymienionych warunków [Pikas 2006]. Według Arthura Meadowsa formalna komunikacja naukowa umożliwia dostęp do treści dużej grupie odbiorców przez długi czas, natomiast komunikacja nieformalna jest często efemeryczna i dostępna tylko dla ograniczonej grupy odbiorców [Meadows 1998, s. 7]. Wydaje się, że Internet sprzyja zacieraniu granic między oboma rodzajami komunikacji naukowej, chociaż granice te nigdy nie były jednoznacznie wytyczone.

popularnonaukowej. Wybór ten jest usprawiedliwiony zauważalnym różnicowaniem się systemu obiegu informacji naukowej i innej [Feather 2003, s. 208]. Ze względu na to ograniczenie, swoje rozważania wiąże głównie z instytucjami nauki, w tym z bibliotekami naukowymi, które wśród innych bibliotek mają największy wpływ na funkcjonowanie GBC. Tam, gdzie omawiane zagadnienia dotyczą różnych bibliotek, nie tylko naukowych, zaznaczyłem to przez użycie terminu „biblioteka” bez przymiotnika „naukowa”.

Powodem refleksji jest konieczność dostosowania wymiany informacji naukowej do współczesnych możliwości technologicznych, oferowanych przez narzędzia informatyczne, a w szczególności sieci rozległe. Komunikacja naukowa rozumiana jest jako zespół formalnych i nieformalnych procesów, służących wymianie wiedzy, nowo powstałej w wyniku badań własnych uczonego, opartych na obserwacji i weryfikowalnych faktach, co stwarza płaszczyznę dyskursu naukowego [Owen 2007, s. 27], dzięki dokumentowaniu, ocenie, redagowaniu, formatowaniu, dystrybucji, organizowaniu, udostępnianiu, archiwizacji, wykorzystaniu i transformacji tej wiedzy. Opis i modelowanie tych zjawisk jako części procesu badawczego ułatwia dostosowanie komunikacji naukowej do bieżących i przyszłych potrzeb nauki bez względu na dziedzinę wiedzy, w której proces jest realizowany. Komunikację naukową należy odróżnić od publikowania naukowego, które jest częścią tej komunikacji, a polega na przekazywaniu informacji przy użyciu nośnika, służącego utrwaleniu wiedzy [Hahn 2008].

Na potrzeby przedstawionej publikacji przyjąłem następujące tezy, których poprawność ma być zweryfikowana w trakcie realizacji badań:

- Przyszłością komunikacji naukowej jest środowisko cyfrowe, choć dla jego pełnego wykorzystania niezbędne jest rozwiązanie wielu problemów (na przykład udoskonalenie metod długotrwałej archiwizacji).
- Cyfryzacja komunikacji naukowej powoduje istotne zmiany w jej organizacji.
- Zmiany w komunikacji naukowej powiązane są ze zmianami w organizacji nauki, szczególnie z budową informatycznej infrastruktury nauki oraz oddolnymi inicjatywami Nauki 2.0.
- Zmiany powodują między innymi powstawanie jednolitych (z punktu widzenia zaspokajania potrzeb użytkowników) zasobów informacji, które stanowią podstawę globalnej biblioteki cyfrowej (GBC).
- W wyniku cyfryzacji komunikacji naukowej zatarciu ulegają różnice między komunikacją formalną i nieformalną, realizowaną w GBC.

Analizy, służące potwierdzeniu lub zaprzeczeniu postawionych tez, wiązały się z poszukiwaniem odpowiedzi na następujące pytania badawcze:

- Jakie zmiany powoduje realizacja procesów komunikacji naukowej w środowisku cyfrowym?
- Czy globalna biblioteka cyfrowa (GBC), jako narzędzie realizacji komunikacji naukowej, pozwoli na unifikację praktyk w tym zakresie, prowadząc do połączenia wysoko sformalizowanych zasad publikowania zapewniającego ocenę jakości oraz bardziej nieformalnych i zindywidualizowanych form komunikowania, opartych na wewnętrznych i zróżnicowanych zasadach przyjętych przez środowisko uczonych?
- Jakie jest miejsce technologii informacyjno-bibliotecznej, w tym metadanych, ontologii, identyfikatorów w przyszłym modelu komunikacji naukowej?
- Jaka jest przyszłość publikacji elektronicznych i sposobów ich wyszukiwania w sieciach globalnych?

Nowy model komunikacji naukowej, przedstawiony w książce, może pomóc w usprawnieniu wielu procesów informacyjnych w istniejącym obecnie okresie przejściowym od mediów tradycyjnych do elektronicznych. Model ten ma stanowić połączenie modelowania procesów tworzenia dokumentów prymarnych (oryginalnych) oraz pochodnych (opisów dokumentów prymarnych tworzonych w ośrodkach informacji, metadanych). W ten

sposób zostaną równocześnie uwzględnione procesy należące do działalności naukowej (publikowanie) i informacyjnej (tworzenie zasobów informacyjnych). Dotychczas oba te elementy komunikacji naukowej badane, opisywane i modelowane były oddzielnie.

Realizacja moich badań może wspomóc także usprawnienie komunikacji naukowej poprzez skonkretyzowanie tendencji i prognoz jej rozwoju. Uzasadniają to następujące przesłanki:

1. Zmiany w nauce, w tym konieczność i możliwość podejmowania przez pracowników nauki współpracy o zasięgu światowym, czego jednym przejawem są tzw. kolaboratoria, stanowiące element informatycznej infrastruktury nauki (IIN) a innym tworzenie nieformalnych grup w obrębie Nauki 2.0.

2. Kryzys komunikacji naukowej, objawiający się niezadowolającą efektywnością jej dotychczasowych narzędzi, głównie tradycyjnych czasopism naukowych.

3. Szybkie zmiany w komunikacji naukowej, widoczne w krajach gospodarczo rozwiniętych (a od pewnego czasu także w Polsce), przejawiające się takimi zjawiskami, jak cyfryzacja publikacji naukowych, tworzenie różnego rodzaju repozytoriów oraz bibliotek cyfrowych, Open Access, Semantyczny Web, Web 2.0 i Nauka 2.0.

4. Integracja w globalnych sieciach komputerowych, rozproszonych wcześniej elementów nauki i komunikacji naukowej.

5. Zmiany w sposobach finansowania publikacji oraz nieadekwatność zasad prawa autorskiego.

Dla bibliotekoznawstwa i informacji naukowej właściwe, aktywne uczestnictwo w tych zmianach ma znaczenie zasadnicze, gdyż wymusza całkowitą zmianę paradygmatu działania, aczkolwiek bibliotekarze od dawna wykonują wiele swoich zadań w skali globalnej, a przynajmniej ponadlokalnej [Pindłowa 2005, s. 47]. Wydaje się, że współpraca na wielu poziomach, w tym globalnym, jest jednym z głównych paradygmatów współczesnego bibliotekarstwa. Mamy do czynienia ze zmianą cywilizacyjną, przejściem do globalnego sieciowego społeczeństwa informacyjnego, do której należy się odpowiednio przygotować.

Zmiany te prowadzą do powstania globalnej (naukowej) biblioteki cyfrowej, przez którą rozumieć ogólnosiwiatowy zasób naukowych obiektów cyfrowych, gromadzony, opracowywany, przechowywany i udostępniany w sposób rozproszony, nieinstytucjonalny¹⁵ i zdalny w sieci globalnej (zob. też p. 1.4). Globalna biblioteka cyfrowa jest siecią – bez Internetu nie może istnieć. Sieciowy i globalny charakter biblioteki cyfrowej powoduje zmiany w sposobach realizacji podstawowych funkcji biblioteki naukowej, które są ściśle powiązane z kierunkami i tempem zmian w komunikacji społecznej w ogóle, a naukowej w szczególności. Jej cyfrowy i sieciowy charakter powoduje, że niezbędne są także nowe narzędzia i metody realizacji tych funkcji. Opis funkcji oraz sposobów ich realizacji, przy zastosowaniu odpowiednich metod i narzędzi, stanowi główną część niniejszej książki.

Treść książki można podzielić na dwie części. W rozdziale pierwszym przedstawiłem założenia, dotyczące funkcjonowania współczesnej nauki i jej infrastruktury informatycznej oraz miejsca globalnej biblioteki cyfrowej w tej strukturze, jako narzędzia komunikacji naukowej. Przedstawiłem koncepcję globalnej biblioteki cyfrowej, jej definicję oraz elementy składowe.

Na strukturę drugiej części książki (rozd. 2 – 6) bezpośredni wpływ miał model 5S bibliotek cyfrowych, opisany w dalszej części rozdziału. Scharakteryzowałem w niej sieciową komunikację naukową, realizowaną w strukturze, nazwanej przeze mnie globalną biblioteką cyfrową. W każdym z czterech rozdziałów omówiłem wybrane aspekty GBC.

¹⁵ O „nieinstytucjonalności” mówimy w takim sensie, że brak określonej instytucji zarządzającej globalną biblioteką cyfrową jako całością, chociaż jej poszczególne zasoby mogą mieć charakter instytucjonalny.



Spśród trzech warstw Internetu: fizycznej, logicznej i treści [Hofmokl 2009, s. 123] główną uwagę skupiłem na tej trzeciej warstwie, decydującej o kształcie GBC, w razie konieczności przedstawiając także elementy warstwy logicznej.

Zastosowanie modelu 5S do opisu GBC ułatwiło holistyczne potraktowanie zagadnień współczesnej komunikacji naukowej, z drugiej jednak strony spowodowało potrzebę powtórzeń, na co wskazują w tekście poprzez odesłania do innych części (rozdziałów) książki. Powtórzeń takich nie można byłoby zapewne uniknąć zupełnie, gdyż bez nich trudno jest przy pomocy linearnego tekstu książki opisać sieciową strukturę GBC.

W rozdziale 2 przedstawione zostały najbardziej podstawowe elementy GBC, począwszy od prostych zestawów bitów i bajtów, poprzez zagadnienia formatów, pozwalających na odczyt treści zapisanych w sposób binarny, języki kodowania treści w Internecie, po protokoły umożliwiające przekazywanie danych i metadanych, głównie w trakcie wyszukiwania informacji. Dalszą część tego rozdziału poświęciłem sposobom organizacji informacji w struktury różnych poziomów, począwszy od obiektów cyfrowych, prostych i złożonych, poprzez dokumenty i publikacje, a skończywszy na strukturach metadanych.

W kolejnym rozdziale opisałem przestrzenie, w jakich funkcjonuje GBC. Przez analogię do tradycyjnej biblioteki naukowej konkretyzuję miejsce zajmowane przez globalną bibliotekę cyfrową. Miejsce to definiowane jest, między innymi, przez przestrzenie nazw i identyfikatory obiektów cyfrowych, stosowane w cyberprzestrzeni. Mowa jest także o miarach, służących ocenie GBC oraz interfejsach, wyznaczających sposób postrzegania przestrzeni GBC przez użytkowników.

W rozdziale 4 przedstawiłem podstawowe procesy, realizowane w ramach GBC i jej serwisów. Działania te, z wyszukiwaniem i udostępnianiem informacji (którym poświęciłem najwięcej miejsca) na czele, stanowią zasadniczy sens istnienia globalnej biblioteki cyfrowej, podobnie jak każdej innej biblioteki. Inaczej niż w bibliotece tradycyjnej, częścią GBC są także procesy publikowania informacji poprzez ich udostępnianie w Sieci. W środowisku cyfrowym szczególnego znaczenia nabierają także takie procesy, jak zapewnienie integralności i autentyczności obiektów cyfrowych, współdziałanie, archiwizacja i ocena, opisane w tym rozdziale z uwzględnieniem ich roli w globalnym środowisku informacyjnym. Mowa jest także o serwisach GBC nowego rodzaju, realizujących idee Semantycznego Webu i Webu 2.0.

Rozdział 5 poświęciłem społecznościom funkcjonującym w obrębie GBC, ich nowym rolom, a także nowym możliwościom i zagrożeniom, jakie niesie dla nich pełna cyfryzacja komunikacji naukowej i funkcjonowanie GBC. Omówione zostały także zagadnienia ekonomiczne i prawne, warunkujące funkcjonowanie i rozwój społeczności w GBC.

W ostatnim rozdziale przedstawiłem model funkcjonowania komunikacji naukowej w GBC, uzupełniony o wnioski dotyczące obecnego stanu i możliwości rozwoju globalnej biblioteki cyfrowej, powstałe w trakcie pracy badawczej. Model opisuje stan obecny komunikacji naukowej, jest więc ujęciem, fotografią stanu przejściowego w okresie dynamicznych zmian. Przez porównanie ze stanem poprzednim może służyć przewidywaniu kierunków aktualnie zachodzących zmian.

W Zakończeniu odniosłem się do rozwiązań proponowanych przez autorów modelu 5S opisując możliwe kierunki rozwoju sieciowej komunikacji naukowej w wyniku umocnienia obecnych tendencji, głównie w formie globalnej biblioteki cyfrowej. Opisałem możliwe rozwój takich elementów GBC, jak: zasoby danych i informacji, narzędzia wyszukiwawcze, serwisy i usługi oraz społeczności.

Przedstawione przeze mnie wyniki opierają się na wcześniej realizowanych **badaniach bibliotek cyfrowych**. Jak stwierdza Christine Borgman, badania te zawsze miały charakter interdyscyplinarny [Borgman 2003a, s. 36]. Informatycy zajmują się głównie stosowanymi technologiami, w szczególności sieciowymi. Bibliotekoznawcy i specjaliści z zakre-

su informacji naukowej skupiają się na treściach, organizacji informacji, zachowaniach użytkowników i zasadach (modelach) publikowania. Specjaliści w zakresie socjologii i ekonomii koncentrują zaś uwagę na zagadnieniach społecznych i modelach ekonomicznych. Takie zagadnienia, jak interakcja człowiek-komputer, projektowanie interfejsów lub dostarczanie usług, interesują natomiast przedstawiciele wszystkich wymienionych dyscyplin. Realizacja projektów, związanych z tworzeniem bibliotek cyfrowych, też opiera się zazwyczaj na doświadczeniach i wynikach badań, uzyskiwanych w wielu dyscyplinach naukowych.

Christine Borgman jest najbardziej znaną autorką, piszącą od wielu lat z perspektywy socjotechnicznej o globalnej bibliotece cyfrowej. Nie ona pierwsza jednak wystąpiła z ideą „globalnego systemu informacyjnego”. Jego początki niektórzy autorzy upatrują już w XV w., wraz z wynalezieniem ruchomych czcionek i odkryciami nowych kontynentów [Neelameghan, Tocatian 1985]. Inni piszą nawet o Arystotelesie i Tomaszu z Akwinu jako osobach, które pierwsze wyrażały wizję możliwości dostępu do całości wiedzy ludzkiej [Veltman 2005, s. 17]. W artykule opublikowanym w 1883 r. Charles Cutter próbował przewidywać wygląd biblioteki za 100 lat, czyli w 1983 r. Pisał o czytelnikach zasiadających za biurkami wyposażonymi w małe klawiatury, za pomocą których łączą się z centralnym katalogiem. Pisał także o sieciach bibliotek połączonych sieciami telekomunikacyjnymi, dzięki czemu dostęp do ich zasobów jest tak łatwy, że „wszystkie biblioteki w kraju stają się w praktyce jedną biblioteką” [Wright 2007, s. 182]. Paul Otlet w swoich pismach na początku XX w. przedstawiał podobne idee powszechnego dostępu do wiedzy, na przykład sieci bibliograficzne, dzięki którym wiązany jest z koncepcją hipertekstu¹⁶ [Rayward 2005]. Tuż przed Vanneverem Bushem opisał i konstruował technologię o nazwie „Mundaneum”, która umożliwić miała taki dostęp. Była to maszyna, pozwalająca użytkownikowi, siedzącemu w swoim fotelu, na poznawanie świata poprzez patrzenie w ekran [Otlet 1935, s. 390-391]. Umożliwiać miała nie tylko wyszukiwanie dokumentów, ale także tworzenie relacji¹⁷, sieci połączeń między nimi, w wyniku czego powstaje jedna, Uniwersalna Książka. Tworzenie relacji wspomaganie było schematem klasyfikacji. W 1937 r. model maszyny matematycznej, służącej głównie modelowaniu obliczeń matematycznych przedstawił Alan Turing. W 1945 r., gdy dopiero powstawały pierwsze elektroniczne maszyny łączące (MARK I, ENIAC), Vannevar Bush opisał możliwości do realizacji model zwiększający efektywność komunikacji naukowej, o nazwie „Memex”, stwierdzając, że możliwe jest utworzenie nowego, globalnego systemu magazynowania i wyszukiwania informacji naukowej [Bush 1945]. Za najcenniejszą cechę swojego systemu uważał Bush możliwość łączenia dokumentów za pomocą dowolnych asocjacji, co można uznać za kolejną zapowiedź hipertekstu. W odróżnieniu od linków Otleta, jego połączenia funkcjonować miały wyłącznie pomiędzy dokumentami. W jego pracach znalazły się także wyobrażenia komputerów osobistych, narzędzi indeksujących oraz wizualizacji informacji. Pomimo tego, że Memex nigdy nie został zbudowany, Bush jako jeden z pierwszych dostrzegł możliwości tkwiące w połączeniu komputerów z pamięciami masowymi i technikami indeksowania [Gawrysiak 2008a, s. 175].

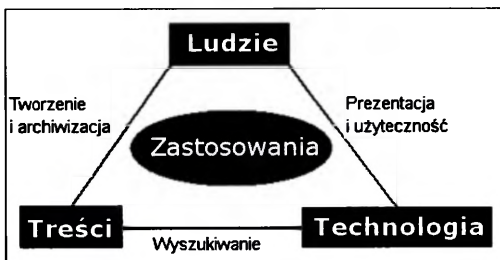
IFLA przez kilka dziesięcioleci promowała programy „powszechnego dostępu do pu-

¹⁶ Termin „hipertekst” po raz pierwszy pojawił się w pracach Theodora Nelsona w latach sześćdziesiątych XX wieku. Odnosił się do formy tekstu elektronicznego, nowej technologii informacyjnej oraz sposobów publikowania. Nelson za hipertekst uważał pisanie nieliniarne, tekst rozgałęziający się, umożliwiający dokonywanie wyborów przez czytelnika. Autonomiczne części tekstu łączone są odnośnikami, pozwalając użytkownikowi na wybór własnej ścieżki poznawania tekstu [Nelson 1993]. Hipertekst konceptualizowany jako tekst nieliniary, zawierający połączenia wewnątrztekstowe pomiędzy włączonymi w tekst odnośnikami. Cechą szczególnie odróżniającą hipertekst od innych tekstów jest jednak automatyczne wyszukiwanie konceptualnie powiązanych tekstów, co pozwala na ich nieograniczone kombinacje [Tredinnick 2007, s. 181]. Kombinacje te dają czytelnikowi nieograniczone możliwości wyboru własnej drogi poznania treści powiązanych dokumentów.

¹⁷ Relacje te Paul Otlet nazwał „linkami”.

blikacji” i „powszechnej rejestracji bibliograficznej” – UBC (Universal Bibliographic Control), a obecnie kontynuuje prace w kierunku „globalnych wspólnot informacyjnych” [Byrne 2005]. W celu koordynacji międzynarodowej działalności informacyjnej funkcjonuje wiele podobnych programów, dzięki czemu możliwa jest identyfikacja i lokalizacja materiałów publikowanych, szarej literatury, zasobów archiwalnych i obiektów muzealnych [Borgman 2003a, s. 237].

Początki zainteresowania bibliotekami cyfrowymi znaleźć można w USA, gdzie w latach osiemdziesiątych ubiegłego wieku NSF¹⁸ rozpoczęła finansowanie projektów, dotyczących badań nad rozwiązaniami sieciowymi i infrastrukturą informatyczną, głównie superkomputerami. Efektem tych prac, w powiązaniu z działaniami międzynarodowymi i na poziomie rządowym, było powstanie Internetu jako światowej sieci. Wraz z rozwojem systemów wyszukiwania informacji, baz danych, sztucznej inteligencji, badań nad interakcjami człowiek-komputer, na początku lat dziewięćdziesiątych, rozpoczęły się bardziej zintegrowane, interdyscyplinarne badania w tym zakresie. Na potrzebę prowadzenia takich badań zwracali uwagę członkowie DELOS-NSF Working Group, proponując interdyscyplinarne programy badawcze realizowane według zagadnień przedstawionych symbolicznie na rys. 1 w postaci boków i wnętrza trójkąta [Chen, Kiernan 2002]. Obszar badawczy pomiędzy wierzchołkami tego trójkąta, symbolizującymi ludzi i treści, jest odzwierciedleniem tworzenia treści i ich archiwizacji. Obszar pomiędzy treściami i technologią obejmuje efektywne wyszukiwanie treści z wykorzystaniem technologii. Badania dotyczące prezentacji i użyteczności mają wpływ na wzrost dostępności treści. Efektywne zastosowanie wyników badań pozwala na integrację trzech powiązanych ze sobą obszarów.



Rys. 1. Interdyscyplinarne badania nad bibliotekami cyfrowymi (wg [Chen, Kiernan 2002])

Współautorem modelu przedstawionego na rys. 1 jest Ching-chih Chen, która od 1993 r. prezentuje kolejne koncepcje globalnej biblioteki cyfrowej [Chen 1998; Chen 2003; Chen 2007]. Początkowo przedstawiała ona technologiczne możliwości łączenia odrębnych kolekcji zasobów cyfrowych, posiadanych głównie przez biblioteki tradycyjne. Później zajmowała się problemami tworzenia zasobów cyfrowych i sposobami ich wykorzystania. Chen uważa, że globalna biblioteka cyfrowa nie jest pojedynczą biblioteką cyfrową, ale siecią wielu rozproszonych zasobów, z których każdy może zawierać wiele wysokiej jakości multimedialnych obiektów cyfrowych.

Jak stwierdzono w raporcie NSF z 2003 r., biblioteki cyfrowe wymyślone i zaprojektowane zostały w latach dziewięćdziesiątych jako odpowiedź na powstanie sieciowego

¹⁸ NSF – National Science Foundation (<http://www.nsf.gov>), niezależna agenda federalna USA utworzona przez Kongres w latach pięćdziesiątych XX w. Finansuje około 20% badań wspomaganych centralnie, realizowanych przez amerykańskie uczelnie i instytucje naukowe.

środowiska wiedzy. Od tego czasu dokonano znacznego postępu w ich tworzeniu, jednak cel zdawał się oddalać wraz z postępem prac, w dodatku okazał się znacznie bardziej dynamiczny, niż początkowo przewidywano [Larsen, Wactlar 2003, s. 9].

Jednym z pierwszych projektów badawczych była Digital Library Initiative (DLI), realizowana w latach 1994-1998 przez NSF, DARPA i NASA, która przygotowywała wizję rozwoju w oparciu o cztery cele:

- Cyfrowa sieć systemów wiedzy – połączenie zasobów obliczeniowych, informacyjnych i ludzkich.
 - Zestaw technologii umożliwiających rozwój – służących tworzeniu, rozpowszechnianiu i stosowaniu wiedzy w zróżnicowanym środowisku multimedialnym, ukierunkowanym na realizację potrzeb użytkowników.
 - Nowe usługi informacyjne – w zakresie edukacji w sieci, handlu, ochrony zdrowia, transportu, działalności władz, wykraczające poza zakres tradycyjnych usług bibliotek i ośrodków informacji.
 - Powszechność, otwartość i personalizacja – dostęp przez całą dobę poprzez sieć.
- Badania prowadzone do końca XX wieku zostały rozszerzone poza obszary zagadnień społecznych, behawioralnych i instytucjonalnych, dominujących w bibliotekach tradycyjnych. Projekty dotyczące bibliotek cyfrowych angażowały wielu partnerów, reprezentujących różne dyscypliny, jednocześnie przyczyniając się do rozwoju wielu gałęzi nauk technicznych i informatyki.

Dla zwiększenia efektywności realizowanych projektów stało się konieczne gromadzenie i rozpowszechnianie uzyskiwanych wyników. W 1995 r. dla tych celów powstało czasopismo elektroniczne D-Lib Magazine¹⁹, stając się integralną częścią działań DLI. Ten (początkowo) miesięcznik, a obecnie dwumiesięcznik, stał się profesjonalnym wydawnictwem, przedstawiającym wyniki badań i dyskusje nad wszystkimi problemami, z którymi stykali się uczeni, zajmujący się bibliotekami cyfrowymi.

Projekt DLI kontynuowany był w latach 1998-2003 jako DLI-2. Do jego realizatorów dołączyły Library of Congress i National Library of Medicine (USA). Finansowanych było 36 projektów badawczych, obejmujących pełen zakres zagadnień związanych z bibliotekami cyfrowymi, takich jak badania podstawowe, tworzenie treści i zasobów cyfrowych, aplikacje dziedzinowe, instalacje eksperymentalne i rozwiązania funkcjonalne [Chen 2004, s. 160-161]. Projekty dotyczyły zagadnień związanych z całym cyklem życia informacji – jej tworzeniem, udostępnianiem, rozpowszechnianiem, wykorzystaniem i archiwizacją, a także tworzeniem wskaźników obrazujących poprawność rozwiązań. Wielką zaletą DLI-2 było interdyscyplinarne podejście do realizowanych projektów, wysoki poziom ich integracji oraz ścisła współpraca nauki z sektorem prywatnym [Griffin 2005].

NSF nadal finansuje badania bibliotek cyfrowych za pomocą kilku programów. Jednym z nich jest NSDL (National Science Digital Library)²⁰, oparty na osiągnięciach DLI-2; jego celem jest utworzenie zasobów dla edukacji w naukach ścisłych, technice i matematyce. W ramach projektu tworzony jest m.in. DLESE (Digital Library for Earth System Education)²¹, w którym współpracują uczeni, nauczyciele i studenci.

Z końcem lat dziewięćdziesiątych XX w. rozpoczęto badania nad nowymi rodzajami informacji i użytkowników. Do prac prowadzonych w USA włączyli się specjaliści z Wlk. Brytanii (JISC²²) i Niemiec (DFG²³), nowe międzynarodowe projekty powstawały

¹⁹ <http://www.dlib.org/dlib/>.

²⁰ <http://nsdl.org/>.

²¹ <http://www.dlesc.org/library/index.jsp>.

²² JISC – Joint Information Systems Committee (<http://www.jisc.ac.uk/>), zajmuje się promowaniem badań dotyczących nowych technologii informacyjnych, w tym rozwojem JANET, dostępem do źródeł elektronicznych, nowego środowiska dla nauczania, edukacji i badań naukowych.

²³ DFG – Deutsche Forschungsgemeinschaft (German Research Foundation – <http://www.dfg.de>), jest niemiecką, centralną instytucją finansującą badania naukowe w publicznych instytucjach naukowych.

w Unii Europejskiej i w Azji. Najbardziej znanym projektem finansowanym przez JISC był, realizowany na przełomie wieków, eLib Programme²⁴. W jego ramach prowadzono badania nad możliwymi zmianami w światowym systemie komunikacji i publikowania naukowego oraz miejscem bibliotek w tych przemianach. Był także znacznie wyraźniej, niż badania realizowane w USA, ukierunkowany na budowę scentralizowanych, narodowych projektów.

Według Marcosa Gonçalvesa w publikacjach odnoszących się do bibliotek cyfrowych brak jest wciąż formalnych modeli dotyczących tej metody publikowania elektronicznego [Gonçalves i in. 2004, s. 275]. Ten być może zaskakujący fakt niektórych tłumaczy szczególnie stopniem skomplikowania tego zagadnienia. Jedną z pierwszych prób zmiany tego stanu rzeczy były prace Bing Wang [Wang 1999]. Zaproponował on tzw. podejście hybrydowe, definiując bibliotekę cyfrową jako połączenie specjalnej bazy danych oraz interfejsu użytkownika opartego na hipermediach.

David Levy i Catherine Marshall przedstawili model konceptualny dla bibliotek cyfrowych, zawierający trzy powiązane ze sobą elementy: dokumenty, technologię i pracę [Levy, Marshall 1995]. Zakwestionowali oni powszechnie spotykane opinie dotyczące bibliotek cyfrowych: to, że zawierają one tylko dokumenty niezmiennie, przechowywane permanentnie; że wykorzystywane są przez osoby pracujące indywidualnie; że funkcjonują w nich tylko technologie cyfrowe. Ich analizy uwidoczniły ograniczenia nakładane na biblioteki cyfrowe, jeśli konceptualizowane są na tych samych zasadach, co tradycyjne wyszukiwanie informacji.

W modelu przedstawionym przez Dagoberta Soergela system biblioteki cyfrowej wychodzi znacznie poza zwykłą obsługę tekstów i obrazów [Soergel 1998]. Zawiera on wiele rodzajów obiektów oraz bogatą sieć powiązań – od pojęć i ludzi do organizacji, projektów, dokumentów; od zapytań do dowolnego obiektu wyszukanego lub wybranego przez użytkownika; od projektów do dokumentów; od części dokumentu do części innego dokumentu. Później Dagobert Soergel przedstawił model konceptualny zawierający trzy podstawowe zasady i jednocześnie zagadnień szczegółowych [Soergel 2002].

Gary Marchionini twierdził, że należy podkreślać rolę bibliotek cyfrowych w organizacji grupowego rozwiązywania problemów i rozpowszechniania wiedzy [Marchionini 1999]. Tego typu model biblioteki cyfrowej nazwał rozprzestrzenianiem (*sharium*).

Leonid Kalinichenko przedstawił kanoniczny model systemów informacyjnych oraz podejście kompozycyjne, które zastosował w celu opisu sposobów współdziałania bibliotek cyfrowych [Kalinichenko i in. 2000]. Wyróżnione zostały narzędzia modelowania metainformacji, przydatne w bibliotekach cyfrowych, takie jak informacje tekstowe i multimedialne, heterogeniczne bazy danych, informacja ontologiczna, informacja nieustrukturyzowana i częściowo ustrukturyzowana.

W artykule Michaela Biebera i in. zaproponowano punkt widzenia na biblioteki cyfrowe nazwany Collaborative Knowledge Evolution Support System (CKESS) [Bieber i in. 2002]. CKESS został opisany jako repozytorium cyfrowe, w ramach którego społeczność użytkowników współpracuje w celu wymiany i modyfikacji swojej wiedzy.

Howard Besser przedstawił model konceptualny, w którym podkreślił rolę i problemy wynikające z funkcjonowania współdziałających repozytoriów dokumentów ponad wyizolowanymi kolekcjami [Besser 2002]. Współdziałanie odnosi się do synchronicznego wyszukiwania w wielu repozytoriach lub asynchronicznej agregacji informacji z wielu repozytoriów do jednego repozytorium przez inteligentnych agentów.

William Arms, badacz bardzo zasłużony dla prac nad bibliotekami cyfrowymi, stosował między innymi metodę analizy punktu widzenia, wykorzystywaną podczas tworzenia oprogramowania [Arms 2005]. Wyróżnił on trzy punkty widzenia biblioteki cyfrowej: organizacyjny, techniczny i użytkownika. Twierdził, że pełny, holistyczny obraz biblioteki

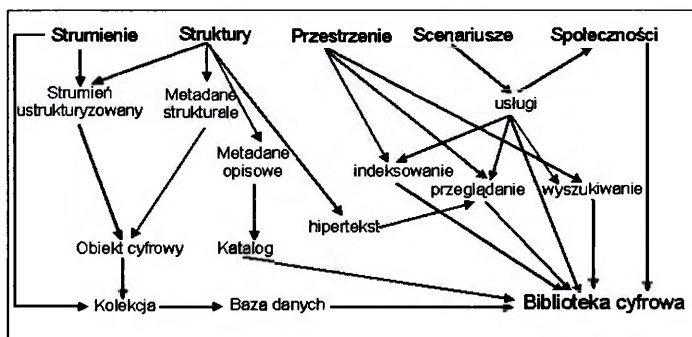
²⁴ <http://www.ukoln.ac.uk/services/elib/>.

cyfrowej mogą przynieść badania, rozpoczynające się od użytkownika. Wyniki takich badań prowadzą do wniosku, że tak jak ponad 20 lat temu odrębne, niezależne sieci komputerowe zaczęły być łączone w jeden, zunifikowany Internet, podobnie obecnie być może nadszedł czas, aby takiej samej przemiany dokonać w stosunku do bibliotek cyfrowych, tworząc z nich jedną globalną bibliotekę cyfrową.

Zgodnie z pracami prowadzonymi w ramach projektu DELOS²⁵ biblioteka cyfrowa jest rozwijającą się organizacją, która powstaje w wielu etapach. W ten sposób tworzone są trzy podstawowe, odrębne systemy: biblioteka cyfrowa, system biblioteki cyfrowej i system zarządzania biblioteką cyfrową [Candela i in. 2007]. Biblioteka cyfrowa gromadzi, zarządza i archiwizuje treści cyfrowe oraz oferuje swoim użytkownikom odpowiedniej jakości, wyspecjalizowane funkcje realizowane na tych treściach, zgodnie z przyjętą polityką. System biblioteki cyfrowej jest to oprogramowanie oparte na zdefiniowanej architekturze i zapewniające funkcjonalność niezbędną dla danej biblioteki cyfrowej. System zarządzania biblioteką cyfrową jest to oprogramowanie wyższego poziomu, które zapewnia odpowiednią infrastrukturę programową dla tworzenia i zarządzania systemem biblioteki cyfrowej oraz integracji dodatkowego oprogramowania, dającego możliwość implementacji nowych funkcji.

Obecnie w zakresie bibliotek cyfrowych funkcjonuje model 5S, nazwany od pierwszych liter angielskich nazw podstawowych elementów modelu – Streams (Strumienie), Structures (Struktury), Spaces (Przestrzenie), Scenarios (Scenariusze) i Societies (Społeczności) [Fox 2005]. Zgodnie z tą strukturą biblioteka cyfrowa zawiera repozytorium, zasoby metadanych, usługi oraz społeczność użytkowników; autorzy założyli, że minimalnym zestawem elementów dla biblioteki cyfrowej, a właściwie dla każdego systemu informacyjnego są zbiory i funkcje [Fox, Kipp, Mather 1996]. Zbiory są kolekcjami unikalnych (niepowtarzalnych) jednostek, mających charakter obojętny. Funkcje są działaniami mogącymi tworzyć i usuwać zbiory, a także wprowadzać i usuwać jednostki do/ze zbiorów.

Strumienie i Struktury służą konstrukcji obiektów cyfrowych, Przestrzenie opisowi kolekcji obiektów cyfrowych i ich wzajemnych powiązań, Scenariusze określają sposób zmiany stanu systemu w wyniku działania usług, a Społeczności pozwalają opisać połączenia ról i działań w obrębie grup użytkowników (rys. 2).



Rys. 2. Metamodel biblioteki cyfrowej wg zasad 5S (wg [Fox 2005])

²⁵ <http://www.delos.info/>.

Niektórzy autorzy wiążą rozwój bibliotek cyfrowych z modernizacją infrastruktury nauki. Ch. Borgman proponuje uznać za globalną bibliotekę cyfrową konstrukt, który obejmuje wszystkie lokalne zasoby cyfrowe, połączone i dostępne poprzez globalną infrastrukturę informacyjną, tak, aby użytkownik korzystający z tej infrastruktury miał możliwość wyszukiwania we wszystkich zasobach [Borgman 2003a, s. 48]. Obecnie rolę tej infrastruktury pełni odpowiednia część IIN. Autorka wyróżnia cztery obszary-kierunki badań tej infrastruktury z perspektywy zagadnień informacyjnych:

- Zagadnienia związane z treściami naukowymi w długiej perspektywie czasowej.
- Zrównoważenie potrzeb i wymagań w skali lokalnej i globalnej.
- Oddzielenie treści od narzędzi i usług.
- Określenie narzędzi, technologii oraz mechanizmów instytucjonalnych, zapewniających odpowiedni poziom spójności i kontroli treści [Borgman 2007, s. 245].

Twierdzi, że nie jest to spójny plan badań infrastruktury informacyjnej służącej badaniom naukowym, a jedynie właściwy początek do tworzenia infrastruktury dla informacji (*for information*), w odróżnieniu od infrastruktury informacyjnej (*of information*).

Chern Li Liew przedstawiła badania nad bibliotekami cyfrowymi w latach 1997-2007 [Liew 2009, s. 249]. Nie zajmowała się ona zagadnieniami technicznymi, a wyłącznie społecznymi. W tym zakresie najwięcej było badań dotyczących zagadnień organizacyjnych i ekonomicznych (38,5% odnalezionych artykułów), użytkowników bibliotek cyfrowych (34,5%), problemów prawnych (prawie 20%) oraz etycznych i społeczno-kulturowych (po ok. 3,5%). Ponad 20% artykułów opublikowano we wspomnianym już czasopiśmie *D-Lib Magazine*; jego znaczenie potwierdzają też inne badania [Singh, Mittal, Ahmad 2007, s. 345]. Wskazują one także na wysoką pozycję badaczy z USA (ponad 47% artykułów w bazie LISA na temat bibliotek cyfrowych) i z Wlk. Brytanii (19%). Na kolejnym miejscu są uczeni niemieccy (6,5%).

Alireza Isfandyari-Moghaddam i Behrooz Bayat badały piśmiennictwo dotyczące bibliotek cyfrowych z trzech punktów widzenia: oceny bibliotek cyfrowych, w tym określenie kryteriów oceny; zarządzania, w tym zagadnienia selekcji zasobów, ich gromadzenia, formatów, struktury i organizacji, udostępniania, bezpieczeństwa i archiwizacji; edukacji personelu i użytkowników [Isfandyari-Moghaddam, Bayat 2008, s. 856].

O globalnych systemach i infrastrukturach piszą także polscy autorzy. Mirosław Muraszkiewicz wskazuje na kolejne zmiany paradygmatów obowiązujących w informatyce na przestrzeni stosunkowo krótkiego okresu kilkudziesięciu lat: od podejścia sytuującego na głównym miejscu oprogramowanie, przez uwzględnienie roli danych, sieci lokalnych aż po rewolucję internetową, w wyniku której użytkownik postrzega sieć jako jeden, wielki komputer, wyposażony w ogromne zasoby obliczeniowe i informacyjne [Muraszkiewicz 2002; Gawrysiak 2008, s. 249] (por. idea Gridu, p. 1.3). Najbardziej przyszłościowe zastosowania informatyki wynikają z połączenia systemów informacyjnych, w tym głównie systemów baz danych, z sieciami transmisji danych.

Mirosław Górny pisze natomiast o globalnej infrastrukturze cyfrowej i miejscu bibliotek cyfrowych w tej strukturze. W związku z cyfryzacją zasobów przewiduje poważne zmiany w funkcjonowaniu bibliotek, które znacznie bardziej niż dotychczas nastawione będą na organizowanie dostępu do lokalnych zasobów cyfrowych. Będą jednak musiały jeszcze przez długi czas funkcjonować jako biblioteki hybrydowe, gromadząc zarówno zasoby cyfrowe (obce i tworzone lokalnie) oraz zasoby tradycyjne [Górny 2005].

- Dalszy rozwój bibliotek cyfrowych wymaga badań w takich obszarach, jak:
 - Rozpoznanie sposobów wykorzystania olbrzymich zasobów danych gromadzonych przez sensory urządzeń badawczych.
 - Dalszy rozwój skali, złożoności i różnorodności zasobów cyfrowych oraz algorytmów służących ich przetwarzaniu. Na przykład wyszukiwanie obiektów trójwymiarowych

(3D)²⁶ ma zasadnicze znaczenie w projektowaniu leków czy zastosowaniach inżynierskich (CAD). Wciąż brak sposobów na automatyczne, jednoczesne wyszukiwanie wielu mediów, chociaż w tzw. wyszukiwaniu uniwersalnym odnotowuje się postępy.

- Badania potrzeb i zachowań użytkowników, pozwalające na lepsze wykorzystanie bibliotek cyfrowych w nauce i edukacji. Systemy muszą zarówno przystosowywać się do potrzeb użytkowników, jak i najpełniej wykorzystywać uzyskiwaną informację.
- Zarządzanie zasobami i tworzenie modeli ekonomicznych, w sytuacji, gdy wszyscy chcą korzystać z dostępnych zasobów informacji, ale niewielu chce pokrywać koszty ich utrzymania.
- Rozwój współpracy interdyscyplinarnej i międzynarodowej.
- Wzrost poziomu inwestycji w biblioteki cyfrowe przez udostępnienie ich kolejnym typom użytkowników i zastosowaniom, od nauki do edukacji i zastosowań komercyjnych [Larsen, Wactlar 2003, s. 7].
- Rozwój możliwości wyszukiwawczych przez uwzględnianie „kontekstu” w trakcie wyszukiwania informacji;
- Sposoby włączenia zasobów informacyjnych w rozwiązywanie problemów dnia codziennego, w tym selekcja danych i przetwarzanie informacji, na podstawie których podejmowane są właściwe działania;
- Wzrost produktywności w wyniku efektywnego dostępu do informacji.

Pierwsza dekada badań nad bibliotekami cyfrowymi ukazała jasno, że możliwości tworzenia i gromadzenia danych znacznie przewyższają zdolność do ich właściwej organizacji, zarządzania i efektywnego wykorzystania. Badania te ukierunkowane były bezpośrednio na zagadnienie „przeładowania informacją”, gromadzoną w zasobach sieciowych. Wiele wskazuje na potrzebę uwzględnienia w dalszych badaniach centralnej roli użytkownika na wszystkich poziomach zastosowań danych, informacji i wiedzy. Pod uwagę brany jest też fakt, że zarówno osoby indywidualne, jak i organizacje, posiadają obecnie wielkie zasoby zapisów cyfrowych, a także to, że zasoby te są połączone poprzez sieć z innymi zasobami.

W trakcie realizacji programów badawczych bardzo wartościowe okazały się prototypy serwisów informacyjnych, szczególnie funkcjonujące w skali rzeczywistej. W kolejnych badaniach, związanych z bibliotekami cyfrowymi, praca na prototypach musi być uzupełniana nowymi działaniami w zakresie eksperymentów dotyczących modelowania i symulacji. Rozwiązania często dotyczą procesów informacyjnych – organizacji, zarządzania, rozpowszechniania i archiwizacji informacji. Poprawnie realizowane usługi przez biblioteki cyfrowe nazywane są transparentnymi, gdyż pozostają niemal niewidoczne dla użytkowników, co jednak nie oznacza, że nie istnieją. W rzeczywistości ich odpowiedni poziom osiągnąy jest dzięki długoterminowym pracom badawczym.

²⁶ Nie chodzi tu o wyszukiwanie grafiki za pomocą słów kluczowych (jak to ma miejsce w Google), ale na przykład przez narysowanie obiektu podobnego, wybór z zestawu kategorii obiektów, czy nawet przez porównywanie elementów obiektów graficznych (np. twarzy).

1. KOMUNIKACJA NAUKOWA W SIECI

W tym rozdziale przedstawię rozwój komunikacji naukowej od czasu wynalezienia druku przez Gutenberga w XV wieku, po najnowsze tendencje w organizacji nauki i komunikacji naukowej, takie jak informatyczna infrastruktura nauki i – będąca jej częścią – globalna biblioteka cyfrowa. Zamierzam opisać tendencje rozwojowe poprzez analogie do wcześniejszych wydarzeń, które mogą ułatwić zrozumienie współcześnie występujących zjawisk w zakresie komunikacji naukowej. Dzięki temu możliwe będzie nie tylko łatwiejsze zrozumienie aktualnych procesów, ale także prognozowanie przyszłości, przynajmniej tej najbliższej.

1.1. Rozwój komunikacji naukowej

Zmiany spowodowane technologiami informacyjnymi w komunikacji naukowej na przełomie XX i XXI wieku, w szczególności wynikające z cyfryzacji i rozpowszechnienia sieci komputerowych, często określane są mianem rewolucyjnych. Porównuje się je do efektów, wywołanych wynalezieniem prasy drukarskiej w XV w., co zapoczątkowało tzw. rewolucję Gutenberga. Według wielu autorów komunikacja naukowa znajduje się obecnie w okresie przemian porównywalnych do tych, które zachodziły w czasach zaraz po wynalezieniu druku i w następujących wiekach [Landow 2006, s. 49-51]. Wynalazek ten nie był oceniany jednoznacznie: książka drukowana, wychodząca w stosunkowo dużej liczbie egzemplarzy, nie cieszyła się tym nimbem rzadkości i unikalności, co rękopisy, a jej stosunkowo niska cena tym bardziej obniżała jej wartość w oczach zbieraczy i czytelników. Podobnie dziś traktowane są media elektroniczne – przekształcające świat w globalną wioskę, a jednocześnie obwiniane o szerzenie pornografii, przemocy i rzeczy bezwartościowych, co nie przeszkadza im rozwijać się w tempie zasługującym na miano „rewolucyjnego”, co opisywałem w jednej ze swoich poprzednich publikacji [Nahotko 2007a].

Próbę wyjaśnienia procesów zmian technologicznych podjął Richard Foster [Foster 1986]. Jego teoria, znana jako „technologiczne krzywe S” opisuje sposoby dokonywania się zmian technologicznych poprzez zastępowanie starej technologii nową. Zgodnie z tą teorią, nowa technologia początkowo z trudem dorównuje starej, ustabilizowanej technologii. Jednak po okresie powolnego rozwoju następuje jej gwałtowny wzrost, w wyniku którego nowa technologia całkowicie dominuje na rynku, aż do momentu powstania i rozwoju kolejnej technologii. Oznacza to, że istnieją okresy współlistnienia technologii, w trakcie których następuje przesilenie i zanik starej technologii. Okres przesilenia jest szczególnie trudny dla użytkowników, gdyż z jednej strony istnieje duże ryzyko w przypadku zbyt szybkiego porzucenia starej technologii, ale także szansa dużych korzyści z wczesnego stosowania nowej.

Biorąc pod uwagę teorię Fostera, oczekiwanie, że dziś nastąpi gwałtowne porzucenie dawnych sposobów komunikacji i natychmiastowa transformacja komunikacji naukowej przez technologie informacyjne w kierunku cyfryzacji, może być złudne. Obecnie również zachodzi konieczność dostosowania formy wypowiedzi do możliwości, jakie oferują nowe technologie [Górska 2009, s. 138]. Zaskakująca jest siła przywiązania do starych

rozwiązań, w środowisku elektronicznym nie tylko nie mających już sensu, ale wręcz przeszkadzających. Jak twierdzi Terrence Brooks, nie ma żadnych technicznych powodów, dla których strony Web miałyby wyglądać jak dokumenty drukowane, oprócz chęci zachowania konwencji kulturowych, do których użytkownicy tych stron są przyzwyczajeni [Brooks 2003].

Dla oceny roli i przyszłych kierunków już rozpoczętych zmian, właściwe wydaje się ich odniesienie do rozwoju komunikacji naukowej w perspektywie historycznej. Z pewnością pierwotną formą komunikacji jest sposób „z ust do ust”, gdzie informacja jest przekazywana bezpośrednio i synchronicznie jednemu lub kilku odbiorcom, w komunikacji naukowej przez mistrza jego uczniom. Uczniowie następnie rozpraszają się po świecie, sami zostają mistrzami znajdując swoich uczniów, dzięki czemu rozprzestrzeniają dalej zdobytą wiedzę, często wzbogacaną o własny dorobek. Taki sposób komunikacji naukowej stosowany jest od czasów Sokratesa po dzień dzisiejszy tam, gdzie funkcjonują lokalne „szkoły”, oparte na koncepcjach rozwijanych i przekazywanych przez mistrza-profesora [Landow 2006, s. 47].

Szybko jednak okazało się, że sprawna komunikacja naukowa wymaga bardziej efektywnego sposobu przekazywania informacji w czasie i przestrzeni. Bardzo wcześnie, bo w VII w. p.n.e., powstały pierwsze naukowe przekazy piśmienne, dokumentujące postęp nauki w ówczesnym, cywilizowanym świecie [Vickery 1997, s. 1-4]. Pomimo wszystkich negatywnych cech, dostrzeganych już przez starożytnych¹, pismo miało w sobie wystarczający potencjał, aby stałe się rozpowszechniać. Utrudnia wprawdzie interakcję, ale utrwała przekaz, a także, ze względu na swój abstrakcyjny charakter, jest forum argumentacji i kalkulacji logicznej [Delany, Landow 2008, s. 79]. Utrwalanie dzieł w piśmie, zarówno literackich, jak i filozoficznych, stawało się normą².

Początkowo rola rękopiśmiennych dokumentów naukowych oraz ich drukowanych następców ograniczała się do scalania i przekazywania istniejącej, autorytarnej wiedzy, gdyż tak wówczas pojmowana była rola uniwersytetów. Komunikacja naukowa w takiej formie, jak się ją rozumie obecnie, funkcjonowała na zasadzie indywidualnej wymiany informacji pomiędzy pracownikami nauki, opierając się na korespondencji lub podróżach (czyli kontaktach bezpośrednich). Aż do połowy XVII wieku drukarnie nie uczestniczyły w procesie modernizacji komunikacji naukowej, polegającej na wymianie idei, pomysłów i efektów badań empirycznych, jak również poddawaniu ich debacie publicznej za pomocą periodycznej, prenumerowanej publikacji. Kiedy publikacje takie (pierwsze czasopiśma naukowe) zaczęły powstawać, ich twórcami były nie uniwersytety, lecz towarzystwa naukowe.

Dla zrozumienia natury tych procesów należy przyjrzeć się początkom nowoczesnej

¹ Ryszard Tadeusiewicz pisze o reakcji celtyckich druidów, przechowujących tradycję w formie ustnych przekazów, na próby wprowadzenia pisma. Uważali oni, że pismo ogłupi ludzi, ponieważ przestaną kształcić swoją pamięć. To zaś spowoduje cofnięcie kultury i cywilizacji do czasów barbarzyństwa [Tadeusiewicz 2006, s. 39]

² Wydaje się, że można zaryzykować twierdzenie o zbliżeniu form sieciowej komunikacji elektronicznej do modelu komunikacji poprzez dialog, preferowanego przez starożytnych filozofów. Co prawda głównym środkiem przekazywania informacji pozostaje pismo, jednak znaczne przyspieszenie cyklu obiegu informacji i znaczna interaktywność nowych mediów pozwalają tak bardzo przyspieszyć proces komunikacji, że w porównaniu z drukiem wydaje się on być dialogiem. Jest on podobnie jak dialog bardziej ulotny: treści w Web zmieniają się stale i stosunkowo szybko. Zgodnie z informacjami podanymi przez Terrence Brooksa połowa stron Web w domenie edu znika po czterech miesiącach, a połowa URL po około dwóch latach; po dwóch latach zmienia się też połowa treści zawartych w Web [Brooks 2003]. W efekcie strony Web traktowane są jak skrózyszy (dokumenty zintegrowane), w które w razie potrzeby wkłada się i wyjmuje dowolne części [Zawado 2002, s.88], modyfikując na bieżąco tok „dialogu”. Przez to zawartość Webu znajduje się gdzieś pomiędzy dwiema tradycyjnymi strategiami przechowywania i rozpowszechniania informacji: efemeryczną, niezapisywaną oraz utrwalaną w sposób bardziej stabilny za pomocą pisma [Koehler 1999].

nauki, wywodzącej się z okresu tzw. rewolucji naukowej³, rozpoczętej w XVI w. i kontynuowanej w kolejnym stuleciu. Nowoczesna nauka rozwinęła się bowiem w trakcie XVII-wiecznych ruchów oświeceniowych [Owen 2007, s. 29], zwiastowanych przez dzieło Kopernika, a później łączonych z takimi nazwiskami, jak Galileusz i Newton.

Praktyka naukowa przed rewolucją naukową opierała się na gromadzeniu, analizowaniu i scalaniu otrzymanej, wcześniej istniejącej wiedzy, w tym w dużej mierze tłumaczeniu na łacinę tekstów greckich, hebrajskich i arabskich. Początkową rolą zarówno rękopisów naukowych i następujących po nich od XV wieku publikacji drukowanych było zachowanie i przekazywanie tego typu wiedzy. Rewolucja naukowa polegała między innymi na przejściu od publikowania kompendiów wszelakiego rodzaju wiedzy, typowych dla Renesansu, do praktyki opartej na badaniu, odkrywaniu i tworzeniu nowej wiedzy. Loet Leydesdorff opisał tę zmianę jako przejście od normatywnego systemu wierzeń do systemu opartego na racjonalnych przesłankach, w którym „możliwe jest dochodzenie prawdy, przez co jej poszukiwanie może funkcjonować jako kod stanowiący podstawę komunikacji” [Leydesdorff 2001]. W efekcie nauka ze społecznego punktu widzenia została zbudowana jako dyskursywny system racjonalnych założeń. Oświeceniowy program badawczy kładł główny nacisk na poznanie natury i sposobu funkcjonowania wszechświata. Podstawą poznania i rozumienia świata materialnego miała być metoda empiryczna, doskonale spełniająca swoją rolę w naukach ścisłych. Nie podlega dyskusji, że opisane zmiany miały olbrzymie znaczenie dla funkcjonowania ówczesnego społeczeństwa, nie tylko nauki. W takim sensie nazywanie ich „rewolucyjnymi” jest usprawiedliwione. Nie oznacza to jednak, że był to proces szybki czy gwałtowny, jak przywykliśmy myśleć o rewolucjach. Trwał on prawie dwieście lat, rozpoczynając się odkryciami Mikołaja Kopernika (1473-1543) w XVI w.⁴, a swoją kulminację osiągnął on dzięki pracom Isaaca Newtona (1643-1727) i innych, współczesnych mu uczonych.

Od XVII wieku przyjęło się myśleć o nauce jako o „otwartym” systemie komunikacyjnym, opartym na możliwie szerokiej wymianie idei, hipotez i osiągnięć, pozwalającym na ich nieograniczoną analizę, krytykę i debatę. Pociąga to za sobą potrzebę wytworzenia mechanizmu zapewniającego swobodne tworzenie i wymianę idei i powszechną dostępność wyników badań naukowych bez względu na miejsce i czas. Za taki mechanizm trzeba uważać zorganizowany system komunikacji naukowej, realizujący określone funkcje (np. dystrybucję, zapewnienie dostępu, archiwizację) oraz działający jako system społeczny w takim sensie, że ktoś, kto jest wykluczony z systemu komunikacji, nie może funkcjonować jako członek społeczności naukowej.

Tworzenie takiego systemu rozpoczęło się jeszcze przed początkiem rewolucji naukowej, wraz z powstawaniem pierwszych uniwersytetów. Stanowiły one element systemu komunikacji posiadający wiele charakterystycznych cech współczesnego systemu nauki. Uniwersytety funkcjonowały jako miejsca wymiany informacji, zarówno istniejącej w formie słownej, jak i piśmiennej. Brały one także udział w społecznym umiejscowieniu nauki dzięki częstym wymianom personalnym pomiędzy uczelniami a społeczeństwami, w których one działały. W odróżnieniu od wcześniejszych, klasztornych ośrodków nauki, cechowały się one interdyscyplinarnością, charakterystyczną także dla czasów rewolucji naukowej; została ona zniszczona dopiero przez XIX i XX-wieczną specjalizację w badaniach naukowych.

Technologia druku jest jednym z pierwszych przykładów technologii informacyjnej i komunikacyjnej. Z pewnością miała ona wieloraki wpływ na rozwój nauki. Przed wynalezieniem ruchomej czcionki, naukowe księgi rękopiśmienne, przeznaczone do roz-

³ O rewolucyjności zmian w nauce w XVII w. dość szybko zdawano sobie sprawę: jako pierwszy mówił o nich w ten sposób Jędrzej Śniadecki w 1781 r. [Ratajewski 1993, s. 91].

⁴ *De revolutionibus orbium coelestium* jak wiadomo opublikowane zostało w roku śmierci astronoma, w Norymberdze, powodując przewrót w nauce i ówczesnym światopoglądzie.

powszechniania wiedzy, były tworzone przez specjalnych pracowników na podstawie egzemplarzy zatwierdzonych przez uniwersytet. Po wynalezieniu druku ta praktyka nie uległa od razu zmianie [Vickery 2000, s. 60]. Na początku druk funkcjonował jako rodzaj mechanizacji rękopiśmiennictwa. Z jednej więc strony ilościowo produkcja gwałtownie wzrosła, z drugiej zaś zasady i funkcja komunikacyjna dzieł drukowanych nie zmieniła się bardzo w stosunku do manuskryptów.

Wpływ druku na środowisko akademickie był początkowo bardziej ilościowy niż jakościowy, powodując raczej konserwację stanu istniejącego niż innowacyjność⁵. W takim sensie mówienie o „rewolucji Gutenberga” wydaje się nie do końca usprawiedliwione, przynajmniej w odniesieniu do komunikacji naukowej. Oprócz stworzenia szerszego grona odbiorców efektów działalności naukowej, druk nie spowodował natychmiastowego powstania nowego typu działalności. Takie zmiany wymagały znacznie więcej czasu. Nie można oczywiście zaprzeczyć, że w dłuższej perspektywie druk (wraz z innymi czynnikami, na przykład odejściem od łaciny na rzecz języków narodowych) doprowadził do poważnych zmian, zapoczątkowujących rewolucję naukową. Brak natychmiastowego efektu innowacyjności i „rewolucyjności” druku tłumaczony jest przez Johna Mackenzie Owena tym, że ważną funkcją druku nie było ułatwianie rozpowszechniania wyników badań, lecz organizacja dostępu do źródeł na wczesnym etapie pracy badawczej (głównie gromadzenie literatury) [Owen 2007, s. 33]. Biorąc pod uwagę czas niezbędny do stworzenia poważnego zasobu danych w formie drukowanej, łatwiej jest zrozumieć, że „rewolucja” druku w nauce była stosunkowo powolnym procesem.

Sytuacja w zakresie współczesnej digitalizacji jest w pewnym zakresie podobna do druku w opisanym wczesnym stadium rozwoju tej technologii. Cyfryzacja i digitalizacja powodują wzrost liczby publikacji i tempa rozpowszechniania rezultatów badań, a także dają zwiększone możliwości interakcji (podobnie jak druk w początkowym okresie). Następuje masowe przenoszenie druków do nowej postaci (digitalizacja), które może pomóc znaleźć nowych czytelników wielu zapomnianym treściom [Górska 2009, s. 134, 136]. Jednak wpływ digitalizacji na działalność naukową może w większym stopniu zależeć od sposobu, w jaki zasoby danych są dostępne w trakcie realizacji badań. Być może wpływ cyfryzacji i digitalizacji na dostępność danych jako źródeł do badań staje się ważniejszy niż jej oddziaływanie na sposób rozpowszechniania wyników badań (publikowania). Może to mieć szczególne znaczenie w naukach humanistycznych, gdzie digitalizacja źródeł informacji jest kosztowna i czasochłonna, przez co, przynajmniej w zakresie tych dwóch kryteriów, nie może konkurować z drukiem offsetowym.

Można wskazać także inne analogie do czasów Gutenberga. Po rozpowszechnieniu się wynalazku ruchomych czcionek, drukarze przede wszystkim rozpoczęli „przedrukowywanie” najważniejszych rękopisów, z Biblią na czele; jednocześnie jednak masowo powstawać zaczęły druki o bardzo niskiej jakości, jak literatura dewocyjna i procratwa⁶. W początkowym okresie książki drukowane naśladowały manuskryty krojem czcionki, układem strony i innymi cechami zewnętrznymi. Szybko jednak zaczęły powstawać nowe dzieła, dostępne wyłącznie w druku i dla niego specjalnie przeznaczone. Teksty te uwzględniały nowe możliwości, ale też wymagania nowej technologii, w której miały być rozpowszechniane. Gdy książki kopiowano w klasztorach w pojedynczych egzemplarzach, nie było możliwości (oraz potrzeby) stosowania wspólnych standardów. Po upowszechnieniu się druku używanie wspólnych norm stało się możliwe i przydatne. Większość z nich wyznaczana była zresztą przez technologię. W książkach mogło pojawić się

⁵ Innowacyjność polega na rozpoznaniu przydatności do nowych celów posiadanej już wiedzy lub informacji. Typowo odbywa się ona w trakcie rozpowszechniania się nowych sposobów wykonywania jakiejś działalności i wiedzy o tych sposobach, czyli dyfuzji innowacji. Dyfuzja jest rodzajem komunikacji [Rogers 2003, s. 5].

⁶ Więcej zjawisk upodabniających zmiany wynikające z wynalazku druku i publikacji cyfrowych znaleźć można w publikacji Małgorzaty Górskiej [Górska 2007].

tylko to, co można było wydrukować na sprzęcie, który był do dyspozycji. Technologia decydowała na przykład o sposobach stosowania ilustracji (częściej o ich braku), kroju czcionki, doborze proporcji elementów tekstu. Przez pewien czas obie technologie: rękopiśmienna i druk funkcjonowały równolegle, po czym zawód skryby zaginął. Nie był to jednak proces natychmiastowy: w XVII w. doszło, z różnych przyczyn społecznych, wręcz do porzucenia druku na rzecz rękopisów, a ostatnie scriptoria w klasztorach funkcjonowały do końca XVIII w. [Głombiowski, Szwejkowska 1971, s. 160].

Pomimo swojej innowacyjności, książka drukowana nie mogła jednakże zaspokoić innych potrzeb charakterystycznych dla nowej nauki, takich jak wzrost tempa rozwoju, mnogość form działalności oraz międzynarodowy charakter. Potrzeba nowej formy publikacji zaznaczyła się na początku XVII wieku, najpierw w formie rozwoju komunikacji naukowej opartej na sieciach współpracy personalnej, tworzonych w wyniku coraz częstszych podróży uczonych do ośrodków badawczych i bezpośrednich spotkań z kolegami, rozwoju czytelnictwa oraz wymiany listów, co umożliwiły nowo powstające usługi pocztowe.

Listy stały się ważnym medium komunikacyjnym w społecznościach związanych z rewolucją naukową z dwóch powodów. Po pierwsze były prekursorem naukowego artykułu w czasopiśmie: role artykułów, publikowanych w pierwszych periodykach naukowych pełniły listy, przygotowane do druku przez redaktora. Po drugie, były one kopiowane w wielu egzemplarzach i rozpowszechniane między uczonymi i zainteresowanymi amatorami, co nieco przypomina dzisiejsze listy dystrybucyjne poczty elektronicznej. Niektórzy uczeni specjalizowali się w kompilacji tych pism naukowych, wykonując zadania „zbieraczy wiedzy”.

Komunikacja bezpośrednia była następnie instytucjonalizowana w formie organizowanych towarzystw naukowych, takich jak słynne The Royal Society w Londynie. Poprzez towarzystwa naukowe rozpowszechnianie listów o treści naukowej i technicznej zostało zinstytucjonalizowane w formie oficjalnych zestawów informacji (biuletynów), wydawanych przez towarzystwa i rosyłanych do ich członków.

The Royal Society of London powstało w 1660 r., utworzone przez grupę uczonych spotykających się już wcześniej regularnie dla dysput filozoficznych. We Francji kardynał Richelieu utworzył w 1635 r. Académie Française, która zajmowała się poprawnością języka francuskiego. Paryska Académie des Sciences, formalnie założona w 1666 r., powstała jako kontynuacja stałych spotkań uczonych w bibliotece królewskiej [Ratajewski 1993, s. 89]. Był to tylko początek ruchu służącego tworzeniu instytucji, zwykle prywatnych, często narodowych, skupiających uczonych i ludzi oświeconych; powstały one niemal we wszystkich krajach europejskich⁷, również w Polsce.

Zmiany zachodzące w europejskiej nauce w XVI i XVII wieku: stosowanie nowych metod badawczych, rewolucja naukowa, utworzenie sieci komunikacji naukowej i zakładanie towarzystw naukowych, tworzyły warunki do powstania nowego nośnika informacji i komunikacji, jakim jest czasopismo naukowe. Na początku roku 1665 r. ukazał się w Paryżu pierwszy numer czasopisma „Journal des Sçavans”, opublikowany przez Denisa de Sallo (1626-1669). Pismo zawierało głównie recenzje książek opublikowanych w Europie. Poza tym publikowano w nim dyskusje naukowe i raporty z badań oraz z od-

⁷ Kolejne towarzystwa i akademie nauk powstawały we Włoszech (1683), Rosji (1725), Szwecji (1739), Szkocji (1783), Irlandii (1785), Holandii (1808). Na ziemiach polskich w 1743 r. powstało w Gdańsku Towarzystwo Przyrodnicze; w l. 1800-1831 działało Towarzystwo Warszawskie Przyjaciół Nauk w Warszawie, którego działalność kontynuuje od 1907 r. Towarzystwo Naukowe Warszawskie; od 1815 r. działało Towarzystwo Naukowe Krakowskie, z którego w 1872 r. powstała Akademia Umiejętności; od 1820 r. istnieje Towarzystwo Naukowe Płockie; od 1827 r. działało Radomskie Towarzystwo Naukowe; od 1851 r. działa Towarzystwo Biblioteki Słuchaczy Prawa Uniwersytetu Jagiellońskiego; od 1857 r. funkcjonuje Poznańskie Towarzystwo Przyjaciół Nauk; 1875 r. to data powstania Towarzystwa Naukowego w Toruniu.

bytych konferencji. W tym samym, 1665 r., pochodzący z Bremy Henry Oldenburg (ok. 1619-1677), sekretarz The Royal Society of London utworzył czasopismo „Philosophical Transactions of the Royal Society of London”, które później posłużyło Isaacowi Newtonowi do ogłaszania swoich odkryć. Ostatnie dziesięciolecie XVII w. przyniosły pewną liczbę podobnych przedsięwzięć⁸, realizowanych zarówno przez towarzystwa naukowe (na podobieństwo „Philosophical Transactions”) jak i przez osoby prywatne (wzorowane na „Journal des Sçavans”). Jak pisze John Mackenzie Owen, tych zmian nie można jednak nazywać rewolucyjnymi; dotyczyły one nieco ponad 20 tytułów, zwykle efemeryd szybko kończących swój żywot [Owen 2007, s. 33]. Taka sytuacja trwała do drugiej połowy XVIII w., gdy w latach 1750-1790 powstały 422 nowe tytuły czasopism. Wzrost ten stał się ówczesnie powodem narzekania na zalew literatury naukowej, nie zawsze najwyższej jakości, przypominających współczesne wypowiedzi [Linde 2006, s. 16].

Nawet po stu latach od opublikowania pierwszego zeszytu „Philosophical Transactions” czasopisma naukowe nie osiągnęły poziomu rozwoju wystarczającego do pełnego zaspokojenia potrzeb uczonych europejskich w zakresie komunikacji naukowej. Czas oczekiwania na publikację był bardzo długi, często ponad roczny, szczególnie w redakcjach czasopism akademickich, które zapewniały kontrolę jakości. Kłopotliwe było utrzymywanie nadzwyczaj szerokiego zakresu tematycznego tych czasopism. Rozwiązaniem było tworzenie czasopism specjalistycznych, o coraz węższym zakresie, zawierających materiały różnego rodzaju, pochodzące z wielu źródeł i szybciej publikowane. Pierwsze specjalistyczne czasopisma powstawały dla fizyki, chemii i botaniki: zawierały artykuły, korespondencje, tłumaczone wyjątki z innych źródeł, recenzje książek.

Odzwierciedleniem rozwoju nauki i jej umocnienia instytucjonalnego w uniwersytetach w XIX w. był wyraźny wzrost liczby publikacji, zarówno w sensie liczby czasopism, jak i publikowanych w nich artykułów. Oblicza się, że w tym stuleciu opublikowano około 2 mln artykułów naukowych [Vickery 2000, s. 123]. Musiało to prowadzić do powstania dobrze zorganizowanego systemu rozpowszechniania informacji naukowej, składającego się z instytucji naukowych, wydawców komercyjnych i bibliotek naukowych. Powstało także zapotrzebowanie na wyspecjalizowane usługi informacyjne, takie jak bibliografie, roczne przeglądy, czasopisma przeglądowe, a w końcu czasopisma abstraktowe i indeksowe⁹ (forma „czasopiśmiennienia” bibliografii). W połowie XIX w. powszechne stało się umieszczanie bezpośrednich odsyłaczy bibliograficznych do wcześniej opublikowanych materiałów. Przez cały XIX w. monografia naukowa pozostała jednak głównym sposobem rozpowszechniania nowych teorii naukowych [Evans, James 2008, s. 399], ustępując ostatecznie miejsca artykułom naukowym.

Oświeceniowe podejście do nauki utrzymało się praktycznie aż do XX wieku. Wówczas pojawiły się pierwsze oznaki dezintegracji tego stabilnego, choć niedoskonałego systemu. Pierwszym powodem rozpadu wielowiekowych stosunków był nadzwyczaj gwałtowny wzrost ilości prowadzonych badań naukowych, który nastąpił w trakcie i zaraz po II wojnie światowej¹⁰. Wiedzy naukowej przybywało, dokonywano nowych, ważnych odkryć, które umożliwiały tworzenie nowych wynalazków technicznych. Spowodowało to również gwałtowny wzrost liczby przygotowywanych publikacji, które nie mogły

⁸ Na przykład w 1668 r. w Rzymie powstało czasopismo „Giornale de'Letterati d'Italia”, a w Niemczech (w Lipsku) przez sto lat, od 1682 r. wydawano „Acta Eruditorum”. W Polsce wydawane były: warszawski „Pamiętnik Historyczno-Polityczny” (1782-92) i krakowski „Zbiór Tygodniowy Wiadomości Uczonych” (1784-85).

⁹ Pierwsze abstrakty pojawiły się jednak dużo wcześniej, znane były nawet w średniowieczu.

¹⁰ Okres ten, oprócz nadzwyczajnych cierpień, śmierci i zniszczeń był także czasem nadzwyczajnego rozwoju nauki i techniki. Z punktu widzenia nauki Warren Weaver wskazywał na dwa zasadnicze osiągnięcia: rozwój nowych maszyn cyfrowych i powstanie interdyscyplinarnych zespołów badawczych [Weaver 1948]. Jak to wykazuje w dalszej części książki oba te osiągnięcia są dla nauki bardzo istotne aż do dzisiaj.

być „konsumowane” przez wydawnictwa uczelniane i towarzystw naukowych. Uczni zwrócili się więc ku wydawnictwom komercyjnym, które wcześniej nie interesowały się publikowaniem artykułów naukowych, traktowanym jako przedsięwzięcie niedochodowe [Thorin 2003]. Publikowanie naukowe stało się przemysłem ważnym z komercyjnego punktu widzenia oraz działalnością angażującą wielu ludzi funkcjonujących w różnych formach organizacyjnych.

W XX wieku obserwowaliśmy powszechnie znany olbrzymi rozwój ilościowy publikacji naukowych (zob. p. 1.2), wzrost specjalizacji prowadzący do tworzenia bardzo wyspecjalizowanych czasopism, zajęcie przez artykuł w czasopiśmie pozycji głównego nośnika informacji w komunikacji naukowej, rozpowszechnienie się języka angielskiego jako uniwersalnego języka nauki i techniki oraz umocnienie się (szczególnie w końcu ubiegłego wieku) na rynku niedużej liczby¹¹ wielkich przedsiębiorstw komercyjnych – międzynarodowych wydawnictw, powstałych w wyniku przejścia mniejszych firm, dbających na równi o własne zyski finansowe, co realizację potrzeb nauki.

- Innymi cechami charakterystycznymi komunikacji naukowej na początku XXI w. są:
- publikowanie przez wielu współautorów: badania wskazują na dwukrotny wzrost średniej liczby autorów w drugiej połowie XX wieku (obecnie 4,5 autorów na artykuł), przy jednoczesnym spadku liczby artykułów autorstwa każdego autora do 0,7 artykułów rocznie [Ware, Mabe 2009, s. 22];
 - zmniejszanie się stosunku artykułów czytanych przez przeciętnego pracownika nauki do ogólnej liczby artykułów publikowanych;
 - zmiana sposobu czytania artykułów: od „intensywnego” funkcjonującego do około 1750 r. (uczni czytali wszystko wielokrotnie), przez „ekstensywny” ok. 1800 r. (jednokrotnie czytana była cała, wzrastająca produkcja), po „przeglądanie” lub „skanowanie” obecnie;
 - ograniczenie prenumeraty nieinstytucjonalnej: ze względu na dużą liczbę publikowanych artykułów pracownicy nauki czytają tylko niektóre z nich, więc zamiast prenumerować całe wydawnictwo wolą kserować tylko to, co ich naprawdę interesuje¹²;
 - do niedawna wysoka liczba cytowań starszych publikacji (ponad 15-letnich), co może mieć wpływ na cytowania publikacji elektronicznych (młodszych) [Liu 2003, s. 895]; w ostatnich latach tendencja ta ulega zmianie, w efekcie czego obserwowane jest zjawisko tzw. *tunnel vision*¹³;
 - ograniczenia w dostępie do publikacji w bibliotekach naukowych, spowodowane przejściem od prenumeraty do licencjonowanego dostępu [Feather 2003, s. 87].

Ważnym elementem systemu komunikacji naukowej w XX wieku była jego rola w ocenie jakości i kontroli reputacji, co znalazło swój wyraz w tworzeniu indeksów cytowań, stosowaniu IF¹⁴ jako jednego z narzędzi oceny publikacji i jej autora oraz tworzenie formalnych rankingów czasopism. Równoległe utrzymywany jest system recenzowania, odróżniający publikacje naukowe od innych, w swoich założeniach gotowy już w XVII wieku, w postaci „ocenia redaktor + dwóch anonimowych recenzentów”.

Za Gustavo Cardoso można stwierdzić, że istnieją trzy fazy przekazywania treści, które wciąż koezystują ze sobą, gdyż kolejna faza nie eliminuje w zupełności poprzedniej [Car-

¹¹ Ponad połowa anglojęzycznego czasopiśmiennictwa naukowego z zakresu STM pochodzi od trzech wydawców: Elsevier (28,2% dochodów z rynku), Kluwer i Springer (14%), Thomson (9,5%) [Madras 2008, s. 163]. Tylko około 1/3 rynku posiadają małe i bardzo małe wydawnictwa w liczbie kilku tysięcy. Dziesięciu największych wydawców publikuje 35% tytułów czasopism.

¹² Być może stąd wynika popularność modelu udostępniania czasopism elektronicznych typu „pay per view”.

¹³ Zjawisko to polega na cytowaniu tylko najnowszych publikacji, z pominięciem publikacji starszych, niż kilkuletnie, co może wynikać z łatwiejszego dostępu do źródeł dzięki ich cyfryzacji [Evans, James 2008, s. 396].

¹⁴ IF – Impact Factor (wskaźnik oddziaływania) to stosunek liczby cytowań artykułów z danego czasopisma do liczby wszystkich artykułów opublikowanych w tym czasopiśmie w okresie dwóch lat.

doso 2008, s. 594]. Jest to symboliczna wymiana realizowana oralnie (wykłady i dysputy naukowe a także kazania, przemówienia okolicznościowe), pisemnie (tu można wyróżnić fazę manuskryptów i druku) i elektronicznie. Na każdym z tych etapów relacje pomiędzy językiem i społeczeństwem, ideami i działaniami, a także poczucie własnej odrębności (na przykład rozumianej jako bycie autorem) są różne. Tak jak w XIX wieku druk odgrywał zasadniczą rolę w formowaniu sfery niezależnej i racjonalnej debaty publicznej, co według Jürgena Habermasa stworzyło podstawy pod XX-wieczne demokracje [Habermas 1999], tak samo nowe media, w szczególności Internet, dzięki swoim cechom odgrywają istotną rolę w zróżnicowanym, zdecentralizowanym i rozproszonym rozpowszechnianiu informacji¹⁵.

Komunikacja naukowa od czasu wyodrębnienia się jej nowoczesnych form spełnia kilka zasadniczych funkcji, sprowadzających się do walidacji wyników badań, pozyskiwania współczesnych i przyszłych odbiorców, oraz tworzenia i obsługi zasobu stanowiącego opis stanu wiedzy naukowej. Funkcje te, nazywane w różny sposób, pozwalają na zachowanie ciągłości procesów komunikacyjnych niezależnie od zmian technologicznych. Umożliwiają również zrozumienie zasad podejmowania decyzji przez uczonych co do przedmiotu badań, kręgu współpracowników, miejsca realizowania prac oraz publikowania osiągnięć. Według Christine Borgman wyróżnić można trzy grupy funkcji komunikacji naukowej:

- **Legitymizacja.** Jest to podstawowa funkcja komunikacji społecznej, zapewniająca kontrolę jakości. Innym zadaniem legitymizacji jest ustalenie pierwszeństwa; temu właśnie służą informacje o dacie wplynięcia i przyjęcia tekstu podawane przez wiele redakcji. W wyniku tych procesów zapewniana jest wiarygodność zarówno publikacji, jak i jej autora, poprzez określenie reputacji zarówno jego, jak i zatrudniającej go instytucji [Arms 2002]. Jeżeli uznamy, że komunikacja naukowa odbywa się w wielu formach między w pełni nieformalną do typowo formalnej, to legitymizacja pojawi się w pewnym stałym miejscu tego kontinuum. To miejsce stanowi punkt równowagi, na którym osadzone jest publikowanie naukowe. Wskazanie i określenie zasad powstawania tego punktu jest podstawą tworzenia modeli publikowania naukowego w środowisku cyfrowym.
- **Rozpowszechnianie** opisywane jest takimi terminami, jak komunikacja i rozprzestrzenianie, poinformowanie, przejrzystość czy dyskurs, czyli według powszechnej terminologii, publikowanie. Rozprzestrzenianie i publikowanie związane jest z poinformowaniem innych członków społeczności uczonych. Przejrzystość oznacza odpowiedzialność uczonych za umieszczenie swoich prac w sferze publicznej, zarówno w celu legitymizacji, jak i umożliwienia dostępu do nich. Dyskurs podkreśla rolę publikacji w formalnej komunikacji w społeczności uczonych. Zasadniczą rolę w rozpowszechnianiu odgrywali wydawcy. Obecnie ich rola maleje na rzecz samych uczonych.
- **Dostęp, archiwizacja i obsługa.** Zasób dokumentów naukowych ma służyć obecnym i przyszłym uczonym, studentom i innym użytkownikom. Metody udostępniania, archiwizacji i obsługi tego zasobu zmieniają się, co powoduje zmiany w ich opisie. Wszystkie trzy procesy mogą być postrzegane jako jedna złożona funkcja komunikacji naukowej, służąca zapewnieniu stałej dostępności zasobu dokumentów naukowych. Funkcja ta jest typowa dla działalności bibliotek od początku ich istnienia. W środowisku cyfrowym część aktywności związanej z jej realizacją przejmują poszczególni uczeni, zespoły badawcze, czasopisma i społeczności powiązane z poszczególnymi dyscyplinami [Borgman 2007, s. 66-68].

W ostatniej dekadzie XX w. czasopisma naukowe weszły w nową fazę ewolucji zwią-

¹⁵ Według amerykańskiego medioznawcy, Philipa Meyera jest to już dobrze widoczne na przykładzie prasy codziennej, której najlepszy rozwój przypadł na lata 1890-1920; kolejno radio i telewizja zabierały jej część rynku wiadomości, ale nie naruszyły podstaw modelu biznesowego. Dopiero rozpowszechnienie Internetu zadalo prasie decydujący cios, doprowadzając większość dzienników na próg bankructwa [Meyer 2004, s. 32].

zanej z udostępnianiem informacji w globalnej sieci komputerowej. Cyfryzacja publikacji naukowych, głównie czasopism, przy jednoczesnym braku drożności systemu komunikacji naukowej, opartej na tradycyjnym modelu publikacji naukowych, spowodowała wzrost zainteresowania w świecie nauki rozpowszechnianiem wiedzy, jako jednej z podstawowych funkcji nauki. Większość tego zainteresowania skupiona jest na czasopismach recenzowanych, które migrują do postaci cyfrowej. Migracja jest na ukończeniu, powodując głównie wzrost liczby czasopism hybrydowych¹⁶. Dalszy rozwój publikowania elektronicznego wiązany jest z ruchem Open Access, szczególnie w jego nowych formach¹⁷.

1.2. Współczesna nauka i komunikacja naukowa

Zmianom w sposobach organizacji komunikacji naukowej towarzyszą przemiany w samej nauce. Obejmują one różne procesy, począwszy od zmiany paradygmatów i metodologii naukowej, a kończąc na zagadnieniach organizacyjnych i politycznych. David Kellogg opisuje te przemiany jako odejście w okresie powojennym od idealizmu Roberta Mertona, przejawiającego się w jego czterech zasadach prowadzenia badań naukowych¹⁸ [Kellogg 2006, s. 4]. Są tacy autorzy, jak John Horgan, współpracownik „Scientific American”, którzy mówią nawet o końcu nauki ze względu na wyczerpanie się jej potencjału rozwojowego i poznawczego. Według niego największym zagrożeniem dla nauk ścisłych jest utrata ich specjalnego miejsca w hierarchii dyscyplin naukowych i zredukowanie ich znaczenia do czegoś w rodzaju krytyki literackiej, ze względu na coraz większą liczbę teoretyków zajmujących się „zagadywaniem” teorii, zamiast tworzeniem prac oryginalnych [Horgan 1996].

Brak tu miejsca na szczegółowe opisywanie olbrzymich zmian, jakim podlega nauka od ostatniej dekady XX w. Warto jednak pokrótce przedstawić przynajmniej niektóre z nich, aby ułatwić późniejszą dyskusję o bibliotece cyfrowej oraz nowej infrastrukturze służącej badaniom naukowym i edukacji na poziomie wyższym.

Zmiany te związane są z:

- przejściem od „produkcji” wiedzy w pojedynczych laboratoriach do współpracy na skalę międzynarodową, wręcz globalną, w tym sieciowe współużytkowanie instrumentów i sprzętu laboratoryjnego;
- szerszą dystrybucją wyników badań, w tym nie tylko w postaci publikacji, ale też zestawów nieprzetworzonych danych, gromadzonych w wielkich bazach danych;
- jednoczesnym (paradoksalnie) zamykaniem dostępu do wyników badań, które coraz częściej stają się własnością prywatną;
- wzrostem interdyscyplinarności badań, co wiąże się także z poszerzonymi możliwościami współpracy;
- jednoczesnym (paradoksalnie) wzrostem specjalizacji;
- bliższym związkiem badań naukowych z potrzebami społecznymi (przynajmniej deklaratywnie);
- zastosowaniu nowych technologii informacyjnych w modelowaniu, symulacji i analizie badanych zjawisk.

¹⁶ Liczba czasopism hybrydowych, czyli publikowanych w obu formach: drukowanej i elektronicznej, według prognoz British Library będzie maleć na rzecz czasopism czysto elektronicznych [Shenton 2005, s. 12]. Według tych samych prognoz czasopisma tradycyjne (drukowane) będą w ogóle zanikać.

¹⁷ Mam tu na myśli modele, w których dość daleko odchodzi się od „wrzucania” publikacji do Sieci przez ich autorów; chodzi raczej o pobieranie opłat przez wydawnictwa od autorów publikacji (subsydiowanych w tym celu przez różne źródła, głównie rząd) i udostępnianie ich w trybie OA. Taki model zasadniczo zmienia miejsce biblioteki w komunikacji naukowej.

¹⁸ Zasady te, sformułowane w 1942 r. to: uniwersalizm (brak ograniczeń społecznych i narodowych), „komunizm” (dzielenie się informacjami naukowymi wśród uczonych), bezinteresowność i zorganizowany sceptycyzm. Stanowią one m.in. o traktowaniu publikacji naukowych jako dobra publicznego [Merton 2002, s. 583-591].

Można odnosić wrażenie, że zmiany dotyczą głównie nauk ścisłych i technicznych, bo początkowo to z nich pochodziły symptomy, określane terminem e-nauki lub informacyjnej infrastruktury nauki. W ostatnich latach w tym samym kierunku podążają nauki społeczne, humanistyczne a nawet część obszarów sztuki, gdzie powstaje wiele najbardziej twórczych zastosowań, bazujących na technologiach informacyjnych oraz publikowaniu elektronicznym [Lynch 2006b]. W efekcie częściej używa się bardziej neutralnego terminu e-badania.

Techniki obliczeniowe i komputeryzacja stały się podstawą funkcjonowania wielu dyscyplin nauki do tego stopnia, że mówi się o „nauce obliczeniowej” (*scientific computing* lub *computational science*). Zmieniając sposób funkcjonowania nauki, umożliwiają dalszy rozwój przez udostępnianie nowej metodologii badań; dzięki nim możliwe jest prowadzenie eksperymentów, które byłyby nie do pomyślenia jeszcze 10 lat temu. Powstają też nowego rodzaju dane naukowe, coraz bardziej złożone i powiększające objętość w sposób wykładniczy, co powoduje tzw. efekt skali [Borgman 2007, s. 6]. Wraz ze wzrostem ilości danych pojawiają się nowe technologie informacyjne (na przykład techniki wyszukiwania informacji), umożliwiające utrzymanie dotychczasowej efektywności działalności informacyjnej. Zalew danych wymusza tworzenie nowych sposobów zarządzania danymi i produktów informacyjnych, powstających w wyniku ich analizy. Wraz ze wzrostem zasobów danych i liczby uczonych, zaangażowanych w ich tworzenie, następuje coraz silniejsza tendencja do automatyzacji zarządzania danymi. Zautomatyzowane gromadzenie danych jest mniej elastyczne, gdyż z góry podejmować trzeba wiele decyzji o tym, jakie elementy danych gromadzić i w jaki sposób je opisywać. Umieszczanie danych w repozytoriach umożliwia ich porównywanie i rozpowszechnianie wśród osób spoza własnego zespołu badawczego, wymaga jednak jeszcze większej standaryzacji, która dodatkowo zmniejsza elastyczność zasobu. Te niedogodności rekompensuje fakt, że dane raz zebrane i udokumentowane mogą być wielokrotnie wykorzystywane przez wielorakie narzędzia do analiz i wizualizacji.

Współczesna nauka charakteryzuje się realizacją wielkich, monumentalnych przedsięwzięć (wielkie teleskopy, cyklotrony, badania genomu ludzkiego), wymagających znacznych nakładów i współpracy międzynarodowych grup uczonych oraz pełnego stosowania technik komputerowych. Nowoczesna nauka, będąc obecnie większa niż kiedykolwiek [Griffiths 2000], prowadzi do standaryzacji zasad realizacji procesów i ich produktów. Powstanie bibliotek cyfrowych, repozytoriów czy standardów metadanych jest więc bezpośrednim efektem tego typu organizacji działań naukowych. Dla obsługi przedsięwzięć naukowych na wielką skalę, realizowanych w globalnej współpracy uczonych, niezbędne jest utworzenie infrastruktury, która mogłaby sprostać tym wymaganiom poprzez dostarczanie rozproszonego dostępu do instrumentów, zasobów obliczeniowych oraz zasobów danych. W takim sensie IIN oraz GBC są efektem procesu przechodzenia od małej do wielkiej nauki¹⁰ [Borgman, Wallis, Enyedy 2007, s. 18]. Nawet Internet może być uważany za „produkt uboczny” wielkiej nauki służący usprawnieniu komunikacji naukowej w skali globalnej, gdyż współczesne badania albo są globalne, albo nie są naukowe [Catsells 2008, s. 127].

Globalizacja działań wielkiej nauki ma więc silny wpływ na sposób funkcjonowania

¹⁰ Od lat sześćdziesiątych XX wieku dla obrazowania procesów zachodzących w nauce stosowane są określenia „wielka nauka” i „mała nauka”. Są one używane w dwóch znaczeniach. Pierwotnie za wielką nauką były uważane monumentalne przedsięwzięcia, na miarę aspiracji i możliwości finansowych nowoczesnych społeczeństw [Weinberg 1961, s. 161], jak na przykład wielkie teleskopy, cyklotrony, badania genomu ludzkiego, wymagające wielkich nakładów i współpracy międzynarodowych grup uczonych oraz pełnego stosowania najnowszej techniki komputerowej. W takim sensie IIN również należy do przedsięwzięć dotyczących wielkiej nauki, wymagających wielkich nakładów. Inny nieco pogląd na to zagadnienie przedstawił Derek de Solla Price, który za wielką naukę uważał dojrzale, lecz niekoniecznie wielkie przedsięwzięcia naukowe [Price 1965, s. 100].

komunikacji naukowej. W tym zakresie na przełomie XX i XXI wieku nasiliły się wcześniejsze już obserwowane tendencje, zarówno pozytywne, jak i negatywne. Do tych pierwszych należy zaliczyć wzrost mocy i dostępności technologii komputerowej, prowadzącej do powstania i upowszechnienia się informatycznej infrastruktury nauki, o czym będzie mowa w kolejnych rozdziałach. Nawet książki drukowane i artykuły w czasopiśmie uzupełniane są masowymi danymi, multimediami, oprogramowaniem, symulacjami, narzędziami interaktywnymi, stronami Web, bazami danych oraz innymi, w sposób ciągły aktualizowanymi, zasobami informacji online. Zjawiskiem negatywnym natomiast jest kryzys tradycyjnych sposobów organizacji wymiany i rozpowszechniania wyników badań naukowych [Davidson 2009, s. 36-37].

Jak twierdzi Andrew Odlyzko, u podstaw kryzysu komunikacji naukowej leży wykładniczy wzrost liczby publikacji naukowych [Odlyzko 1994]. Wymownym przykładem jest matematyka, bardzo stara nauka, gdzie akumulacja publikacji następowała przez stulecia. Mimo tego, znaczna większość publikacji w tej dyscyplinie powstała stosunkowo niedawno. W 1870 r. z zakresu matematyki opublikowano jedynie 840 artykułów (średnio 2,3 dziennie). W latach dziewięćdziesiątych XX wieku publikowano około 50 tys. artykułów rocznie (137 dziennie). Obecnie do dziedzinowej bazy danych MathSciNet dodawanych jest 350 opisów bibliograficznych dziennie (ponad 120 tys. rocznie), a sama baza liczy prawie 2 mln opisów. Prostą konsekwencją postępu geometrycznego, charakteryzującego przyspieszenie publikowania w matematyce jest fakt, że większość do tej pory opublikowanych artykułów ukazała się w ciągu ostatnich 10 lat.

Wzrost liczby publikacji dobrze ilustrują dane uzyskane przez Michaela Mabe i Mayura Amina z analizy katalogu Ulrichsa [Mabe, Amin 2001, s. 155]. Autorzy ci przedstawili analizy zmian ilości czasopism w całym XX wieku. Wynika z nich podwajanie się liczby publikowanych, recenzowanych tytułów czasopism naukowych w drugiej połowie XX wieku co 15 lat. Wcześniej (w pierwszej połowie XX w.) wzrost liczby tytułów czasopism przebiegał w przybliżeniu liniowo. Taki sam wynik uzyskał Albert Henderson badając zmiany liczby cytowań w największych, dziedzinowych bazach danych, takich jak Physics Abstracts, Engineering Index oraz Chemical Abstracts Service [Henderson 2002, s. 10].

Podobny wzrost cechuje inne wskaźniki rozwoju nauki, takie jak liczba uczonych. American Mathematical Society (AMS) liczyło 1926 członków w 1930 r., 6725 w 1960 r. i 25 623 w 1990 r., co oznacza podwajanie liczby członków tego towarzystwa średnio co 16 lat. Podobny wzrost notowało Polskie Towarzystwo Matematyczne (PTM), liczące w 1938 r. 216 członków, 555 członków w 1962 r. i ponad 1900 członków obecnie. Od połowy XX wieku do jego końca liczba uczonych w USA wzrosła siedmiokrotnie, z 1 mln do 7 mln osób [Griffiths 2000]. Wielu autorów przewiduje spadek tak olbrzymiego przyrostu liczby pracowników nauki i ich publikacji; w tym pierwszym przynajmniej mają rację. Liczba członków AMS wynosi obecnie około 29 tys. Jak pisze Andrew Odlyzko, przy światowej populacji ludności liczącej 10 mld (czego można oczekiwać w ciągu kilku najbliższych dekad), w drugiej połowie XXI wieku może okazać się, że na świecie będzie 10 razy więcej pracowników nauki, niż obecnie. Nawet, jeżeli w jakiś sposób spowodowalibyśmy zatrzymanie wzrostu ilościowego publikacji naukowych, to nadal liczba bieżąco produkowanych publikacji podwajałaby się co 20 lat [Odlyzko 1994].

Wykładniczy rozwój wielu elementów środowiska nauki nie może trwać w nieskończoność. Obecnie obserwuje się spadek tempa wzrostu liczby publikacji, jednak bezwzględna liczba publikacji nadal rośnie. W krajach, gdzie wzrost był największy, liczba publikacji stabilizuje się i wszystko wskazuje na to, że nie będzie rosta w dotychczasowym tempie. Tego typu zahamowania zdarzały się już jednak w przeszłości (na przykład w latach siedemdziesiątych XX w.), a mimo to silne tendencje wzrostowe powracały. Obecnie największe ośrodki naukowe skupiają się w Ameryce Północnej, Europie (zwłaszcza

Zachodniej) i Japonii. Tam sytuacja stabilizuje się. Gwałtownego wzrostu natomiast należy spodziewać się w takich krajach jak Chiny, Indie i inne kraje Dalekiego Wschodu oraz kraje arabskie. Powstają tam olbrzymie ilości publikacji naukowych, które ich autorzy próbują umieścić w najbardziej prestiżowych czasopismach, zdominowanych dotąd przez środowiska naukowe USA i Europy Zachodniej²⁰. Taka sytuacja niesie bardzo poważne konsekwencje zarówno dla samych uczonych jak i dla wszystkich instytucji funkcjonujących w obszarze komunikacji naukowej, w tym dla bibliotek naukowych.

W wyniku tej sytuacji od dawna żaden pracownik nauki nie jest w stanie choćby tylko dowiedzieć się o istnieniu wszystkich publikacji ze swojej dziedziny (nawet biorąc pod uwagę coraz większą specjalizację), nie mówiąc o ich przeczytaniu. Jedną z podstawowych umiejętności w procesie badań naukowych staje się zatem zdolność do szybkiego przeglądania dużych ilości materiałów i trafna selekcja interesujących dokumentów, z którymi należy zapoznać się bliżej. Stąd wzrastająca rola pracowników informacji, wspomagających badania z pozycji członka zespołu badawczego. Ciężar zapoznania się z bieżącą literaturą ostatecznie spada jednak na pracowników nauki. W rezultacie uczeni zamykają się w małych, współpracujących grupach, zajmujących się tym samym zagadnieniem, posługujących się podobną terminologią specjalistyczną, która przestaje często być zrozumiała dla innych uczonych z tej samej dziedziny, ale spoza grupy.

Efekty pracy uczonych oceniane są głównie na podstawie ich publikacji. Powoduje to, że pracownicy nauki czasem starają się publikować częściej niż byłoby to niezbędne dla przedstawienia rezultatów ich prac, sztucznie na przykład dzieląc jeden tekst na kilka artykułów. Spotkać można twierdzenia, że tendencja wzrostu liczby współautorów publikacji wynika częściowo także z konieczności jak najczęstszego publikowania. Znany jest też negatywny wpływ oceny wyników pracy uczonych za pomocą IF. Starają się oni publikować w czasopismach o najwyższym IF, co znacznie wydłuża czas oczekiwania na publikację w tych źródłach (nawet do kilku lat). W efekcie publikacje takie w wielu dziedzinach, w momencie ukazania się, mają często wartość historyczną. Nie zawsze niezbędne mnożenie tekstów jest także wynikiem wymagań, stawianych kandydatom na pracowników naukowych, którzy muszą legitymować się dorobkiem naukowym w postaci określonej liczby publikacji recenzowanych. Pisanie książek głównie dla uzyskania tytułu czy stanowiska może obniżać poziom tych prac.

Inaczej niż w nienaukowych sferach twórczości (na przykład w literaturze pięknej) publikacje naukowe tworzone są przez uczonych i dla uczonych²¹. Publikowanie naukowe, ze względu na swoją naturę, nie może wykorzystywać zjawiska wzrostu skali rynku (wzrostu liczby uczonych), tak jak pozostała część komercyjnego rynku księgarskiego. Wraz ze wzrostem liczby pracowników nauki i publikacji zwiększa się bowiem tendencja do zawężania specjalizacji, w wyniku czego liczba odbiorców pozostaje stała. Ośrodki zatrudniające uczonych (zazwyczaj uczelnie) nie rozwijają się bardzo szybko w sensie powstawania nowych struktur organizacyjnych i stanowisk pracy, więc pracownicy nauki pozostają coraz bardziej rozproszeni. Bez względu na to należy jednak zapewnić im pełny dostęp do literatury. Prowadzi to do kryzysu objawiającego się brakiem możliwości zakupu wszystkich niezbędnych publikacji ze względu na brak funduszy.

Na zjawisko gwałtownego wzrostu liczby pracowników nauki i publikacji naukowych nałożyły się inne: komercjalizacji rynku wydawniczego i jego konsolidacja, co daje moż-

²⁰ W latach 1995-2005 nastąpił wzrost publikowanych artykułów: w Chinach o 17%, Tajwanie o 16%, Indiach o 4,5%, a w USA o 0,6%, w Europie o 1,8%. Pomimo tych różnic w USA wciąż publikuje się 29% światowych artykułów [Ware, Mabe 2009, s. 21].

²¹ Zauważmy, że stopień prawdziwości tego twierdzenia zależy także od dyscypliny. W dziedzinach praktycznych, takich jak np. medycyna, można wyróżnić publikujących pracowników nauki i znacznie większą grupę lekarzy – praktyków głównie czytających. Inaczej jest np. w matematyce czy fizyce, gdzie autorzy i czytelnicy stanowią tę samą grupę: uczonych.

liwość stosowania praktyk monopolistycznych. Pewne symptomy tej sytuacji na Zachodzie pojawiły się już w latach siedemdziesiątych XX w., gdy uczelnie zaczęły ograniczać subsydia dla własnych, lokalnych wydawnictw i te po kilku latach zmuszone były zaprzestać działalności [Ryan i in. 2002, s. 173]. Biblioteki naukowe, przeznaczając większość swoich funduszy na czasopisma, znacznie zmniejszyły zakupy książek, wydawanych przez wydawnictwa uczelniane. W efekcie te oficyny wydawnicze, a także należące do towarzystw naukowych, czy też tytuły wydawanych przez nie czasopism, wykupywane były przez wydawców komercyjnych. Wydawnictwa uczelniane starały się ograniczać ryzyko finansowe, publikując głównie podręczniki i książki zawierające treści, mogące zainteresować szerszą publiczność. Jednym z objawów wycofywania się z niepewnych przedsięwzięć jest zaniechanie publikowania serii. Największe wydawnictwa komercyjne, po przejściu roli wydawców publikacji naukowych, rozpoczęły konsolidację rynku, poprzez wchłanianie mniejszych firm. Połączeń i wykupów firm jest tak dużo, że trudno za nimi nadążyć. Dla przykładu podam, że Elsevier przejął między innymi takie, działające wcześniej samodzielnie wydawnictwa, jak: Pergamon, North-Holland, Excerpta Medica, The Lancet, LexisNexis, Congressional Information Service, BioMedNet, Beilstein, El Inc. i inne. Już w 2001 r. Elsevier kontrolował 20% rynku czasopism STM²². Wzorem innych przedsiębiorstw komercyjnych, tego rodzaju firmy wykupują nie tylko inne wydawnictwa, ale także przedsiębiorstwa, działające na pokrewnych rynkach, takich jak produkcja baz danych, oprogramowania bibliotecznego czy sieci księgarń i elektroniczne sieci przekazywania zamówień dla księgarń i hurtowni. Warto dodać, że połączenia firm często generują wzrost kosztów; po fuzji Elsewiera z Pergamonem, tytuły publikowane przez Pergamon podrożały o 27%, a Elsewiera o 8%. Podobne efekty dały inne komasacje [McCabe 2002, s. 262].

W pozostałych na rynku towarzystwach naukowych nastąpiły zmiany, które można nazwać pseudokomercjalizacją [Orsdel 2007, s. 193]. Pseudokomercyjni wydawcy towarzystw naukowych przejęli sposób działania wydawców komercyjnych w zakresie wyznaczania cen, marketingu i tworzenia pakietów z tytułów publikowanych przez siebie czasopism. Ostatecznie wiele funkcjonujących wydawców towarzystw naukowych bardziej przypomina instytucje komercyjne, niż naukowe.

Każde czasopismo naukowe jest niepowtarzalną instytucją, posiadającą swoją reputację i pozycję w świecie publikacji naukowych. Dla funkcjonowania nauki w każdej dyscyplinie musi istnieć pewien podstawowy zestaw tytułów, stąd konkurencja pomiędzy czasopismami STM nie funkcjonuje w taki sam sposób, jak na rynku magazynów ilustrowanych czy dóbr materialnych. Powoduje to występowanie niedoskonałego rynku, charakteryzującego się produkcją dotowaną i nieelastycznym popytem, w którym nie istnieje tradycyjna równowaga spowodowana swobodną konkurencją.

Nic dziwnego, że na takim rynku ceny publikacji naukowych (szczególnie czasopism) rosą w zawrotnym tempie. Wydatki bibliotek naukowych USA, zrzeszonych w Association of Research Libraries (ARL) na zakup czasopism wzrosły w latach 1986-2005 o ponad 300%, przy wzroście wydatków na książki o niecałe 60% i inflacji wynoszącej 78%. Również badania Carol Tenopir i Donald Kinga wskazują na wzrost cen czasopism w okresie 1975-1995 ponad inflację 2,6 razy [Tenopir, King 2000].

Szczególna sytuacja panuje w zakresie cen czasopism w najszybciej rozwijających się dziedzinach, takich jak medycyna, chemia, nauki biologiczne, wysokie technologie (ogólnie dotyczy czasopism STM) [Townsend 2003, s. 32]. Tam więc, gdzie informacja powinna być dostępna jak najszybciej i rozpowszechniana jak najszerszej, dostęp do niej jest hamowany przez ceny narzucone przez wydawców. Z badań prowadzonych przez Martę Kyrillidou i Marka Younga wynika, że ceny czasopism STM wzrosły w latach 1980-2002

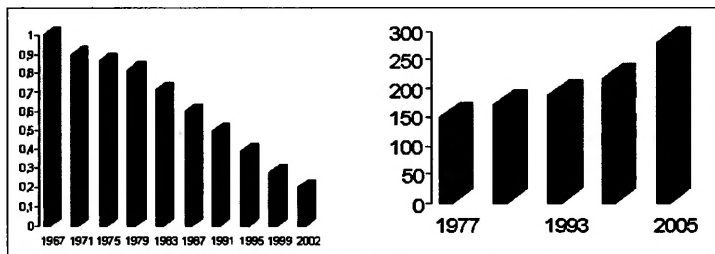
²² STM – skrót od ang. Science, Technology, Medicine, dotyczy czasopism z zakresu nauk ścisłych techniki i medycyny, dziedzin najszybciej się rozwijających, w których czasopisma osiągają najwyższe ceny.

o 600%, a w niektórych dziedzinach, jak chemia nawet o 750%. W tym samym czasie wolniej drożały periodyki z historii czy literaturoznawstwa: o około 250% [Kyrillidou, Young 2006, s. 17]. Ceny te nadal rosną; w 2008 r. średnio za roczną prenumeratę czasopisma z fizyki i chemii trzeba było zapłacić ponad 3000 USD, a z medycyny 1330 USD [Orsdel, Born 2008].

W wymienionym okresie czasopisma drożały o około 11% rocznie, podczas gdy roczna inflacja wynosiła 2,5%. Tylko niewielką część podwyżki cen można wytłumaczyć większą liczbą opublikowanych stron²³. Również zmiana kosztów produkcji czy wprowadzenie technologii cyfrowej nie tłumaczy takiego wzrostu. Prenumerata czasopism u wydawców non profit (na przykład uczelni lub towarzystw naukowych) kosztuje o około 50-75% mniej, niż u wydawców komercyjnych.

Wzrost cen i kosztów powoduje ograniczenie rozmiarów prenumeraty czasopism naukowych w bibliotekach. Ponieważ tzw. koszt pierwszego egzemplarza (koszty, jakie ponosi wydawca, aby przygotować teksty do opublikowania) jest stały i wysoki, wydawcom, którzy chcą, przy malejącej prenumeracie, zachować zyski na dotychczasowym poziomie, pozostaje dalsze podwyższanie cen czasopism. Powstaje błędne koło komunikacji naukowej:

- Aby funkcjonować w nauce, autorzy muszą wydawać swoje publikacje.
- Aby czerpać zyski wydawcy muszą przygotowywać nowe wydawnictwa.
- Liczba publikacji danego wydawcy rośnie, wzmacniając jego pozycję na rynku.
- Dla wydania tych publikacji tworzone są nowe tytuły czasopism, co podnosi sumaryczne koszty stałe (które muszą być ponoszone dla każdego czasopisma).
- Dla wyrównania rosnących sumarycznych kosztów stałych i utrzymania dotychczasowych zysków wydawca podnosi cenę jednostkową czasopisma.
- W związku z wyższą ceną jednostkową spada liczba zamówionych egzemplarzy (prenumerata), co powoduje dalszy wzrost kosztów stałych, a to zmusza do podnoszenia cen.



Rys. 3. Porównanie wpływu wzrostu cen czasopism na prenumeratę (lewa strona) [Schlingcn, Kronenfeld 2004, s. 310] oraz wzrostu liczby czytanych artykułów (prawa strona) [Tenopir, King 2008]

Kryzysową sytuację w komunikacji naukowej ilustruje zestawienie dwóch wykresów (rys. 3). Po lewej stronie zaprezentowane są zmiany siły nabywczej bibliotek w odniesieniu do czasopism od 1967 r. (oznaczonej jako 1) do 2002 r. (spadek do 0,2). Zakładając, iż biblioteki otrzymywały corocznie budżet powiększony o wskaźnik inflacji, po 35 latach, z powodu wzrostu cen czasopism, mogły kupić tylko 20% pierwotnej liczby tytułów. Po

²³ W latach 1975-2001 objętość przeciętnego artykułu wzrosła z 7,4 do 12,4 strony, a czasopisma z 820 do 2216 stron rocznie, co wynika także ze wzrostu liczby publikowanych artykułów rocznie w przeciętnym czasopiśmie z 83 do 154 artykułów [King, Alvarado-Albertorio 2008, s. 253].

prawej stronie przedstawiony jest wzrost liczby czytanych w ciągu roku artykułów przez uczonych z USA w naukach medycznych. Wynika z niego stały wzrost liczby czytanych publikacji; w latach 1977-2005 uległa prawie podwojeniu.

Jak widać, biblioteki nie są w stanie sprostać zwiększającym się potrzebom dostępu do literatury przez pracowników nauki. Osiągnięcie sprzedaży 300 egz. tytułu czasopisma naukowego w prenumeracie uznawane jest za sukces wydawniczy. Większa prenumerata dotyczy tylko najważniejszych, światowych czasopism naukowych [Boissy, Feick, Knapp 2007, s. 169]. Dodatkowo nasila się zjawisko rezygnacji z prenumeraty indywidualnej²⁴, co łącznie powoduje gwałtowny wzrost barier w dostępie do publikacji naukowych. Sytuację tę łagodzi nieco organizowanie się bibliotek w konsorcja (które zwykle są w stanie wynegocjować niższe ceny, a koszty rozkładają się na wszystkich członków konsorcjów), jednak nie zmienia to ogólnej tendencji.

W skrócie można więc opisać aktualną sytuację następująco: liczba badań i publikacji rośnie, ale dostęp do nich jest coraz trudniejszy. W efekcie żaden z uczestników procesu komunikacji naukowej nie jest w pełni zadowolony. Autorzy z jednej strony oddają swoje teksty wydawcom za nieadekwatne wynagrodzenie lub wręcz bezpłatnie, w dodatku przekazują im prawa majątkowe (*copyright*) i świadczą inne usługi (jak bezpłatne recenzowanie), a z drugiej strony nie mogą dotrzeć do potrzebnych materiałów, bo nie są one kupowane przez biblioteki, zmuszone do skreślania tytułów z prenumeraty. Wydawcy niezadowoleni są z malejącej prenumeraty, starają się utrzymać odpowiedni zysk (wynoszący od 20 do 40% [Bergstrom Carl, Bergstrom Theodore 2004, s. 898], co stanowi jeden z najlepszych wyników w działalności komercyjnej) poprzez komasację rynku i wchodzenie w jego nowe obszary, usuwając z niego małych wydawców. Pośrednicy tracą rynek, gdyż wydawcy rezygnują z ich usług na rzecz własnych serwisów online. Bibliotekarze w obliczu rosnących cen zmuszeni są redukować liczbę prenumerowanych tytułów, ograniczając tym samym możliwości zaspokajania potrzeb swoich użytkowników; ci z kolei szukając innych rozwiązań, zwracając się na przykład ku internetowym serwisom wydawców typu *pay-per-view*. Instytucje naukowe w poszukiwaniu dodatkowych funduszy podwyższają opłaty za studia i narzuty ogólne na fundusze otrzymywane przez uczonych z grantów indywidualnych.

Pomimo łagodzenia w ostatnich latach negatywnych zjawisk dzięki poprawie dystrybucji treści naukowych w wyniku ich cyfryzacji, przez co spadek oferty bibliotek naukowych został zahamowany, dalsze utrzymywanie się aktualnych tendencji może doprowadzić do utraty możliwości realizacji przez pracowników nauki ich dwóch podstawowych zadań: tworzenia i rozpowszechniania wiedzy. Tworzenia, ponieważ zabraknie dostępu do podstawowych źródeł; a rozpowszechniania, ponieważ coraz trudniej jest o publikację w renomowanym czasopiśmie, a na dostęp do opublikowanych wyników badań wielu nie będzie mogło sobie pozwolić. Widać więc, że aktualna sytuacja charakteryzuje się błędnymi kołami i sprzężeniami zwrotnymi na wielu poziomach. W takich wypadkach jedynym rozwiązaniem jest znalezienie miejsca, w którym błędne koło może zostać przerwane.

Poszczególne grupy uczestników komunikacji naukowej poszukują zadowalającego wyjścia z sytuacji. Jednym z nich jest wykorzystanie nowych rozwiązań technicznych, głównie Internetu, w tym do udostępniania²⁵ Open Access (zob. p. 4.1), co wiąże się ze wzrostem znaczenia komunikacji nieformalnej i powrotem do czasów, gdy w rozpowszechnianiu wyników naukowych nie królowała komercja. Powstają inicjatywy „oddol-

²⁴ W 1977 r. uczeni w USA w 60% przypadków korzystali z prenumeraty indywidualnej, a w 24,8% z artykułów w bibliotekach; w 2005 r. proporcje się odwróciły: 22,6% korzystało z prenumeraty, 62% z bibliotek [Tenopir i in. 2009, s. 17].

²⁵ Open Access jest modelem dystrybucji, a nie kosztów; może być stosowany zarówno przez wydawnictwa komercyjne, jak i inne. Jednak z OA zwykle wiąże się potrzeba stosowania nowych modeli kosztów, gdyż tradycyjny model oparty na prenumeracie jest nieprzydatny jako sprzeczny z definicją OA.

ne”, realizowane przez samych pracowników nauki, takie jak Public Library of Science²⁶ (PLoS), mające na celu usprawnienie komunikacji naukowej poprzez uwolnienie jej z rąk wydawnictw komercyjnych. Inicjatywy te tworzą jednocześnie nowe modele publikowania i obiegu informacji, zmieniające istniejące stosunki. Zmiany te dotyczą nie tylko wydawców komercyjnych, ale także bibliotek naukowych. Według modelu propagowanego przez PLoS, autorzy mają płacić wydawcy nonprofit za opublikowanie swojego tekstu. Zazwyczaj nie pokrywają tych kosztów bezpośrednio z własnych dochodów, ale z dotacji, grantów i podobnych źródeł. Korzystając z tych funduszy wydawcy przygotowują publikację i umieszczają w Sieci, dostępną na zasadzie Open Access. Innymi, równie oddolnymi inicjatywami jest tworzenie różnego rodzaju repozytoriów (dziedzinowych, instytucjonalnych). Jednocześnie formułowane są nowe formy ochrony własności intelektualnej dzieł udostępnianych w Sieci, takie jak Creative Commons, stojące w opozycji do obecnie obowiązującego, bardzo restrykcyjnego prawa copyright (zob. p. 5.3). Jak widać, w tym modelu brak miejsca dla bibliotek; do tej pory to biblioteki głównie uczelni otrzymywały dotacje na zakupy czasopism i rozdysponowywały je w imieniu swoich użytkowników. Teraz użytkownicy (i autorzy jednocześnie) sami decydują o kwotach przekazywanych wydawcom publikacji Open Access na opublikowanie i udostępnienie artykułów naukowych. Fundusze przekazywane dotąd bibliotekom na zakup publikacji trafią więc bezpośrednio od autorów do wydawców z pominięciem bibliotek, których pośrednictwo staje się zbędne. Bibliotekom natomiast może przypaść rola archiwów, gromadzących dawne, drukowane egzemplarze czasopism i książek (zob. p. 5.1).

Wraz z gwałtownym wzrostem liczby publikacji – zalewem – ich rozpowszechnianie w trybie OA może mieć dodatkowy walor. Użytkownik stojący przed koniecznością selekcji dostępnych zasobów wybierze te, które są łatwiej dostępne, a więc w formie cyfrowej i najlepiej bez żadnych dalszych ograniczeń w postaci opłat, haseł dostępu lub rejestracji. W następstwie takich postaw to, co nie jest dostępne w takiej formie zniknie ze świadomości odbiorców [Gawrysiak 2008, s. 347].

Trudno powiedzieć, czy ten model publikowania rozpowszechni się na tyle, aby zmienić sposób funkcjonowania światowego systemu komunikacji naukowej. Sam Open Access nie jest pozbawiony pewnych nierozwiązanych do końca problemów, takich jak kłopoty z zapewnieniem jakości materiałów, udostępnianych w Sieci bezpośrednio przez autora (przed recenzją²⁷) czy poszukiwania niezawodnych sposobów długotrwałej archiwizacji (zob. p. 4.1). Nic dziwnego, że wszyscy poszukują wyjścia z opisanej, kryzysowej sytuacji, często upatrując je w zastosowaniu nowych technologii i zasad organizacyjnych w komunikacji naukowej, na przykład takich, jakie oferuje informatyczna infrastruktura nauki.

1.3. Organizacja nauki w środowisku cyfrowym: Informatyczna infrastruktura nauki

Już dawno negatywnie zweryfikowane zostały poglądy, że stały i szybki rozwój nauk spowoduje rozwiązanie wszystkich problemów. Wiadomo, że odpowiedź na jedno pytanie powoduje potrzebę formułowania następnych. Wiele z nich pozostawało bez od-

²⁶ <http://www.plos.org>.

²⁷ Sam proces tradycyjnego recenzowania także bywa krytykowany, np. Fiona Godlee i Kay Dickersin opisują go jako „kosztowny, powolny, subiektywny, tendencyjny, nieodporny na nadużycia, wyrwykowy w wykrywaniu ważnych błędów metodologicznych i prawie zupełnie bezużyteczny w wykrywaniu oszustw lub zachowań nieetycznych” [Godlee, Dickersin 2003, s. 92]. Można także zauważyć, że w różnych dyscyplinach istnieją zróżnicowane potrzeby co do kontroli jakości, na co wskazuje stosunek artykułów przyjętych do odrzuconych dla różnych dyscyplin (medycyna – bardzo duży odsetek tekstów odrzuconych, nauki ścisłe – znacznie większy odsetek publikacji przyjętych).

powiedzi, a może nie zostałyby nawet postawione, gdyby nie stały postęp w technikach informatycznych i informacyjnych. Zaawansowane techniki obliczeniowe początkowo stosowane były przez ograniczoną liczebnie elitę uczonych, pracujących nad wybranymi zagadnieniami w wyizolowanych ośrodkach badawczych; obecnie stają się podstawą coraz większej grupy badań, stanowiąc warunek przyszłego rozwoju zarówno nauk ścisłych i technicznych, jak również humanistycznych i społecznych [Pindlowa 2008, s. 154].

Gwałtowny rozwój wydajności sprzętu komputerowego, powodujący coroczne podważanie się wielu wskaźników jego efektywności²⁸, spowodował upowszechnienie świadomości, że hardware, nawet najpotężniejszy, nie jest wystarczającym warunkiem do prowadzenia badań wspomaganych komputerowo. Uczeni oczekują dostępu do inteligentnego oprogramowania, narzędzi umożliwiających wizualizację realizowanych procesów, middleware²⁹ i aplikacji naukowych tworzonych i stosowanych przez zespoły interdyscyplinarne. Oprócz niezbędnego postępu technicznego oczekuje się także rozwoju nowych modeli organizacyjnych, określenia zasad realizacji procesów i wyznaczenia brzegowych warunków ekonomicznych.

Nadzwyczajny sukces Internetu ukazuje wielką rolę, jaką odgrywa zaawansowana infrastruktura informatyczna, służąca rozwojowi innowacyjności w najszerszym znaczeniu, ułatwiając współpracę, wymianę danych, modelowanie i symulacje, komunikowanie idei, łączenie badań z ich zastosowaniami oraz innowacji technologicznych ze społecznymi. Jak twierdzi Brian Kahin, rolę Internetu w tym zakresie można porównywać tylko z siecią elektryczną, stanowiącą przykład równie rozwiniętej infrastruktury [Kahin 2007]. Transformacje Internetu powodowane są współdziałaniami i synergią z innymi formami infrastruktury – infrastrukturą nauki, telekomunikacyjną i dla potrzeb handlu. Sieć jest więc połączeniem infrastruktur: komercyjnej, publicznej i społecznej. Internet powoduje transformacje każdej z nich, w efekcie czego kształtowane i przekształcane jest sieciowe społeczeństwo informacyjne³⁰. Transformacje te są podobne do zmian powodowanych przez rozwój innych infrastruktur, chociaż modyfikacje zachodzą w tym przypadku szybciej i mają większy zakres.

Integracja, współdziałanie i współzależności zadań współczesnej nauki w zakresie działań informacyjnych wymuszają traktowanie jej w kategoriach infrastruktury, która tworzona współcześnie w celu wykorzystania gwałtownie zmienionej i wciąż zmieniającej się technologii informacyjnej w badaniach naukowych nazywana jest informatyczną infrastrukturą nauki (IIN), cyberinfrastrukturą (*cyberinfrastructure*) czy e-Nauką (*e-Science*).

Początkowo termin „infrastruktura” stosowano dla instalacji wojskowych, później

²⁸ W ciągu nieco ponad 25 lat (1980-2006) powstało osiem generacji procesorów komputerowych. W tym okresie czas dostępu do RAM spadł z ok. 50 ns do 2 ns (25 razy), a liczba tranzystorów w CPU wzrosła z ok. 29 tys. w procesorze 8086 do 300 mln w Core 2 Duo (podważanie co ok. 18 mies.). Podobne dane, dotyczące wydajności układów scalonych, podaje także Manuel Castells [Castells 2008, s. 54]. Wzrost tego typu (wykładniczy) obserwuje się w rozwoju większości elementów sprzętu komputerowego, co zgodnie jest z tzw. prawem Gordona Moore'a, współzałożyciela Intelu, sformułowanym w 1965 r. i aktualnym do dziś.

²⁹ Middleware (oprogramowanie pośredniczące) jest to rodzaj oprogramowania służącego do komunikacji aplikacji użytkownika z bazami danych lub innymi serwerami/usługami. Znajduje się ono pomiędzy siecią a aplikacjami. Oprogramowanie to pozwala na połączenia pomiędzy zasobami sieciowymi, które bez niego są odseparowane od siebie. Wykorzystanie middleware ułatwia tworzenie aplikacji, gdyż uniezależnia je od zastosowanego rodzaju bazy danych. Tworzone jest w celu ułatwienia korzystania i rozpowszechniania zasobów, takich jak komputery, dane, sieci, instrumenty oraz do wspomnienia współpracy i komunikacji w badaniach naukowych i edukacji. Obecnie, wobec rozwoju aplikacji wielowarstwowych, middleware zyskuje na znaczeniu.

³⁰ Według Piotra Sienkiewicza nowe struktury tego społeczeństwa powstają równolegle do kontinuum dane-informacja-wiedza-mądrość (zob. p. 2.4). Nowe klasy społeczne obejmują „proletariat”, funkcjonujący na poziomie danych, posługujący się ICT w celu realizacji podstawowych funkcji; „profesjonalistów”, „klasę średnią”, działających na poziomie informacji, dla których ICT jest narzędziem pracy; oraz „nomenklaturę”, dla której dostęp do wiedzy i mądrości jest środkiem realizacji dalekosiężnych celów [Sienkiewicz 2006, s. 69].

dowszelkich struktur, mających podstawowe znaczenie dla rozwoju wszelkich dziedzin gospodarki. Cechą charakterystyczną każdej infrastruktury jest jej głębokie zakorzenienie w innych strukturach, społecznych i technologicznych, w sposobach wykonywania działań i pracy, przez co jest ona bardzo transparentna w takim znaczeniu, że nie musimy zastanawiać się nad sposobami jej stosowania czy współdziałania z nią. Podczas prowadzenia samochodu jesteśmy „zakorzenieni” w fizycznym systemie dróg i usług towarzyszących (np. stacje napraw i benzynowe), ale także w społecznym systemie przyznawania prawa jazdy i funkcjonowania zasad kodeksu drogowego. Infrastruktura ta staje się bardziej widoczna dopiero w wyniku awarii; jej istnienie jest lepiej uświadamiane, gdy przestaje działać. Infrastrukturę buduje się na powierzchni zainstalowanej bazy; jest ona tworzona w postaci kolejnych, modułowych przyrostów, nie jednocześnie w każdym miejscu. Elementy infrastruktury mogą współdziałać ze sobą oraz rozwijać się dzięki tworzeniu i stosowaniu standardów i uzgodnień dotyczących konwencji społecznych. Przykładem może być obecnie istniejąca globalna infrastruktura telekomunikacyjna, powstająca na wcześniejszej działającej bazie, która budowana była w wielu miejscach przez ponad sto lat, a globalny zasięg uzyskała kilkadziesiąt lat temu.

Z kilku powodów infrastruktura informatyczna jest mniej widoczna niż infrastruktura fizyczna. Wiele jej elementów składowych, takich jak szybkie sieci i zaawansowane ośrodki obliczeniowe, jest rozproszonych i ukrytych przed użytkownikiem końcowym. Inne elementy infrastruktury informatycznej, na przykład programowanie, proces projektowania, umiejętności ludzi i know-how, nie ma charakteru fizycznego, namacalnego. Infrastruktura ta często nie jest jeszcze wystarczająco dojrzała do osiągnięcia takich cech charakterystycznych, jak zakorzenienie, transparentność czy globalny zasięg. Istotny jest także fakt, że infrastruktura informatyczna powstawała i była wdrażana dość gwałtownie (w krótkim czasie), w związku z czym konieczne było zapewnienie szerokiego udziału pracowników nauki i innych specjalistów w jej konstruowaniu. Podobnie jak dla innych rodzajów infrastruktury, ważne jest podejmowanie właściwych decyzji już w trakcie planowania i budowy IIN, gdyż po stworzeniu infrastruktury informatycznej będzie dużo trudniej zmienić lub poprawić zasady jej działania.

Amerykańska National Science Foundation (NSF) stosuje termin *cyberinfrastructure* dla infrastruktury rozproszonej technologii komputerowej, informacyjnej i komunikacyjnej [Atkins i in. 2003]. IIN w USA pozwała na integrację w sieci (szczególnie w Internecie) różnego rodzaju narzędzi informatycznych, takich jak sprzęt obliczeniowy (do przetwarzania danych i sieciowy), sensory cyfrowe, obserwatoria, urządzenia służące prowadzeniu eksperymentów i inne przyrządy badawcze oraz współdziałające usługi i narzędzia softwareowe stanowiące zestawy oprogramowania aplikacyjnego i middlewaru. Równie ważnym elementem IIN, jeśli nie najważniejszym, są odpowiednio przygotowani ludzie: interdyscyplinarne zespoły i specjaliści od technologii informacyjnej, mający doświadczenie w zakresie projektowania algorytmów, analiz systemowych i tworzenia aplikacji, mogący w pełni wykorzystać możliwości IIN do tworzenia, rozpowszechniania i archiwizacji naukowych danych, informacji i wiedzy. Stwarza ona więc warunki do współdziałania technologii, praktyk społecznych, organizacji pracy oraz standardów. Twórcy amerykańskiej cyberinfrastruktury stwierdzają, że podobnie jak infrastruktura fizyczna niezbędna jest dla ekonomii przemysłowej, tak infrastruktura informatyczna konieczna jest dla ekonomii wiedzy.

Raport Daniela Atkinsa wskazuje na dwa poziomy IIN: techniczny i organizacyjny, opisując je następująco:

- Infrastruktura techniczna składa się m.in. z middleware, aplikacji, protokołów wymiany danych i jest (lub może być) stosowana we wszystkich dziedzinach wiedzy.
- Infrastruktura organizacyjna tworzona jest przez politykę instytucjonalną w obrębie nauki i ulega większym wpływom zasad prawa i norm społecznych – wymienić tu można prawo dotyczące własności intelektualnej, tworzenie i adaptację standardów w obrębie

danej dziedziny i poza nią, narzędzia i usługi oraz edukację i szkolenia profesjonalne osób tworzących i stosujących IIN [Atkins i in. 2003, s. 12].

Jak pisze Peter Freeman, IIN jest podstawą, abstrakcyjną konstrukcją, na bazie której tworzone są inne. Można je nazywać infrastrukturami wyspecjalizowanymi lub ekologiaми wiedzy. IIN jest złożonym systemem, projektowanym jako system otwarty [Freeman 2007]. Składa się ona z tysięcy częściowo nakładających się na siebie społeczności kolaboratorów lub Gridów³¹ (zob. też p. 1.4), tworzonych dla wybranych dziedzin lub zastosowań, kastomizowanych na poziomie aplikacji, ale w bardzo szerokim zakresie współużytkujących jednolitą IIN, która powinna zawierać (zob. rys. 4):

- Gridy centrów obliczeniowych,
- różnego rodzaju zasoby obiektów cyfrowych, takie jak na przykład publikacje elektroniczne i oprogramowanie,
- multidyscyplinarne, zarządzane i zintegrowane zasoby danych naukowych,
- tysiące instrumentów online i rozległe matryce sensorów (*sensor arrays*),
- zestawy przyjaznych i efektywnych narzędzi służących wyszukiwaniu zasobów, modelowaniu i interaktywnej wizualizacji,
- narzędzia współpracy pomiędzy fizycznie rozproszonymi zespołami ludzi, stosującymi wymienione wcześniej elementy.

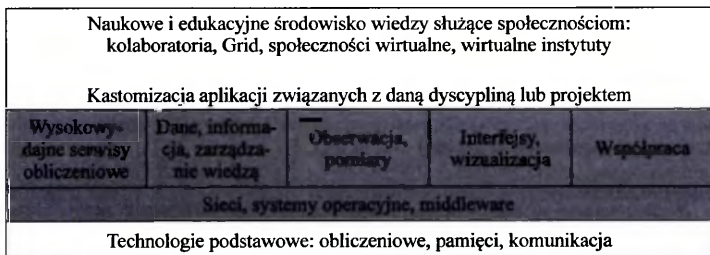
Według Programu Rozwoju Infrastruktury Informatycznej Nauki na lata 2007-2013 infrastruktura dla e-Science w Polsce powinna obejmować:

- Sieci komputerowe wraz z usługami, klastry i komputery dużej mocy, wraz z zaawansowanym oprogramowaniem naukowym i narzędziowym.
- Systemy składowania danych.
- Zasoby informacyjne, dostępne sieciowo bazy danych, także literaturowe i faktograficzne.
- Warstwę wspólną oprogramowania Gridowego zapewniającą łatwy dostęp do zasobów komputerowych oraz możliwość spójnego budowania Gridów dziedzinowych.
- Gridy dziedzinowe o znaczeniu ogólnonarodowym (specjalizowane platformy usługowe, „wirtualne³² laboratoria”), tworzone dla poszczególnych obszarów badań, jednak z wykorzystaniem uniwersalnych elementów *e-Infrastruktury* [Program rozwoju 2007, s. 6].

Na rys. 4. przedstawiłem rodzaje urządzeń i usług dostarczanych w sposób zintegrowany przez IIN (na ciemnym tle). Poziom IIN nadbudowany jest na warstwie zawierającej podstawą technologię obliczeniową, pamięci i komunikację. IIN powinna być tworzona i zarządzana w sposób umożliwiający pracownikom nauki lub projektom traktowanym całościowo korzystanie z własnych, efektywnych i dostosowanych do dziedziny, ale współdziałających środowisk wiedzy, służących badaniom i edukacji.

³¹ Grid (sieć, na wzór sieci elektrycznej) to system, który integruje i zarządza zasobami będącymi pod kontrolą różnych domen (od instytucji po system operacyjny) i połączonymi siecią komputerową; używa standardowych, otwartych protokołów i interfejsów ogólnego przeznaczenia (odkrywania i dostępu do zasobów, autoryzacji, uwierzytelniania) oraz dostarcza usług odpowiedniej jakości. Według Małgorzaty Janiak jest to infrastruktura umożliwiająca współużytkowanie zasobów informacyjnych oraz komputerowych, które pozostają rozproszone geograficznie. Innymi cechami Gridu jest: brak scentralizowania źródeł oraz zróżnicowany i nietrywialny charakter dostarczanych usług [Janiak 2007, s. 227]. Twórcą tego określenia i idei Gridu jest Ian Foster [Foster 2002].

³² Przymiotnik „wirtualny” rozumiem tutaj tak, jak jest ujmowany z punktu widzenia przetwarzania danych: jako potencjalny wszechświat porównywalny przy użyciu cyfrowego modelu i wejścia wspomaganego przez użytkownika (technologia informacyjna związana z modelami cyfrowymi [Lévy 1996, s. 37]). czyli nie w ujęciu potocznym (nierzeczywisty, wyobrażony, istniejący tylko na ekranie komputera [Golka 2008, s. 100]) ani filozoficznym (struktury ontologiczne wirtualnych projektów jako światów możliwych [Konik 2009, s. 83]).



■ = IIN: hardware, software, zasoby, usługi, personel, organizacja

Rys. 4. Usługi IIN tworzące środowisko wiedzy dla badań i edukacji

IIN, podobnie jak większość infrastruktur, nie jest zamierzona, ani budowana w sposób ściśle planowy, pod względem pełnej kontroli tego procesu i jego zarządzania. W związku z tym ostateczny rozwój infrastruktury jest trudny do przewidzenia, a w każdym razie nie wynika wprost z warunków początkowych [Jackson i in. 2007].

Efektywna IIN będzie tworzona z uwzględnieniem przedstawionych poniżej pięciu podstawowych zasad. Ma ona ułatwiać współpracę, umożliwiać innowacyjność, zapewniać długotrwałe funkcjonowanie przyjętych rozwiązań, pozwolić na współdziałanie poszczególnych elementów. Wiedza w niej zawarta powinna być traktowana jako dobro publiczne.

Ułatwienie współpracy

Dzięki nowym technologiom informacyjnym i komunikacyjnym koszty związane z osiągnięciem celów gospodarczych i/lub społecznych metodą współpracy są niższe niż koszty ponoszone w osiągnięciu tych celów drogą rywalizacji [Muraskiewicz 2002, s. 26]. Technologia cyfrowa sprzyja otwartości i współpracy, czego przejawy widoczne są w wielu miejscach i przedsięwzięciach, począwszy od sieci społecznych i Web 2.0 do tworzonych międzynarodowych zespołów badawczych. IIN powinna umożliwić wspólne podejmowanie badań przez specjalistów różnych dyscyplin i krajów. Musi być zaprojektowana w taki sposób, aby rozwijać i wzmocnić współpracę i rozpowszechnianie wyników (szczególnie w przypadkach, gdy współpraca przekracza granice poszczególnych dyscyplin), ułatwiać angażowanie specjalistów spoza środowiska naukowego i stwarzać nowe perspektywy i metody badań naukowych.

Praca w środowisku cyfrowym zmieniła sposoby współdziałania i wzajemnych interakcji w nauce. Normą staje się współpraca przekraczająca ramy poszczególnych instytucji, również międzynarodowa; nadal częściej jest ona domeną nauk ścisłych, technicznych czy biologicznych (w tym medycyny). Dyscypliny humanistyczne i społeczne wciąż cechuje indywidualizm w prowadzeniu badań naukowych. Oprócz współpracy interdyscyplinarnej znacznie bardziej niż dotąd niezbędna jest ścisła kooperacja wewnątrz zespołu badawczego. Oczekuje się pełniejszego współdziałania pomiędzy wszystkimi uczestnikami procesów komunikacji naukowej – uczonymi, Technikami, bibliotekarzami, archiwistami, nauczycielami i wydawcami. Tam, gdzie realizowane czynności z natury są rozproszone, organizacja współpracy pomiędzy poszczególnymi ośrodkami i działania kolektywne zawsze dają dobre efekty; środowisko cyfrowe i stosowanie dokumentów elektronicznych zwiększa potrzeby w tym zakresie.

Decyzja o tym, kiedy należy współpracować podczas tworzenia wspólnych rozwiązań, a kiedy współzawodniczyć w przygotowaniu i usprawnianiu określonych działań, nie jest łatwa. Na przykład technologie Web stwarzają wrażenie, że każdy może publikować w Sieci, nieomal nie ponosząc żadnych kosztów. Rzeczywiście, rozpowszechnienie treści w Internecie prawie nic nie kosztuje. Jednak takie indywidualistyczne podejście do kosztów nie odpowiada rzeczywistości: zakłada ono istnienie dostępnej i niedrogiej infrastruktury, za którą zapłacił ktoś wcześniej.

Innowacyjność

Według podanej wcześniej definicji, innowacyjność to wykorzystanie znanych koncepcji, technik i technologii w nowych zastosowaniach. Jej efektem jest powstawanie nowych rozwiązań. Jest to siła, która zmienia odkrycia, wyniki badań i eksperymentów w ich zastosowania i nowe rozwiązania, przez co uzyskują bezpośredni wpływ na formy, wygląd i struktury funkcjonujące we współczesnym świecie [Muraszkiwicz, Jacobfeuerborn 2008, s. 12]. Innowacje powstają obecnie w sieciach łączących organizacje, w tym naukę z biznesem; zależności między nimi mają charakter systemowy [Klincewicz 2008, s. 6]. Informatyczna infrastruktura nauki z jednej strony powinna być stabilna i niezawodna, z drugiej jednak należy przewidzieć możliwość stałej ewolucji, przez podejmowania działań innowacyjnych. Często podkreślana jest rola indywidualnych osób i badane są ich cechy osobowościowe, składające do innowacyjności [Rijnsoever, Donders 2009, s. 986]. Mieczysław Muraszkiwicz i Bruno Jacobfeuerborn zauważają postępującą demokryzację innowacyjności, w sensie wzrostu liczby pracowników, użytkowników i klientów związanych z działalnością proinnowacyjną [Muraszkiwicz, Jacobfeuerborn 2008, s. 15]. Wydaje się, że wzrost wskaźników charakteryzujących naukę, o czym była mowa, tworzenie IIN oraz GBC wpisują się w tę tendencję.

Podjęcie działań innowacyjnych zawsze niesie za sobą zwiększone ryzyko popełnienia błędu i porażki. Jednak również niepowodzenia, byłe właściwie wykorzystane z punktu widzenia informacyjnego, mają pozytywne znaczenie pozwalając na wybór właściwych dróg posuwania się przez nieznanne obszary, gdzie nie sięga nasza wiedza. Z każdym kolejnym krokiem i następnym, rozwiązaniem problemem, nie tylko pojawiają się nowe, nierozwiązane zagadnienia, ale także łatwo odnieść wrażenie, że tam, gdzie do niedawna wydawało się, iż posiadamy ugruntowaną wiedzę, w rzeczywistości wszystko jest płynne i niebezpiecznie nieznanne [Unsworth 1997].

Długotrwałość

Długotrwałość często rozważana jest w kategoriach finansowych: chodzi o zapewnienie dalszego funkcjonowania projektu po wyczerpaniu pierwotnego źródła finansowania. Cyfryzacja nauki spowodowała uwidocznienie mankamentów w finansowaniu badań, komunikacji naukowej i archiwizacji dorobku naukowego, które wcześniej mogły pozostawać niezauważone dzięki sprawnemu funkcjonowaniu tradycyjnych bibliotek i wydawnictw uczelnianych. Zarówno pracownicy nauki, jak i bibliotekarze, muszą odejść od traktowania projektów związanych z cyfryzacją jako krótkotrwałych czy jednorazowych, a zacząć uważać je za długoterminowe programy przeznaczone do stałej realizacji. Warto również wziąć pod uwagę fakt, że obciążenie użytkownika końcowego publikacji pełnymi kosztami funkcjonowania komunikacji naukowej jest tylko jedną, być może nie najbardziej popularną opcją, zapewniającą długotrwałość. W grę wchodzi więc głównie długoterminowe subsydiowanie tej działalności przez instytucje naukowe, rząd czy przedsiębiorstwa komercyjne. Długotrwałość wykracza jednak poza samo poszukiwanie stałego źródła finansowania, gdyż intelektualna długotrwałość wymaga kapitału ludzkiego. Realizacja projektów związanych z IIN angażuje ludzi wykształconych, odpowiednio wyszkolonych i oddanych swojej pracy, w związku z czym instytucje naukowe muszą tworzyć programy i rozwijać możliwości przygotowania ludzi o takiej wiedzy i umiejętnościach.

Współdziałanie

Współdziałaniem (*interoperability*) nazywa się zdolność dwóch lub więcej systemów lub ich części do wykorzystania informacji, pochodzących z wymiany bez potrzeby realizacji dodatkowych prac w każdym z systemów [Nahotko 2004, s. 14]. Zagadnienia związane z zapewnieniem współdziałania serwisów GBC przedstawione zostały w p. 4.6. Istotnym zadaniem IIN jest zapewnienie bezproblemowego dostępu do danych gromadzonych w dowolnych zasobach, co wymaga tworzenia i stosowania standardowych narzędzi i metadanych, pozwalających uzyskać nie tylko właściwy poziom współdziałania, ale także możliwość wielokrotnego wykorzystania zasobów w różnych zastosowaniach i aplikacjach (*reusability*). Dotychczas dostawcy treści udostępniali wielką liczbę zasobów online, jednak zazwyczaj znajdujących się w niedostępnych „kasach pancernych” z pojedynczym wejściem i wyjściem, z możliwością jednokierunkowego wyszukiwania pionowego. Poziome wyszukiwanie danych nie jest możliwe. Taki sposób funkcjonowania systemów udostępniających zasoby, mało użyteczny dla uczonych, w prosty sposób powiela bariery istniejące w analogowym świecie, w którym informacja z natury jest fizycznie rozproszona i trudna do scalenia przez potrzebujących jej użytkowników.

Odpowiednio rozwinięta IIN będzie funkcjonować jak sieć elektryczna: możemy używać najprzeróżniejszych urządzeń elektrycznych do własnych, indywidualnych celów i urządzenia te podłączane do wspólnej sieci funkcjonują bez zarzutu, pod warunkiem, że spełniają ogólnie przyjęte standardy. Podobnie pracownicy nauki otrzymają łatwy dostęp do mocy obliczeniowej, komunikacji i potrzebnych im zasobów informacji, mogąc jednocześnie wybierać różne ścieżki realizacji zainteresowań i stosować różne narzędzia.

Dobro publiczne

Wielu autorów twierdzi, że informacja cyfrowa ma znaczenie dla rozwoju społeczeństw [Welshons red. 2006]. Jednak fakt ten może się w pełni uwidocznić tylko pod warunkiem, że dostęp do zasobów nauki będzie jak najbardziej otwarty (idea otwartej nauki – *open science*), zarówno w sensie ekonomicznym, jak i intelektualnym, dla wszystkich grup użytkowników [Schroeder 2007, s. 4]. Jak dotąd sukces w tym względzie jest połowiczny. W Webie udostępniane są bez żadnych ograniczeń olbrzymie zasoby ludzkiej wiedzy i dorobku, jednak wiele informacji, w tym o najwyższej jakości, nie jest powszechnie dostępnych. W wielu dziedzinach, na przykład w medycynie, coraz większą rolę odgrywają firmy komercyjne, które digitalizują materiały zarówno będące dobrem publicznym, jak i chronione prawem autorskim; zasoby te stają się dostępne tylko dla dużych instytucji, mogących ponosić poważne opłaty licencyjne (np. duże biblioteki naukowe).

Zasadniczą cechą dobra publicznego jest możliwość powiększenia grona jego konsumentów, bez pogorszenia jakości dobra dostępnego dla innych; jest ono niekonkurencyjne [Benkler 2008, s. 51]. Dobra publiczne rzadko występują na rynku, gdyż zasadą jest tam pobieranie opłat za korzystanie z dóbr. Koszty wytworzenia dobra publicznego mogą być wysokie, jednak poziom jego użytkowania można zazwyczaj podnosić bez ponoszenia dodatkowych kosztów; dlatego raz wytworzone dobro publiczne powinno być jak najczęściej wykorzystywane. Produkcja naukowych publikacji cyfrowych jest oczywiście dobrym tego przykładem. Z tego powodu wśród wielu „modeli biznesowych” stosowanych dla bibliotek naukowych rzadko występują takie, które naprawdę można nazwać biznesem, brak bezpośrednich opłat jest bowiem czynnikiem pozwalającym na rozszerzanie zasięgu społecznego oddziaływania bibliotek [Wojciechowski 2006, s. 16]. Z IIN wiąże się nadzieje na organizację komunikacji naukowej dla dostarczania informacji i wiedzy jako dobra publicznego: projekty realizowane w Webie potwierdziły już, że takie postępowanie może pozwolić na wyodrębnienie nowych modeli ekonomicznych, które dają zaskakujące efekty i dostarczają niespodziewanych korzyści.

Jak uważa Elinor Ostrom, wiedza może być uważana za dobro publiczne z punktu

widzenia dwóch przyjętych przez nią kryteriów: trudności wykluczenia oraz niskiej rywalizacji. Trudno jest wykluczyć kogoś z dostępu do wiedzy, która została mu udostępniona. Wykorzystanie wiedzy przez jedną osobę nie ujmuje możliwości wykorzystania tej wiedzy przez inne osoby. Mowa tu jest o ideach, wiedzy i mądrości zawartych w publikacjach, gdyż same publikacje mogą być uważane za dobro prywatne [Hess, Ostrom 2006, s. 9]. W takim sensie dobro publiczne odróżnić można od dóbr wspólnych (*commons*), od których co prawda również trudno jest wykluczyć, ale wykorzystanie ich przez kogoś utrudnia dostęp innym – przykładem są biblioteki tradycyjne. Z tego wynika, że przejście od biblioteki tradycyjnej do serwisów GBC powoduje, że biblioteka z dobra wspólnego staje się dobrem publicznym. Z drugiej strony nowe technologie mogą służyć także do zawładnięcia tym, co wcześniej było dobrem publicznym.

Interesujące rezultaty może przynieść połączenie zasad funkcjonowania IIN z Nuką 2.0, opartą na serwisach Web 2.0 (zob. p. 4.9). W takich przypadkach, tam gdzie to możliwe, do IIN można zaadaptować podejście typowe dla Nauki 2.0. Nie wszystkie na przykład usługi obliczeniowe muszą działać w powiązaniu z silną autoryzacją dostępu, stosowaną w systemach Gridowych, a złagodzenie wymogów tego rodzaju spowoduje, że uczeni zyskają nowe możliwości, wynikające z rozpowszechnienia oddolnych inicjatyw [Pierce i in. 2006, s. 267]. Takie połączenie nauki zinstytucjonalizowanej, zapewniającej dostęp do największych urzędów badawczych, ze swobodnym, oddolnie organizowanym rozpowszechnianiem wyników badań w serwisach społecznościowych, na przykład Wiki lub blogach, jest charakterystyczne dla współczesnej nauki [Wheeler 2008, s. 109].

1.4. Dane, informacja i wiedza w IIN

Każda infrastruktura nauki musi zarządzać danymi. IIN w szczególny sposób dotyczy danych: sposobów ich otrzymywania, rozpowszechniania, przechowywania oraz przetwarzania w produkty końcowe nauki, takie jak wiedza, odkrycia, edukacja, aplikacje.

Dla celów tej publikacji **dane, informacje i wiedza** uznane zostały za elementy kontinuum związanego z procesem naukowym, którego początkiem są „surowe”, nieprzetworzone i nie poddane analizie dane prymarne uzyskane w efekcie badań naukowych [Machlup 1983, s. 641]. Kolejnymi elementami tego kontinuum są informacje wydobywane z danych w wyniku ich analizy, zazwyczaj publikowane w dokumentach oraz wiedza, wynikająca z prezentowanych informacji i mądrość oznaczającą zdolność zdobywania wiedzy, gromadzenia jej i użycia w praktyce, podczas rozwiązywania problemów [Materska 2007, s. 50]. Jako dane traktowane są m.in. wyniki obserwacji cyfrowych, monitoringu naukowego, odczyty z czujników, metadane, wyniki modelowania i scenariusze, dane z obserwacji behawioralnych, wizualizacje, dane statystyczne. Zazwyczaj pojawiają się na początku procesu badawczego, zapisane na jakimś nośniku danych.

Informacje stanowią konkluzje wynikające z analizy, uporządkowania (struktura) i kontekstualizacji danych, przez co ustala się znaczenie danych w kontekście wypowiedzi lub działania. Stanowią one zasadniczą część publikacji naukowych (dokumentów). Jako takie posiadają swoich nadawców i odbiorców, a celem ich przekazywania jest wywieranie wpływu przez nadawcę na świadomość odbiorcy. Informacje reprezentowane są w systemach informacyjnych, w których idee, ludzie, miejsca i obiekty posiadają etykiety³³ organizowane w listy, schematy hierarchiczne, fasety, chmury tagów i inne struktury,

³³ Termin etykieta w tej książce, podobnie jak tag w języku ang., jest stosowany w wielu, choć zbliżonych znaczeniach. Pojawia się wszędzie tam, gdzie mowa jest o nazywaniu, kodowaniu, oznaczaniu pewnych elementów. W tym znaczeniu etykiety pojawiają się w językach kodowania (HTML, XML), strukturach danych (DC,

służące opisowi i uporządkowaniu dokumentów [Borgman 2007, s. 43]. Jak stwierdza Christine Borgman, istnieje wiele definicji informacji, na podstawie których można ją traktować zarówno jako coś wielkiego i amorficznego, jak w teorii kultury, lub też jako coś bardzo małego i dyskretnego, jak w matematycznej teorii komunikacji Shannona i Weavera [Borgman 2003a, s. 58]. Definicje informacji różnią się bardzo w zależności od dyscypliny i kontekstu [Kisilowska 2009, s. 76]. Ta różnorodność wskazuje na niemożność traktowania informacji w jakiś jeden, określony sposób. Od setek lat filozofowie starają się wyjaśniać znaczenie terminów takich, jak „informacja” i związanych z nim terminów „wiedza” i „znaczenie”; opis tych dyskusji wykracza poza ramy tej pracy.

Informacja bywa często traktowana jako składnik wiedzy. Powoduje ona modyfikację struktur zasobów wiedzy, posiadanej przez odbiorcę informacji [Materska 2008, s. 31]. Wiedza, w odróżnieniu od danych i informacji na charakter intuicyjny, co utrudnia jej definiowanie i analizę. Znajduje się ona w ludzkich umysłach, gdzie powstaje przy wykorzystaniu doświadczenia i intuicji, na podstawie uzyskiwanych i rozumianych danych oraz informacji. Jest zjawiskiem społecznym, nie może więc być wprost zapisana w danych, publikacjach czy usługach globalnej biblioteki cyfrowej, które raczej stosowane są do przesyłania i przetwarzania danych i informacji, niż wiedzy [Golka 2008, s. 163]. Możliwy jest zapis jedynie reprezentacji wiedzy (w dokumentach, systemach informacyjnych) lub jej ucieleśnienie w wartościach, metodach i procedurach [Owen 2001]. Wówczas mamy do czynienia z wiedzą jawną (*explicit*). Często jednak wiedza ma charakter ukryty³⁴; z jej istnienia zdajemy sobie sprawę, potrafimy ją wykorzystać, ale bardzo trudno jest stworzyć jej reprezentację, więc jej formalizacja i przekazywanie są bardzo utrudnione. Dla zmiany wiedzy ukrytej w wiedzę jawną wykorzystuje się różne metody opisywania i organizacji obiektów informacyjnych, w efekcie czego informacja jest publikowana i w ten sposób rozpowszechniana³⁵. Cyfryzacja tych procesów także przyczynia się do „ujawniania” wiedzy ukrytej i zwiększenia jej mobilności [Borgman 2007, s. 167]. Wiedza ukryta jest przekazywana słownie, głównie podczas przeżywania wspólnych doświadczeń, na przykład w kontaktach mistrz-uczeń³⁶. Wiedza zależy także od pamięci, a ta także dzielona jest na dwa rodzaje – pamięć wewnętrzną (służącą do zapamiętywania) oraz zewnętrzną – stosowaną do zapisywania wiedzy przekształconej w informację na zewnętrznych nośnikach, gromadzonych między innymi w globalnej bibliotece cyfrowej.

MARC), identyfikatorach (URI). W nieco innym znaczeniu, zbliżonym do słów kluczowych, etykiety występują w serwisach Webu 2.0, głównie Delicious.com. Etykiety bywają w polskiej literaturze nazywane tagami, znacznikami, metkami, przywieszkami.

³⁴ Autorem pojęcia wiedzy ukrytej (*tacit knowledge*) był w latach sześćdziesiątych XX wieku Michael Polanyi, który wiedzę tę opisywał następująco: „wiemy więcej, niż jesteśmy w stanie powiedzieć, a przez to, że znaczenie ma charakter ukryty, nigdy nie możemy znać skutku, jaki odniesie to, co powiemy” [Polanyi 1998, s. 99]. Wiedza ukryta gromadzi się w każdym człowieku, jednak mimo tego jest bardzo trudna do przekazywania.

³⁵ Podział na wiedzę jawną i ukrytą bywa uważany za błędny; jest to raczej odróżnienie „poczucia sensu” od „formalnego wyrażania” wiedzy. Oba są dwoma wymiarami tej samej aktywności, polegającej na odkrywaniu znaczenia. Formalne wyrażanie nie może zastąpić poczucia sensu, a poczucie sensu jest zawsze niezbędne dla wydobycia znaczenia z formalnego wyrażania. Odkrywanie znaczenia jest czynnością typową dla istot ludzkich, stąd wiedza nie może istnieć poza nimi [Andriessen, Boom 2009, s. 399].

³⁶ Opinie takie są negowane przez konstruktystów, którzy uważają, że wiedza nie jest czymś, co można zdobyć przez obcowanie z książkami i nauczycielami. W myśl konstrukttywizmu wiedza jest tworzona przez jednostkę. Człowiek nie rejestruje informacji, lecz buduje struktury wiedzy z dostępnych informacji. Wiedza nie jest więc czymś istniejącym obiektywnie poza człowiekiem, a rzeczywistość nie istnieje niezależnie i oddzielnie od tego, kto ją poznał.

Komunikowanie wiedzy możliwe jest dopiero po jej transformacji na informację³⁷, która może przybierać różne formy³⁸.

Rozróżnienie pomiędzy danymi, informacją i wiedzą nie jest jednoznaczne i może się zmieniać zależnie od sytuacji. Granice między nimi są rozmyte, powodując przechodzenie jednego stanu w drugi [Materska 2007, s. 56]. W szczególności informacje uzyskane w określonych badaniach mogą stać się danymi wykorzystywanymi w kolejnych. Zjawiska te uległy wyraźnemu nasileniu w wyniku stosowania Internetu. W tym środowisku wartość dodaną uzyskuje się poprzez przetwarzanie informacji i wiedzy, gdyż dane i ich zestawy (a więc także zestawy metadanych) mogą składać się na bazy wiedzy lub być z nich wydobywane. Postępowanie jest więc tu prowadzone w kierunku odwrotnym niż w systemach tradycyjnych: od wiedzy ku danym. Gromadzenie wiedzy jawnej i jej organizowanie w postaci ustrukturyzowanej informacji jest jedną z najważniejszych cech i możliwości, jakie daje Web [Mason 2009, s. 43].

Istotą działalności naukowej są procesy polegające na tworzeniu danych i informacji oraz przekształcaniu ich w wiedzę i mądrość. Procesy te są jednak często rozumiane i postrzegane w skrajnie odmienny sposób. Tacy badacze, jak na przykład Warren Weaver³⁹ twierdzą, że nauka służy rozwiązywaniu problemów w oparciu o podstawowe prawa logiki, a badane zjawiska są w zasadzie możliwe do zmierzenia [Weaver 1948, s. 543]. Z tego punktu widzenia badania naukowe powinny być powtarzalne, bezstronne i oparte na faktach przyrodniczych. Pomimo atrakcyjności takiego poglądu na naukę, inni badacze uważają go za nerealistyczny, gdyż nie bierze pod uwagę roli uwarunkowań społecznych działalności naukowej⁴⁰.

Na potrzebę uwzględniania tych uwarunkowań zwracał uwagę Karl Popper [Popper 2002, s. 71]. Twierdził on, że nie ma możliwości sprawdzenia prawdziwości teorii naukowej, można tylko odrzucić teorię jako fałszywą. Zamiast więc weryfikacji, czyli potwierdzenia, zalecał falsyfikację, czyli szukanie sposobów na obalenie danej hipotezy⁴¹. Podważał on przydatność metody indukcyjnej, gdyż przyjęta teoria wpływa na obserwację. Tej metodzie przeciwstawił metodę dedukcyjno-hipotetyczną jako jedynie logicznie uprawnioną. Ponadto stwierdzał, że warunki realizacji obserwacji i eksperymentu istotnie wpływają na akceptację lub odrzucenie ich wyników. Uważał także, że wiedza jest produktem społecznym, powstającym w efekcie modyfikacji wcześniej posiadanej wiedzy, przez co nie jest tworzona w bezpośredniej konfrontacji ze światem fizycznym.

Możliwe są więc co najmniej dwa, krańcowo sprzeczne poglądy na naukę: pierwszy,

³⁷ Istnieją także opinie, że informacja jest rodzajem wiedzy [Vickery 2005]. Jest ona taką interpretacją doświadczeń przez mózg, że powoduje modyfikację wiedzy w umyśle. Informacja jest więc nową lub zmodyfikowaną wiedzą, która została dodana, czy też zasymilowana do osobistej struktury wiedzy. Z drugiej strony Sabina Cisek relacjonując podejście Sense-Making stwierdza, że informacja nie jest absolutnym, niezmiennym, a nawet obiektywnie istniejącym bytem. Nie jest więc bezwzględnie prawdziwa czy fałszywa, stanowi bowiem tylko odzwierciedlenie subiektywnych opinii ekspertów o rzeczywistości. Z punktu widzenia użytkownika informacja stanowi subiektywną konstrukcję, tzn. wszystko, co informuje użytkownika jest informacją [Cisek 2008, s. 99].

³⁸ Formy te to kodyfikacja: tworzenia kategorii i zasad pozwalających zastosować wiedzę w konkretnym przypadku; abstrakcja: tworzenia struktur oraz konstrukcji logicznych, pozwalających na uogólnienie wiedzy, która dzięki temu może być stosowana w różnych, ale powiązanych sytuacjach; zapis: tworzenia symbolicznej reprezentacji wiedzy skodyfikowanej lub abstrakcyjnej; dyfuzja: komunikowania zapisanej wiedzy innym.

³⁹ Warren Weaver (1894-1978) był współtwórcą ogólnego modelu komunikacji Shannona-Weavera.

⁴⁰ Trzeba jednak przyznać, że sam Weaver także widział pewne uwarunkowania społeczne, twierdząc, że metoda naukowa spełni pokładane w niej oczekiwania, jeżeli osoby posługujące się nią odznaczać się będą szczególną uczciwością, otwartością umysłu oraz umiłowaniem prawdy.

⁴¹ Na podstawie metody falsyfikacji hipotez za naukowe można uznać tylko hipotezy posiadające empirycznie falsyfikowalne konsekwencje. W skład nauki nie może wchodzić hipoteza, która niczego nie zabrania i daje pogodzić się z wszystkimi możliwymi faktami. Jako przykłady takich nienaukowych hipotez Popper podaje teorię Zygmunta Freuda oraz marksizm [Szumilewicz-Lachman 1989, s. IX].

według którego podczas procesu badawczego wszystko można wyjaśnić i zrozumieć; drugi, zgodnie z którym nic właściwie nie może być przez naukę udowodnione, gdyż jest ona „produktem społecznym”. Jak to często bywa, najbardziej poprawne wydaje się stanowisko pośrednie, pomiędzy tymi dwoma, gdyż nauka dąży do przedstawiania faktów, kreowania i udowadniania obiektywnego wizerunku praw rzeczywistych, nawet, jeżeli ograniczana jest kontekstem historycznym i społecznym, wpływającym na wyniki badań naukowych [Chalmers 1997, s. 209]. Thomas Kuhn, dla wyjaśnienia, czym jest nauka, przedstawił kilka jej najbardziej istotnych cech:

- Nauka jako tworzenie wiedzy. Działalność naukowa jest charakteryzowana przede wszystkim przez jej produkty [Kuhn 2001, s. 47].
- Nauka jako proces akumulacji. Nauka jest procesem polegającym na ciągłym ulepszaniu lub wzroście, którego celem jest tworzenie nowych teorii służących przewidywaniu nowych faktów⁴².
- Nauka jako proces współpracy. Tworzenie nowej teorii naukowej rzadko udaje się zakończyć pojedynczemu uczonemu [Kuhn 2001, s. 297].

Informatyczna infrastruktura nauki, którą opisałem w poprzednich rozdziałach, stanowi sprawny i różnicowany, globalny system zarządzania i udostępniania danych i informacji cyfrowych, służących rozwojowi wiedzy. Funkcjonowanie IIN powoduje, że badania naukowe i edukacja na poziomie wyższym stały się dziedzinami znacznie bardziej „danochłonnyimi”. Wymagają znacznych ilości danych, powstających w wyniku działania kolejnych generacji technologii cyfrowych, instrumentów oraz sieci rozległych, za pomocą których dane są tworzone, gromadzone, rozpowszechniane i analizowane. Pracownicy nauki codziennie produkują dane, uzyskują do nich dostęp, analizują je, integrują i przechowują terabajty danych cyfrowych, pozyskiwanych poprzez eksperymenty, obserwacje i symulacje. Według badań Petera Lymana i Hala Variana ilość nowej informacji, zapisanej na papierze, filmie, nośnikach magnetycznych i optycznych podwoiła się między latami 1999 a 2002; wzrost osiągnął wskaźnik ok. 30% rocznie [Lyman, Varian 2003]. Autorzy oszacowali, że w 2002 r. wyprodukowano 5 eksabajtów⁴³ informacji, z czego 92% zapisano na nośnikach magnetycznych⁴⁴. Według innych szacunków w 2006 r. powstało już 180 eksabajtów danych, a prognozy mówią o 10-krotnym wzroście rocznej produkcji informacji cyfrowej w 2011 r. [Gantz i in. 2008]. Wskaźnik ten jest miarodajny dla skali wzrostu danych i dokumentów cyfrowych. Dynamiczna integracja danych, generowanych w wyniku prowadzonych obserwacji i symulacji pozwala na rozwój nowych metod badawczych, dostosowywanych do zmieniających się warunków realizacji badań w celu tworzenia nowej wiedzy. Jest to sprzężenie zwrotne, gdyż szybki wzrost ilości dostępnych i wykorzystywanych w nauce danych, powoduje dalszy rozwój badań naukowych, a także przyspieszenie transformacji wyników badań w dostępne produkty i usługi oraz podnosi efektywność nauczania.

W dynamicznym środowisku badawczym i dydaktycznym dane naukowe są stale tworzone, gromadzone, zapisywane, udostępniane, analizowane i przetwarzane w informacje, w szalonym wyścigu do nowej wiedzy. W przyszłości przywództwo w zakresie badań naukowych i zgromadzonej wiedzy w coraz większym stopniu zależęć będzie od umiejętności i zdolności zapewnienia dostępu do zasobów danych naukowych w formie cyfrowej i transformacji tych danych w informację i wiedzę przy pomocy skomplikowanych

⁴² Thomas Kuhn wskazywał na istnienie okresowych rewolucji naukowych, które charakteryzują się brakiem akumulacji wiedzy. Podczas rewolucji naukowych stary paradygmat zastępowany jest w całości lub części przez nowy. Rewolucje naukowe różnią się od normalnej nauki, która ma charakter akumulacyjny [Kuhn 2001, s. 53].

⁴³ Jak podają autorzy, 5 eksabajtów (jeden eksabajt to miliard miliardów znaków) informacji jest to zasób, który pomieściłby 500 tys. bibliotek wielkości Library of Congress.

⁴⁴ Na papierze zapisane zostało 0,01% informacji.

procesów eksploracji danych⁴⁵ (*data mining*) i stosowania narzędzi służących do ich integracji, analizy i wizualizacji. Tworzenie powiązań między danymi, ukazujących bogate relacje między nimi, pozwalających na ich wielokrotne wykorzystanie i integrację, wymaga znajomości i udokumentowania proveniencji danych, budowy ontologii, tworzenia opracowań i analiz eksperckich. Działania te umożliwią oferowanie usług polegających na tworzeniu wizualizacji, symulacji oraz eksploracji danych i ich modelowaniu, a także innych form reprezentacji i wydobywania wiedzy.

Dane, tworzone z punktu widzenia potrzeb, praktyki i przyzwyczajzeń jednej dyscypliny, mogą być trudne do automatycznego skonwertowania do użytku w innej, nawet pokrewnej [Borgman 2008, s. 31]. Suzanne Thorin pisze wręcz o odrębnych „plemionach” uczonych, zajmujących wybrane dyscypliny jako własne terytoria [Thorin 2003]. Uczeni z jednej strony muszą mieć pewność, że dane, udostępniane przez nich, będą wykorzystywane w sposób właściwy i nie szkodzący interesom ich twórcy, np. w zakresie pierwszeństwa, autorstwa publikacji. Z drugiej strony, pracownik nauki, otrzymujący dane powinien mieć pewność co do ich wiarygodności, szczególnie w sytuacji, gdy wzrasta liczba źródeł danych a wiedza o ich reputacji maleje. W niektórych dyscyplinach z długą tradycją współpracy, jej zasady mogą być od dawna wypracowane. W innych sama potrzeba współpracy może nie być tak oczywista.

W najbliższej przyszłości nastąpi gwałtowny wzrost objętości dostępnych danych, co bywa nazywane „zalewem danych” (*data deluge*). Podstawowe znaczenie dla właściwego wykorzystania tego nadmiaru będzie miał rozwój nowoczesnych narzędzi wyszukiwawczych i związanych z nimi technologii, które muszą zapewnić wzrost aktywności przeszukiwania. Należy zadbać o rozwój następujących obszarów: eksploracja danych i tekstu, wyszukiwanie dokumentów nietekstowych (ruchome obrazy, dźwięk, wyszukiwanie obrazów według ich treści), techniki indeksowe dźwięku, narzędzia dla opisu proveniencji i historii obiektów oraz ich rankingu. Spowoduje to połączenie umiejętności typowych dla informatyki, z tradycyjnymi technikami stosowanymi w bibliotekarstwie i informatyce naukowej. Potrzeba umiejętnego sterowania zalewem danych, poprzez stosowanie zintegrowanego gromadzenia, przetwarzania, przechowywania i udostępniania danych, wskazuje na zasadniczą rolę, jaką odgrywa organizacja i archiwizacja danych (zob. p. 2.6), stanowiąca od zawsze domenę bibliotekarzy.

Wzrost zainteresowania potencjalnymi efektami, wynikającymi z wdrożenia zaawansowanej IIN powoduje projekt utworzenia **Gridu**, infrastruktury informatycznej wolnej od technicznych niedoskonałości obecnego Internetu, takich jak na przykład zawodne połączenia, ograniczona i nierównomiernie rozprowadzona sieć, wrażliwość komputerów na ataki z zewnątrz i samopowielające się, złośliwe wirusy. Użytkownicy komputerowego Gridu, podobnie jak sieci (*grid*) elektrycznej, będą mogli podłączyć się do jakiegokolwiek aplikacji informacyjnej według potrzeb, w każdym miejscu i czasie; będą mieli do stałej dyspozycji moc obliczeniową Gridu, udostępniane dane i narzędzia – wszystko to bez potrzeby znajomości, a nawet wiedzy o istnieniu podstawowej architektury, służącej do lokalizowania i dostarczania tych zasobów [Foster 2003]. Gridem nazywa się także oprogramowanie służące gromadzeniu niewykorzystanej mocy obliczeniowej sieci. Wizja ta, przedstawiająca możliwości bezproblemowego dostępu do wszechobecnych czy „przezroczystych” zasobów komputerowych, bywa uznawana za cokolwiek utopijną. Jednak

⁴⁵ Eksploracja danych (nazywana też wydobywaniem danych, drążeniem danych, ekstrakcją danych czy poszukiwaniem wiedzy) jest jednym z etapów odkrywania wiedzy w bazach danych. Idea eksploracji danych polega na wykorzystaniu szybkości przetwarzania komputerowego do znajdowania ukrytych dla człowieka (z powodu jego ograniczonych możliwości przetwarzania) prawidłowości w danych zgromadzonych w hurtowniach danych. Eksploracja danych dobiera się z wykorzystaniem takich technik jak: wizualizacja danych, metody statystyczne, sieci neuronowe, algorytmy genetyczne, uczenie maszynowe, logika rozmyta czy zbiory przybliżone. Część z nich stanowi metody sztucznej inteligencji [Wojanik 2007, s. 239].

w ten sposób często bywają określane nowe koncepcje rozwiązań technicznych, żeby wspomnieć Semantyczny Web (zob. p. 4.9). W analogii do tego ostatniego mówi się także o Semantycznym Gridzie, który od Gridu różni się tym, czym Semantyczny Web od Webu [Roure, Jennings, Shadbolt 2001].

Celem projektantów Gridu jest utworzenie współdziałającego, powszechnego, niezawodnego i nieograniczonego dostępu do zasobów obliczeniowych, udostępnianych przez komputery w sieci⁴⁶. Grid nie jest kolejną aplikacją dostępną poprzez Internet. Jest to raczej rodzaj systemu operacyjnego dla Internetu. Jednym z jego elementów jest middleware, oprogramowanie niezależne od zróżnicowanego sprzętu komputerowego, dzięki któremu aplikacje mogą pomijać to hardwareową różnorodność, przez co tworzenie wspomnianych aplikacji staje się łatwiejsze.

Usługi Web mogą być uważane za pierwszy etap na drodze od Internetu do Gridu. Terminem tym obejmuje się wszystkie wysiłki podejmowane w celu zwiększenia kompatybilności i standaryzacji, służące osiągnięciu współdziałania poszczególnych sieci komputerowych. Usługa jest definiowana jako działająca w sieci jednostka, która dostarcza pewnych funkcjonalności, takich jak obliczenia, przechowywanie danych, programy aplikacyjne dla symulacji, przetwarzanie transakcji⁴⁷ [David 2004]. Usługi Gridowe są usługami Web dostarczającymi użytkownikom odpowiednich interfejsów i funkcjonującymi zgodnie z odpowiednimi protokołami (zgodnymi z interfejsami).

Jednym z celów IIN jest tworzenie platform softwareowych, pozwalających na pracochłonne przetwarzanie danych i informacji na potrzeby przestrzennie i instytucjonalnie rozproszonej nauki i techniki. Według nieco dowolnej interpretacji, IIN zawiera elementy, związane z tworzeniem Gridu, realizowanym przez instytucje nauki i techniki (osoby i zespoły) lub przez **kolaboratoria**. Według bardziej ograniczonej interpretacji, IIN obejmuje połączenie badań nad Gridem i kolaboratoriami.

Te ostatnie zyskują coraz większe znaczenie. Kolaboratoria definiowane są jako laboratoria zawierające i łączące segmenty „wirtualnej rzeczywistości”: pracowników nauki z przyrządami badawczymi oraz dane, rozlokowane w sposób rozproszony w przestrzeni – mogących współpracować ze sobą w sposób interaktywny, w czasie rzeczywistym. Kolaboratoria skracają fizyczny dystans między uczonymi, wspomagają współpracę między specjalistami w zakresie obszarów wspólnego zainteresowania oraz pozwalają na udostępnianie danych, obiektów i narzędzi niezbędnych do realizacji zadań badawczych. Zasadnicze znaczenie dla rozwoju rosnącej zbiorowości uczonych i projektów, głównie w zakresie zaspokojenia potrzeb informacyjnych mają nowe rodzaje organizacji nauki i otoczenia wspomagającego („laboratoria bez ścian”). Są tworzone pod różnymi nazwami, jak kolaboratoria (*collaboratory, co-laboratory*), wirtualne laboratoria (*virtual laboratory*), społeczność Gridowa (*grid community*), społeczność e-nauki (*e-science community*), społeczność wirtualna (*virtual community*). angielska nazwa *collaboratory* powstała z połączenia przez Williama Wulfa w 1999 r. słów *collaboration* i *laboratory* [Kling, McKim, King 2001].

Kolaboratoria są cyfrowym odpowiednikiem „niewidzialnych uniwersytetów” (*invisible colleges*), których koncepcja pochodzi z XVII w., gdy powstało The Royal Society, skupiające uczonych, działających na niezbyt rozległym obszarze i posiadających podobne zainteresowania, ale jednak nie skupionych w jednej instytucji. Pisał o nich także Derek de Solla Price, mając na myśli nieformalne grupy uczonych z wielu instytucji, także rozproszonych (czasem znaczenie) terytorialnie [Price 1965]. Już w 1994 r. John Gresham

⁴⁶ W ten sposób realizowana jest idea tzw. metakomputera. Termin ten oznacza zasoby obliczeniowe transparentnie dostępne dla użytkowników w środowisku sieciowym. Jest to sieć heterogenicznych zasobów komputerowych powiązanych za pomocą oprogramowania w taki sposób, że użytkownik może nie dostrzegać różnicy między pracą z zasobami lokalnymi i metakomputerem [Smarr, Catlett 2003, s. 825].

⁴⁷ Przykładem usługi tego typu może być bankowość internetowa lub biblioteki cyfrowe.

zauważył przejście od tradycyjnego niewidzialnego uniwersytetu do czegoś, co nazwał „uniwersytetem w przestrzeni cybernetycznej” (*cyberspace college*). Twierdził, że jest to nowa forma nieformalnej sieci badawczej, jednak nie przewidywał, że zastąpi niewidzialne uniwersytety; uważał, że obie formy istnieć będą obok siebie [Gresham 1994].

Początkowo funkcje kolaboratorów ograniczone były do zadań wykonywanych przez laboratoria naukowe. Ich działalność wiązano ze zdalnie sterowanymi i kontrolowanymi instrumentami oraz przyrządami wirtualnymi. Umożliwiają one dostęp do wyposażenia i baz danych, które tradycyjnie zaliczane są do laboratoryjnej organizacji nauki oraz do kolegów i współpracowników, bez względu na oddalenie geograficzne. Michael Nentwich przytacza kilka definicji kolaboratorium; m.in. mówi, że jest to połączenie technologii, narzędzi i infrastruktury, pozwalające uczonym pracować z odległymi urządzeniami i sobą nawzajem tak, jakby byli oni w jednym miejscu [Nentwich 2003, s. 188]. Wirtualne środowisko kolaboratorium pozwala na współpracę każdego rodzaju – ludzi z ludźmi, ludzi z maszynami i maszyn z maszynami. Współpraca ta polega na wymianie danych i informacji, w wyniku czego powstaje środowisko wiedzy.

Tam, gdzie nauka nie jest zorganizowana w sposób laboratoryjny, mówi się o wirtualnych instytutach, podobnych do kolaboratoriów w naukach technicznych i przyrodniczych. Służą one także instytucjonalizacji współpracy na odległość. Wirtualne instytuty również zapewniają dostęp do infrastruktury informatycznej pozwalającej na wymianę danych i wirtualne spotkania.

Kolaboratoria powodują nie tylko rozwój współpracy (wzrost liczby wspólnych projektów), ale także zmiany w sposobach jej realizacji. Nowe wzorce współpracy obejmują w szczególności:

- Rozwój sieci kontaktów osobistych: wzrasta liczba osób, z którymi uczone może współpracować, czasem nawet nie znając współpracowników osobiście.
- Powiększanie się grup współpracujących pracowników nauki: nowe narzędzia stwarzają możliwości organizowania znacznie większych zespołów współpracowników. Niezbędne jest zapewnienie im nie tylko technicznych możliwości komunikacji, takich jak sprzęt, sieci i dane, ale także uwzględnienie zjawisk społecznych, takich jak odmienne tradycje, sposoby i priorytety pracy, czy ograniczenia budżetowe.
- Łatwiejszą kontynuację współpracy: dzięki nowym narzędziom (np. poczcie elektronicznej) łatwiej jest utrzymywać kontakty nawet po przeniesieniu się współpracowników w inne miejsce pracy, a nawet do innego kraju.
- Lepsze wykorzystanie kompetencji: dzięki kolaboratoriom dobór współpracowników do zespołów badawczych nie musi opierać się na kryteriach geograficznych czy personalnych, główną rolę odgrywają ich kompetencje.
- Specjalizację: możliwość łatwego włączenia się uczonych do współpracy na skalę światową może umocnić tendencje do dalszego wzrostu specjalizacji, gdyż w cenie będą bardzo specyficzne umiejętności i doświadczenia, trudne do uzyskania lokalnie.
- Nowe formy współpracy: współpraca w kolaboratoriach może przyjmować formy działalności polegającej na tworzeniu wspólnych zasobów danych i baz wiedzy.
- Standaryzację sposobów pracy: oprogramowanie wykorzystywane grupowo może prowadzić do standaryzacji sposobów pracy. Technologia (oprogramowanie do pracy grupowej, interfejsy baz danych) będzie wymuszać na użytkownikach akceptację tych samych sposobów pracy, tzn. postępowania według tych samych wzorców, wykonywania tych samych czynności w tej samej kolejności. Z jednej strony może to mieć pozytywny wpływ na badania, z drugiej może ograniczać kreatywność.
- Intensyfikację procesów komunikowania: o ile tradycyjne środki komunikacji były stosunkowo mało poręczne i nieefektywne (powolne i mało symultaniczne), nowe środki cybernetyczne są łatwe w użyciu, co może zwiększyć częstotliwość komunikowania się pomiędzy rozproszonymi współpracownikami kolaboratoriów.

Michael Nentwich, w związku z funkcjonowaniem kolaboratoriów wskazuje na kilka

zjawisk, zarówno pozytywnych, jak i negatywnych, takich jak możliwości rozwoju współpracy, powstawanie nowych wzorów współpracy, ale jednocześnie wzrost izolacji spowodowany ograniczeniem bezpośrednich kontaktów interpersonalnych [Nentwich 2003, s. 203 i nast.]. Oprócz ewidentnych korzyści, wynikających z rozwoju współpracy, kolaratoria (paradoksalnie) mogą także prowadzić do większej izolacji pracowników nauki, zasiadających codziennie przed ekranem komputera. Samotna praca przed monitorem komputerowym jako „partnerem komunikacyjnym” w długim okresie może prowadzić do izolacji. Na szczęście technologia komputerowa umożliwia nie tylko komunikację typu człowiek-maszyna (np. wyszukiwanie w bazach danych), ale także człowiek-człowiek, tzn. z innymi badaczami, jakkolwiek nie bezpośrednio. Przy tym, jak już wspomniałem, rozmiary komunikacji między pracownikami nauki (przynajmniej w formie elektronicznej) powiększają się, więc izolacja (w sensie zmniejszonych kontaktów i mniejszych rozmiarów komunikacji z kolegami uczonymi) jest mało prawdopodobna. Potwierdzają to badania wskazujące, że nowe techniki komunikacji nie odłączają ludzi od wspólnot. Komunikacja w Sieci poszerza istniejące wspólnoty, ułatwiając rozwój kontaktów tam, gdzie wcześniej nie istniały [Barney 2008, s. 190].

1.5. Globalna biblioteka cyfrowa jako część IIN

Nowe zasady organizacji nauki, których przejawem jest IIN, wymagają także nowego podejścia do gromadzenia, opracowania, rozpowszechniania i archiwizowania danych oraz informacji naukowej, a więc do funkcjonowania komunikacji naukowej. Szansą wyjścia z kryzysu komunikacji naukowej, o którym pisałem w poprzednich rozdziałach tej książki, jest rozwój mediów elektronicznych i ich zastosowanie w procesach komunikacji w nauce. Cyfrowa komunikacja naukowa realizowana jest na różnych zasadach ekonomicznych; często nadzieje wiąże się z rozpowszechnieniem w niej zasad Open Access, w różnych jego odmianach, także tych propagowanych przez wydawnictwa komercyjne. W tym kontekście mówi się o publikacjach elektronicznych i bibliotece cyfrowej.

Termin „biblioteka cyfrowa”, podobnie zresztą jak termin „biblioteka”, można rozumieć rozmaicie⁴⁸. Może to być instytucja lub kilka (wiele) instytucji, bądź też zespół rozwiązań technologicznych, służących digitalizacji. Biblioteki cyfrowe można także potraktować jako nowy sposób realizacji komunikacji naukowej i rozprzestrzeniania wiedzy, poprzez jej globalne gromadzenie, organizowanie, informowanie o niej i udostępnianie. Wówczas rozumienie biblioteki cyfrowej wybiega daleko poza procesy polegające na digitalizacji publikacji drukowanych, nawet jeżeli towarzyszą jej pewne działania organizacyjne, służące uzupełnieniu dotychczasowych zasobów ich wersjami elektronicznymi. Także fakt współpracy i wspólnej digitalizacji, prowadzonej przez kilka instytucji nie powoduje zmiany jakościowej. Jeżeli biblioteka umieszcza w sieci zasoby cyfrowe, to przestaje być biblioteką lokalną (zinstytucjonalizowaną), a staje się częścią globalnego obiegu informacji. W wyniku bowiem realizacji wymienionych funkcji, typowych przecież dla wszystkich bibliotek, ale w cyfrowym środowisku sieci rozległych, powstaje nowa jakość, którą nazwać można globalną biblioteką cyfrową⁴⁹.

⁴⁸ Zob. na przykład definicje przedstawione przez Marka Nahotko [Nahotko 2004, s. 37] oraz Dariusza Grygrowskiego [Grygrowski 2001, s. 159].

⁴⁹ Jak stwierdza Wanda Pindłowa należałoby raczej w tym przypadku mówić bardziej o uniwersalizmie niż globalizmie, a więc o uniwersalnej bibliotece cyfrowej [Pindłowa 2005, s. 45]. Uniwersalizm to dążenie do powszechności danego poglądu, do objęcia działalnością wszystkich ludzi czy ogarnięcia pewnej całości. Zastępowanie uniwersalizmu globalizmem jest mylące. Uniwersalizm w ostatnim czasie został usystematyzowany jako metafizyka, czyli teoria i perspektywa syntetyzująca wszystkie nauki, ważne kierunki filozoficzne, kultury narodowe i kontynentalne. Mimo pokusy użycia słowa „uniwersalna” w odniesieniu do biblioteki cyfrowej pozostałem jednak przy słowie „globalna” ze względu na jego znaczne rozpowszechnienie w literaturze.

Część definicji bibliotek cyfrowych, pomimo tego, że dotyczyły jeszcze zinstytucjonalizowanych bibliotek cyfrowych, zawiera elementy dobrze przystające do koncepcji globalnej biblioteki cyfrowej. Przykładem może być dwuczłonowa definicja, stworzona przez wieloosobowy zespół specjalistów w 1996 r. [Borgman i in. 1996]. Niewielka modyfikacja tej definicji wystarczy do przystosowania jej do potrzeb globalnej biblioteki cyfrowej. Wówczas globalna biblioteka cyfrowa to:

1. Zasoby elektroniczne i związane z nimi możliwości techniczne służące tworzeniu, wyszukiwaniu i wykorzystywaniu informacji naukowej. W tym znaczeniu jest to rozwinięcie i udoskonalenie systemów wyszukiwania informacji naukowej, pozwalające na manipulację danymi cyfrowymi w każdym formacie (tekst, multimedia, symulacje) funkcjonującymi w rozproszonym środowisku sieciowym. Globalna biblioteka cyfrowa zawiera dane naukowe (prymarne i przetworzone) oraz różnego rodzaju metadane (w tym opisowe i strukturalne, m.in. metadane w OPAC bibliotek naukowych, będące „przełącznikiem” między zasobami drukowanymi i cyfrowymi).

2. Zbiór serwisów, tworzonych przez lokalne społeczności uczonych (stanowiących łączną globalną społeczność użytkowników) i świadczących dla nich usługi informacyjne. Realizowane funkcje służą zaspokajaniu potrzeb informacyjnych tej społeczności. Jest ona częścią tej społeczności, w której poszczególne osoby, ich grupy i społeczności lokalne współdziałają ze sobą w trakcie wykorzystywania serwisów globalnej biblioteki cyfrowej, usług, zasobów i systemów danych, informacji i wiedzy. Globalna biblioteka cyfrowa jest rozwinięciem i integracją wielu serwisów, zarządzanych przez instytucje i osoby prywatne, które tworzą, selekcionują, gromadzą, organizują, archiwizują i udostępniają zasoby dla nauki, służąc realizacji potrzeb użytkowników⁵⁰.

Definicja ta łączy dwie uzupełniające się idee. Pierwszy jej człon akcentuje fakt, że GBC jest kontynuacją i rozwinięciem istniejących systemów informacyjno-wyszukiwawczych, funkcjonujących w obszarze nauki, zawierających wszelkiego rodzaju dane i metadane; w drugiej jej części podkreślono, że praktyka GBC powinna uwzględniać kontekst społeczny, w którym funkcjonuje. Wskazano tu na konwergencję informacji i technologii komunikacyjnych, funkcjonującą w sieciach globalnych i globalnej IIN.

Tak rozumiana GBC, w odróżnieniu od biblioteki tradycyjnej, może być utożsamiana z całym systemem komunikacji naukowej, w tym z jego częściami służącymi komunikacji formalnej i nieformalnej. Globalna biblioteka cyfrowa nie jest bowiem biblioteką naukową w tradycyjnym sensie, ale raczej globalną organizacją, infrastrukturą serwisów sieciowych, w których pracownicy nauki i badacze, odgrywając różne role i postępując zgodnie ze skomplikowanymi wzorcami współdziałania, stosują nowoczesne i zaawansowane technologie do tworzenia i rozpowszechniania informacji w Sieci [Fuhr i in. 2007, s. 24]. Jest składnikiem globalnej informatycznej infrastruktury nauki, rozumianej jako struktura, w której sieci komunikacyjne dostarczają wysokiej jakości usług, służących komunikowaniu się ludzi oraz udostępnianiu informacji naukowej. Właśnie GBC jest tą częścią IIN, która służy zapewnieniu dostępu do informacji naukowej⁵¹. Rozpowszechniana informacja może stanowić opis wyników badań, może jednak zawierać także materiały źródłowe, surowe dane naukowe lub niepublikowane wyniki bieżących badań. Taka infrastruktura posiada swoją strukturę (architekturę). Szczególnie istotną cechą architektury globalnej biblioteki cyfrowej jest jej otwartość, pozwalająca wszystkim na wzajemne

⁵⁰ Autorzy definicji z 1996 r. w jej dalszej części wymieniają różne instytucje, których biblioteka cyfrowa jest rozwinięciem i integracją: biblioteki tradycyjne, muzea, archiwa, szkoły, biura, pracownie uczonych, mieszkania, miejsca publiczne. Jeżeli dodamy do nich wydawców (oryginalna definicja nie przewidywała tworzenia danych w bibliotece cyfrowej), to uzyskamy wyliczenie większości serwisów globalnej biblioteki cyfrowej. Wyliczenie to w mojej definicji zostało pominięte ze względu na malejącą rolę zinstytucjonalizowania globalnej biblioteki cyfrowej.

⁵¹ Dlatego też wytyczenie granicy pomiędzy GBC a IIN wymaga właściwego zdefiniowania pojęcia „informacja”, czego dokonałem w p. 2.3.

połączenia i wymianę danych [Borgman 2003a, s. 26]. Zasadnicze znaczenie mają tam struktury hipertekstu, rozumianego jako medium dla formułowania i komunikowania myśli [Conkling 1988, s. 454].

Globalna biblioteka cyfrowa jako globalna infrastruktura, służy rozpowszechnianiu wiedzy, będąc siecią uczonych, którzy tworzą i udostępniają informacje w formie cyfrowych obiektów informacyjnych. Przez analogię do IIN, stanowiącą pewien abstrakcyjny konstrukt, na który składają się konkretne wyspecjalizowane systemy (o czym była mowa w p. 1.3), również GBC, rozpatrywana na najwyższym poziomie, jest abstrakcyjną bazą, na której budowane są współdziałające systemy i serwisy, z których każdy jest projektowany, tworzony i zarządzany lokalnie (na niższych poziomach GBC).

Niezbędne jest odróżnienie serwisu GBC od realizowanych przez niego usług, gdyż w terminologii stosowanej przeze mnie oba te terminy oznaczają co innego. Ten pierwszy jest miejscem w Internecie, w którym znajdują się zasoby obiektów cyfrowych i wykonywane są funkcje wymienione w drugiej części definicji globalnej biblioteki cyfrowej. Usługi natomiast rozumiane są podobnie jak w zastosowaniach Gridu; według Roberta Kahna i Roberta Wilensky'ego są to dowolne funkcje, które mogą być wykonywane na lub z jednym bądź wieloma obiektami cyfrowymi znajdującymi się w serwisie GBC [Kahn, Wilensky 1995]. Zarówno więc umieszczanie, udostępnianie obiektu cyfrowego w serwisie GBC (na przykład w repozytorium), jak i tworzenie i wysyłanie zapytań oraz wyszukiwanie, są przykładami usług oferowanych przez ten serwis. Realizacja usług w serwisie GBC zazwyczaj tworzy wartość dodaną do obiektu cyfrowego. Serwisy GBC przejmują wiele funkcji (a więc i usług) od tradycyjnych bibliotek, muzeów i archiwów, często łącząc funkcje wielu tradycyjnych instytucji w jednym serwisie; z tego punktu widzenia serwisy, takie jak na przykład wyszukiwarki Web mogą być traktowane jako nowe wcielenie bibliografii [Haider, Sundin 2010]. Jeffrey Pomerantz dzieli usługi⁵² na dwie grupy:

- Usługi techniczne, które, jako funkcje realizowane na dokumentach, zmieniają stan lub warunki ich stosowania. Funkcje te są niezbędne dla działania serwisu GBC. Usługi te mogą być jednak dla użytkowników niezauważalne, gdyż korzystają z nich tylko pośrednio.
- Usługi użytkowników – te są funkcjami realizowanymi na dokumentach, w wyniku czego stan albo warunki stosowania dokumentu zmieniają się tylko pośrednio lub pozostają niezmiennie. W wyniku realizacji tych usług zmienia się natomiast stan lub warunki funkcjonowania użytkownika [Pomerantz 2008]. Przykładem może być wyszukiwanie, lokalizacja, zamawianie i dostarczanie [NISO Web Services and Practices WG 2006, s. 10-11].

Sama GBC, jako złożona infrastruktura jest raczej wynikiem specyficznych procesów historycznych, współzależności, innowacji i łączenia w całość mniejszych części na zasadzie kołażu. Nie należy myśleć o niej jako o maszynie czy systemie, który można zbudować czy zaprojektować. Takie myślenie umniejsza znaczenie rozlicznych kwestii natury społecznej, organizacyjnej, prawnej, kulturowej i innych nietechnicznych, które muszą być rozwiązywane w sposób rozproszony, ale spójny. Paul Edwards przestrzega przed próbami ścisłej kontroli struktur organizacyjnych, powstających dzięki możliwościom technologicznym: nawet, jeżeli takie pomysły wyglądają interesująco, to takie struktury po prostu nie działają [Edwards i in. 2007, s. 6].

⁵² Czasem odróżnienie serwisu GBC od realizowanych przez niego usług jest trudne. Tak jest na przykład w przypadku systemów społecznościowych, funkcjonujące w obrębie Web 2.0, takich jak Flickr, który jest serwisem (usługą?) służącym do gromadzenia, kategoryzacji i udostępniania fotografii cyfrowych. Jeffrey Pomerantz uważa, że Flickr nie jest serwisem GBC, ale raczej usługą wspomagającą użytkowników w organizowaniu prywatnych kolekcji fotografii, umożliwiającą tworzenie dodatkowego poziomu usług przez dodawanie etykiet do fotografii innych użytkowników [Pomerantz 2008].

W globalnej bibliotece cyfrowej obiekty tworzone i zarządzane (publikowane, archiwizowane) są lokalnie (choć często w wyniku współpracy w skali ponadlokalnej, czasem wręcz globalnej), natomiast udostępnia się je globalnie, bez ograniczeń czasoprzestrzennych, chociaż z ograniczeniami ekonomiczno-prawnymi: część zasobów jest udostępniana odpłatnie. Takie zderzenie zasobów tworzonych lokalnie i udostępnianych globalnie oraz masowych zasobów zewnętrznych, funkcjonujących w globalnej sieci, nie musi być destruktywne, a w nauce, która ma charakter globalny powinno być wręcz pożyteczne. Zgodnie z opiniami przedstawionymi przez Jacka Wojciechowskiego, treści globalne oraz lokalne nie muszą się nawzajem eliminować, bo mogą podlegać wspólnej asymilacji. Miejscowy kontekst oraz lokalna świadomość stanowią wówczas układ odniesienia, według którego przyswajane są treści zewnętrzne. Następuje konwergencja treści, poprzez dodanie do treści lokalnych przetworzonych składników zewnętrznych. Oba obszary komunikacji, lokalny i globalny, koegzystują ze sobą bez potrzeby wzajemnego konkurowania. Technologie sieciowe pozwalają na masowe dostosowywanie się do indywidualnych potrzeb [Barney 2008, s. 82]. Zjawisko to nazywane jest globalizacją [Wojciechowski 2008b, s. 231-232] lub fragmetacją (*fragmentation + integration*) [Bard, Söderqvist 2006, s. 173]. W obrębie GBC odbywa się komunikacja zapośredniczona, w której część komunikatów jest dostępnych bez potrzeby dzielenia wspólnego miejsca [Mikułowski-Pomorski 2006, s. 102].

W GBC istnieją wyłącznie obiekty cyfrowe, jednak znajdują się w niej także opisy (metadane) dokumentów tradycyjnych, dzięki czemu może ona pełnić rolę biblioteki hybrydowej, udostępniającej bezpośrednio (najlepiej bezpłatnie, ale także komercyjnie) obiekty cyfrowe oraz za pośrednictwem metadanych (znajdujących się w OPAC i bazach danych) dokumenty tradycyjne. Dzięki temu możliwe jest stosowanie GBC do identyfikacji i lokalizacji obiektów fizycznych, przechowywanych w bibliotekach naukowych, archiwach, muzeach, należących do władz różnego szczebla, przedsiębiorstwach i wszelkich innych miejscach, w których znajdują się zasoby ważne dla nauki. Dodatkowo niemal każdy rodzaj dokumentu tradycyjnego może być i jest digitalizowany. W efekcie powstają mniej lub bardziej złożone obiekty cyfrowe, co powoduje stały wzrost zasobów cyfrowych kosztem analogowych, o czym była już mowa w p. 1.3.

Obiekty cyfrowe stają się coraz bardziej odmienne od dokumentów tradycyjnych, takich jak książki i czasopisma drukowane. Wciąż składają się głównie z tekstu, ale zawierają także zbiory danych, materiały audiowizualne i multimedialne, symulacje, aplikacje software'owe i inne obiekty nietekstowe. Mogą charakteryzować się dowolnym formatem, często modyfikowanym dynamicznie. Powstają wciąż nowe rodzaje dokumentów, co zmusza do tworzenia nowych technologii służących ich dystrybucji, archiwizowaniu, wizualizacji i wyszukiwaniu.

Wielu autorów, mówiąc o bibliotece cyfrowej, stosuje do niej pojęcia z okresu bibliotek tradycyjnych. Bibliotekarze zazwyczaj wyobrażają ją sobie w sposób zinstytucjonalizowany, jako lokalny zasób cyfrowych, często zdigitalizowanych (a więc skopiowanych z drukowanych oryginałów) dokumentów, do którego dołączone są metadane i narzędzia wyszukiwawcze [Arms 2000]. Nawet, gdy mówią o globalnej bibliotece cyfrowej, zazwyczaj mają na myśli wiele instytucji powiązanych umowami, czy inaczej sfederowanych w sposób formalny, z centralną instytucją koordynującą [King 2004]. Tego typu zasób można określić mianem biblioteki cyfrowej w węższym znaczeniu i zaliczyć raczej do obszaru elektronicznego publikowania, wraz z czasopismami elektronicznymi, repozytoriami⁵³ i innymi sposobami upowszechniania dokumentów w Internecie, którymi zajmę

⁵³ Zauważmy, że nasze biblioteki cyfrowe, takie jak PBI czy używające oprogramowania dLibra różnią się od typowych repozytoriów jedynie rodzajem publikowanych materiałów (w bibliotekach cyfrowych dominują dane przetworzone, które można określić jako publikacje, w repozytoriach większą część zasobów stanowią nieprzetworzone dane prymarne) oraz osobą umieszczającego tekst w repozytorium: zamiast autora (twórcy),

się w dalszej części książki. Stanowią one rozproszone serwisy w sumie składające się na globalną bibliotekę cyfrową. W takim sensie mówi się np. o istnieniu wielu bibliotek cyfrowych, podczas gdy właściwie jest traktowanie wszystkich, zróżnicowanych zasobów jako jednej, światowej „biblioteki” – globalnej biblioteki cyfrowej we właściwym znaczeniu. Taka biblioteka nazywana bywa również biblioteką wirtualną, zapewne dla zaznaczenia jej „niematerialnego” charakteru, co nie do końca jest prawdziwe. Globalna biblioteka cyfrowa składa się więc z wielu rodzajów publikacji elektronicznych oraz serwisów udostępniających je w bardzo różny sposób z punktu widzenia organizacji dostępu – od całkowicie nieformalnych, „prywatnych” stron internetowych i blogów uczonych, aż po bardziej zinstytucjonalizowane i sformalizowane formy organizacji, takie jak repozytoria instytucjonalne (realizujące często politykę władz instytucji) czy biblioteki cyfrowe (dla których tworzy się plany digitalizacji podobne do polityki gromadzenia zbiorów wcześniej istniejącej w tradycyjnych książnicach). Ze względu na to, że jest wiele miejsc – serwisów, gdzie przechowywane są informacje cyfrowe (przechowywanie zorganizowane jest w sposób rozproszony), w każdym miejscu udostępniającym te dane należy prowadzić spójne i jednolite metody udostępniania tych zasobów.

Z punktu widzenia użytkownika w miarę możliwości powinien istnieć jeden system biblioteki cyfrowej; użytkownik postrzega cały Internet jako jeden wielki komputer, wyposażony w ogromne zasoby obliczeniowe i informacyjne, uniwersalną bazę danych [Dobrowolski 2004, s. 76], z których można korzystać w zjednoczony i bez mała nieograniczony sposób [Muraszkiwicz 2002, s. 30; Hofmokl 2009, s. 188]; takie podejście zgodne jest ze wspomnianą wcześniej ideą tzw. metakomputera. Dużą rolę w takim postrzeganiu Internetu ma mocno ujednolicony sposób prezentacji informacji, a więc spójny interfejs użytkownika, pozwalający na ujednoczenie wyszukiwania i wyświetlania informacji [Lagoze, Fielding, Payette 1998, s. 134]. W zakresie układu stron Web i nawigacji pomiędzy nimi oraz wyglądzie takich detali, jak przyciski i terminy stosowane w menu, istnieje wiele konwencji stylistycznych wspólnych dla wszystkich GUI. Standaryzacja w tym zakresie, nawet, jeśli nieformalna, pozwala na szybkie rozpoczęcie korzystania z dowolnego serwisu globalnej biblioteki cyfrowej, również w przypadku, gdy wcześniej użytkownik nigdy nie miał z nim styczności. Jednak pomimo tego, że konwencje te są bardzo istotne, funkcjonują jedynie na powierzchni Web. Użytkownik, który będzie próbował rzeczywiście pracować z informacją dostępną online, szybko stwierdzi, że pozornie podobne serwisy BBC różnią się znacznie w warstwie semantycznej. Iluzja jednolitości usług jest jednak bardzo przydatna i użyteczna w zdecentralizowanym, rozproszonym środowisku, jakim jest BBC.

Globalna biblioteka cyfrowa, podobnie jak globalny rynek, nie jest więc tworem zinstytucjonalizowanym, organizowanym hierarchicznie przez jakikolwiek pojedynczy „ośrodek władzy”, chociaż w jej skład wchodzi także instytucje, takie jak biblioteki naukowe, muzea, archiwa, szkoły i uczelnie. Jest ona raczej organizowana przez inicjatywy oddolne specjalistów z określonej dziedziny, grupy osób zdefiniowane na przykład przez wspólne zainteresowania lub posługiwanie się tym samym językiem (np. językiem danej dyscypliny). Tworzą one mechanizmy organizacyjne, realizujące ich potrzeby. Z milionów podejmowanych przez nich niezależnych, choć powiązanych ze sobą decyzji, tworzony jest nadzwyczaj jednolity układ. Jego elementy mogą, pomimo braku tradycyjnie pojmowanej „instytucjonalności”, być uważane za nowy rodzaj instytucji, typowy dla funkcjonowania Sieci; Kazimierz Krzysztofek nazywa je *webnative* [Krzysztofek 2008, s. 14]. Są one w coraz większej części produktami technologii kooperacji, samoorganizującymi się sieciami technoludzkimi, nie podlegającymi sterowaniu [Bojar 2009, s. 18], takimi jak wspólnoty *computing*u Gridowego czy kolektywy wiedzy.

jak to najczęściej (ale nie zawsze) dzieje się w repozytorium, teksty w bibliotekach cyfrowych umieszczają bibliotekarze, którzy są wydawcami ich wersji cyfrowych.

Tak jak całym Internetem, również GBC rządzi raczej procesy typowe dla systemów chaotycznych. Takie procesy, często występujące w samoorganizujących się systemach, oparte są na naturalnych zasadach i prawdopodobieństwie przypadków. Systemy te powstają w wyniku wzajemnie dostosowujących się procesów, związanych z wielokrotnymi interakcjami pomiędzy obiektami i tworzonymi w ten sposób relacjami. Spontanicznie samoorganizujące się systemy, których złożoność powstaje z prostoty, a wyższy porządek z chaosu, uzyskują nadzwyczajne wyniki [Castells 2008, s. 82]. Samoorganizację Webu widać w zagregowanych strukturach odnośników (*link*) tworzących grupy (*clusters*⁵⁴) stron Web wydzielonych tematycznie, jednocześnie obrazujących zainteresowania grup ich użytkowników. Tego typu agregacje odnośników mogą dotyczyć działalności badawczej, na przykład grupy powiązanych ze sobą stron Web uczonych i ich projektów badawczych, publikacji i instytucji nauki funkcjonujących w danej dziedzinie. Inne grupy stron Web mogą zawierać poszczególne serwisy GBC, takie jak portale czy repozytoria. W tym sensie GBC jest podobna do złożonych sieci społecznych, które nie posiadają zaprojektowanej architektury, lecz samoorganizują się na podstawie lokalnych interakcji wielkiej liczby członków i ich grup [Björneborn 2004, s. 3].

Pojęcie odległości w przestrzeni mocno zmieniło swój sens. Obecnie mało ważne bowiem jest to, kto umieścił dokument w sieci i w jakim konkretnym miejscu, na jakim serwerze go posadził. Dużo ważniejsze natomiast jest, aby dokument ten mógł być odnaleziony i udostępniony potrzebującemu go użytkownikowi, w czym znacznie większą rolę odgrywają spójne protokoły i inne narzędzia wspomagające wyszukiwanie i prezentację informacji.

Taki obraz globalnej biblioteki cyfrowej odpowiada koncepcji tzw. obiektów transgranicznych (*boundary objects*), czyli takich, które zamiast umacniać odmienności pomiędzy grupami społecznymi i społecznościami (na przykład pomiędzy grupami uczonych), pozwalają na porozumienia poszczególnych grup ponad tylnymi granicami i odmiennosciami. Są one definiowane jako obiekty, które jednocześnie przynależą do kilku grup (społeczności) oraz realizują potrzeby informacyjne każdej z nich [Bowker, Star 2000, s. 30]. Mogą one być łatwo dostosowywane (kustomizowane) do potrzeb dowolnej grupy, co oznacza, że wykazują się znaczną elastycznością. Jednocześnie zapewniają one jednolitość działań w obrębie wielu zastosowań. Jest to możliwe dzięki słabemu ustrukturyzowaniu obiektów transgranicznych na poziomie ogólnym, z zachowaniem możliwości uzupełnienia i uszczegółowienia struktury na poziomie poszczególnych usług. Takie obiekty ułatwiają komunikację, jednak bez zacierania różnic pomiędzy grupami; te bywają nawet uwypuklane. Mogą to być obiekty materialne, formy organizacyjne, procedury czy przestrzenie konceptualne. Obiektami transgranicznymi są takie formy organizacyjne, jak standardowe struktury danych, metadanych oraz narzędzia, które stosują te standardy. Umożliwiają wymianę danych i dokumentów pomiędzy poszczególnymi grupami. Odgrywają także istotną rolę w procesach zmian, dokonywanych w obrębie grup; zmiany w obiektach transgranicznych powodują modyfikację infrastruktury społecznej jednej grupy, a to pociąga za sobą możliwości zmian także w innych grupach, wspólnie użytkujących dany obiekt.

Globalna biblioteka cyfrowa jest znakomitym przykładem obiektu transgranicznego, gdyż jej zawartość może być wykorzystywana przez wiele społeczności. Dane gromadzone w ramach jednego serwisu z myślą o potrzebach określonej grupy badaczy mogą być wykorzystywane przez inne grupy, których członkowie często prowadzą badania w innej dziedzinie. Każda z tych grup interpretuje dane w swoisty sposób, pomimo tego, że pobiera je ze wspólnego zasobu [Borgman i in. 2005]. W odniesieniu do GBC model ten wskazuje na dwa poziomy wymiany wiedzy pomiędzy różnymi kulturami; na pozio-

⁵⁴ Używane w jęz. angielskim słowo *cluster* oznacza grupę (klasę, gniazdo, klaster) obiektów, które są do siebie podobne pod względem wybranych cech.

mie syntaktycznym wspólny język i słownik dołączenia tych kultur, a na semantycznym – bazy danych i tezaury w celu likwidacji barier komunikacyjnych pomiędzy kulturami. Efektem znoszenia tych przeszkód jest powstawanie nowych kultur, w których pojawiają się nowatorskie sposoby postępowania, ułatwiające przełamywanie barier i przekraczanie granic. W takim sensie GBC jest nie tylko zasobem informacyjnym, ale głównie miejscem, w którym stale tworzona jest nowa wiedza [Mason, Hart 2007].

Pomimo braku tradycyjnych struktur organizacyjnych, a więc i możliwości zatrudniania specjalistów, bibliotekarze i pracownicy informacji także znajdują swoje miejsce w globalnej bibliotece cyfrowej, w poszczególnych jej serwisach. Mają swój udział w tworzeniu analiz i katalogowaniu, zarówno rzeczowym, jak i formalnym, co oznacza tworzenie metadanych, które wykorzystywane są do indeksowania przez wyszukiwarki internetowe. Dotyczy to metadanych o dokumentach publikowanych w Web, w różnego rodzaju repozytoriach i bibliotekach cyfrowych (ich tworzenie jest kolejnym zadaniem bibliotekarzy), a także metadanych przechowywanych w bibliotecznych OPAC, a dotyczących dokumentów tradycyjnych. Wynika z tego, że stawiana czasem teza o zagrożeniu, jakie wyszukiwarki stwarzają dla przyszłości bibliotek, a więc i dla biblioteki cyfrowej, jest błędna. Przeciwnie – zarówno zasoby i serwisy biblioteki cyfrowej tworzone przez bibliotekarzy, jak i nowe narzędzia indeksujące powinny osiągać korzyści ze swego współistnienia i współpracy. Współistnienie to powinno dać interesujące efekty, przewidywane przez niektórych autorów w postaci tzw. Amazoogle⁵⁵. Jednym z nich może być całkowita zmiana sposobu pracy bibliotekarza (specjalisty od informacji naukowej), gdyż w środowisku cyfrowym najrozsądniej jest, aby wykonywali oni swoje obowiązki w sposób rozproszony, bez sztywnych ram organizacyjnych, ale w ścisłym związku z grupą, na rzecz której realizują czynności, w tym także w formie telepracy⁵⁶.

W tabeli I. zestawiałem podstawowe elementy i funkcje naukowej biblioteki tradycyjnej, bibliotek cyfrowych jako przykładu serwisu GBC i globalnej biblioteki cyfrowej. Jak można zauważyć, funkcje wykonywane w bibliotece każdego rodzaju są podobne, ich realizacja właśnie stanowi o wartości dodanej, wytwarzanej w bibliotece. Wpływ na funkcje mają takie stałe elementy bibliotek, także naukowych, jak:

- użytkownicy wraz z badaniem i realizacją ich potrzeb,
- zasoby, bez względu na formę, nośnik i lokalizację,
- bibliotekarze (pracownicy informacyjni); ich rola w olbrzymim, rozproszonym środowisku GBC zmienia się, ale nie maleje.

Różnice wynikają głównie z dwóch przyczyn:

- zastosowania nowego, elektronicznego medium dokumentów cyfrowych, które stwarza zupełnie nowe możliwości w zakresie realizacji funkcji bibliotecznych;
- zastosowania połączonych sieci globalnej i lokalnych (intranety, ekstranety) do udostępniania tych dokumentów.

Wymienione przyczyny bardzo wpływają na zmianę roli niewymienionego dotąd stałego elementu biblioteki (tradycyjnej), jakim jest budynek biblioteczny i jego wyposażenie. Budynek biblioteczny, pozbawione fizycznych zasobów oraz fizycznej obecności użytkowników i personelu, zmieniają swoje funkcje, stając się miejscem spotkań, obiektami

⁵⁵ Według Lorcana Dempseya z OCLC jednym z przejawów zmian w bibliotekarstwie są nowe oczekiwania użytkowników spowodowane możliwościami, jakie dostarczają takie narzędzia jak Google i Amazon. Dempsey stawia pytanie – „dlaczego narzędzia wyszukiwania informacji stosowane w bibliotece nie mogą funkcjonować w taki sposób jak Amazon/Google?”. Tego typu zmiany w oczekiwaniach nazwał on efektem „Amazoogle” [Holmström 2004]. O łączeniu efektów pracy wyszukiwarek (indeksowanie pełnotekstowe Google) z katalogiem bibliotecznym (WorldCat OCLC) pisał także Jerzy Franke [Franke 2007, s. 125].

⁵⁶ Telepraca niestety nie musi oznaczać mniejszego wymiaru pracy – wręcz przeciwnie, jak pisze Marian Niezgoda, wraz z rozwojem telepracy kategoria czasu wolnego staje się jeszcze bardziej iluzoryczna [Niezgoda 2006, s. 113].

dydaktycznymi i biurowymi a nawet punktami gastronomicznymi. Świadczy to o malejącej roli biblioteki jako miejsca [Sennyey, Ross, Mills 2009, s. 253].

Przejście od bibliotek tradycyjnych do GBC polega więc w mniejszym stopniu na zmianie tego, co jest w nich wykonywane (funkcje i czynności), a większym tego, w jaki sposób jest to wykonywane (organizacja, metody, narzędzia, standardy). Na ten aspekt zwraca uwagę Jane Greenberg, która opisała podobieństwa między bibliotekami a Semantycznym Webem w zakresie realizowanych funkcji; termin „gromadzenie” zastąpiła ona terminem „selekcja zasobów Webu”, „katalogowanie” – „semantyczną reprezentacją Webu”, „informowanie” – „usługami Semantycznego Webu” a „udostępnianie” – „wykorzystaniem zasobów Webu” [Greenberg 2007, s. 215 i nast.]. Należy zwrócić uwagę na istotną rolę nowych technologii w procesie przejścia od zasobów do usług. Ważniejsze stają się umiejętności informacyjne i kompetencje technologiczne pracowników informacji, niż ich wiedza o treści zasobów informacyjnych [Sennyey, Ross, Mills 2009, s. 256].

Tak rozumiana globalna biblioteka cyfrowa wpisuje się w powstałą na przełomie lat osiemdziesiątych i dziewięćdziesiątych XX w. koncepcję nowej organizacji badań naukowych. Staje się ona częścią kolaboratorium, które rozumiane jest jako nowa struktura organizacyjna dla działalności naukowej, szczególnie związanej ze współpracą realizowaną za pośrednictwem sieci komputerowej. Koncepcja ta zakłada funkcjonowanie wirtualnej nauki, wspomaganą przez globalną bibliotekę cyfrową.

Globalna biblioteka cyfrowa będzie znacznie mocniej włączona w procesy badawcze i dydaktyczne, niż to się dzieje obecnie. Stanie się częścią wirtualnego laboratorium, a nie dodatkiem lub uzupełnieniem do instytucji, realizujących procesy badawcze, co zasadniczo zmieni także role poszczególnych uczestników komunikacji naukowej. Te ulegną wymieszaniu, gdyż granice między tworzeniem informacji, publikowaniem, gromadzeniem i udostępnianiem, zacierają się.

Bibliotekarze muszą „wmieścić się” w skład zespołów badawczych i dydaktycznych, stając się ich integralną częścią [Marlino, Sumner 2008, s. 194]. Nowa rola bibliotekarza cyfrowego będzie wiązać się z bezpośrednim uczestnictwem w procesie badawczym i dydaktycznym. Nastąpi pewnego rodzaju dyfuzja funkcji biblioteki cyfrowej i jednostki badawczej/dydaktycznej. Bibliotekarz stanie się badaczem i dydaktykiem nie tylko z powodu odpowiedniego zapisu w Ustawie, ale z racji faktycznie realizowanych zadań. Podobne funkcje i zadania wcześniej były przypisywane bibliotekarzom dziedzinowym, działającym w scentralizowanej bibliotece tradycyjnej. Bibliotekarz cyfrowy stanie się członkiem wirtualnego kolaboratorium, zespołu badawczego funkcjonującego w cyberprzestrzeni, z zadaniem obsługi informacyjnej zespołu: zarówno dostarczania informacji niezbędnej do właściwej pracy zespołu, jak i przekazywania informacji o jego pracach i osiągnięciach; te zmiany już się rozpoczęły w postaci tzw. bibliotekarzy uczestniczących (*embedded librarians*) [Siess 2009]. Tak, jak wynalazek Gutenberga uwolnił skrybów klasztornych od pracowitego przepisywania tekstów, podobnie biblioteka cyfrowa pozwala bibliotekarzom skoncentrować pełną uwagę na cyfrowym publikowaniu, wyszukiwaniu informacji, jej udostępnianiu i w ogóle na pracy z użytkownikiem.

Środowisko biblioteki cyfrowej i elektroniczne publikowanie wzmacnia rolę uczonych w stosunku do pozostałych uczestników łańcucha komunikacji naukowej. Przykładem tej tendencji może być Public Library of Science, gdzie pracownicy naukowcy są nie tylko autorami, recenzentami i redaktorami, co zdarza się również w środowisku tradycyjnym, ale zajmują się także wydawaniem i dystrybucją publikacji naukowych. Tak więc wcześniejszy model łańcucha informacji, w którym każdy uczestnik miał określoną rolę do odegrania, przestał być aktualny [Owen 2002, s. 276]. Zróznicowanie funkcji autora, wydawcy i bibliotekarza zanika, wzrasta natomiast zakres odpowiedzialności samych uczonych.

Tab. 1.

Elementy i funkcje biblioteki tradycyjnej, bibliotek cyfrowych i GBC.

	BIBLIOTEKA TRADYCYJNA	BIBLIOTEKI CYFROWE – SERWISY GBC	GLOBALNA BIBLIOTEKA CYFROWA
ZBIORY	Lokalne, scentralizowane, tradycyjne, niewielka liczba zasobów cyfrowych, głównie na nośnikach niesciekających; własność biblioteki	Lokalne, mało rozproszone; umieszczone na serwerach jednej lub kilku instytucji; obiekty cyfrowe są własnością danej biblioteki cyfrowej	Rozproszone w sieci, duże zróżnicowanie form (od tradycyjnych do wczesniej nieistniejących), wiele form własności, w tym Open Access
BIBLIOTEKARZE	Praca scentralizowana, hierarchiczne struktury organizacyjne, wiedza zw. ze specyfiką procedur bibliotecznych	Praca scentralizowana, struktury podobne jak w bibliotece tradycyjnej, nowe specjalności zawodowe	Telepraca, wolny zawód, bibliotekarze cyfrowi, „niezinstytucjonalizowani”, członkowie zespołów badawczych, wiedza informacyjna
UŻYTKOWNICY	Określona grupa, z dobrze zdefiniowanymi potrzebami, łatwo identyfikowalni (nieanonimowi)	Dobre określona grupa celowa użytkowników, możliwość pojawienia się użytkowników przypadkowych	Użytkownicy potencjalni (każdy przyłączony do Internetu), rozproszeni geograficznie, ze zróżnicowanymi potrzebami, trudni do zidentyfikowania, współpracujący ze sobą
POMIESZCZENIA	Określone miejsce w przestrzeni fizycznej, wyspecjalizowane budownictwo biblioteczne	Wyspecjalizowane pomieszczenia (np. serwerownie); brak pomieszczeń dla obsługi użytkownika	Decentralizacja, pracownice bibliotekarzy cyfrowych dowolnie rozproszone geograficznie; biblioteka jako portal, „miejsce” w sieci
GROMADZENIE	Polityka gromadzenia zbiorów; pozyskiwanie zasobów drukowanych, dostępnych głównie w handlu księgarskim i antykwarycznym	Tworzenie kolekcji cyfrowej na podstawie zbiorów drukowanych; selekcja do digitalizacji (polityka); uzupełnianie kolekcji zbiorami „born digital”	Brak polityki gromadzenia w sensie tradycyjnym – wyszukwanie i lokalizowanie w sieci zasobów spełniających określone kryteria jakościowe; „kolekcja” o zasięgu globalnym, wspólne gromadzenie zasobów przez zainteresowanych
OPRACOWANIE (FORMALNE) I WYSZUKIWANIE	Katalogi biblioteczne, w tym automatyzowane, dane bibliograficzne w jednym z niewielu formatów – standardów światowych, np. MARC 21, wskazujące na konkretny obiekt fizyczny przez znak miejsca	Bazy metadanych, tworzonych przez bibliotekarzy; zwiększona liczba schematów metadanych; metadane indeksowane przez wyszukiwarki	Rozproszone zasoby metadanych (wielkie schematów), wykorzystywane przez różne narzędzia wyszukiwawcze; powiązanie ze zbiorami przez odnośniki; terminy wyszukiwawcze ukazwane w kontekście; nawigacja w zasobach globalnych; indeksowanie pełnotekstowe wyszukiwarek jako proces integrujący zasoby metadanych
OPRACOWANIE RZECZOWE	JiW jako podstawa katalogu rzeczowego, ChWd i ChWz tworzone w wybranych ych)	JiW przejęte z bibliotek tradycyjnych; ontologie	Pełny tekst jako podstawa indeksowania. Każde słowo z tekstu i metaaksu (metadane) elementem wyszukiwawczym; JiW i inne narzędzia stosowane lokalnie w serwisach GBC; nowe narzędzia typu folksonomii
PRZECHOWYWANIE	Lokalne, w magazynach biblioteki	Lokalne, na dyskach biblioteki cyfrowej	Rozproszone w serwisach GBC różnego rodzaju, wyspecjalizowane serwisy archiwizujące
UDOSTĘPNIANIE	Lokalne, tradycyjne, bezpłatne; wypożyczalnia, czytelnie. Wypożyczanie międzybiblioteczne. Wielość interfejsów (OPAC, bazy danych, zasoby analogowe i cyfrowe)	Globalne, poprzez ujednolicony interfejs przeglądarki internetowej; indywidualnie zaprojektowane możliwości wyszukiwawcze	Zdalne i rozproszone, 7/24, wg zróżnicowanych zasad (bezpłatnie, odpłatnie wg różnych licencji); ujednolicony interfejs; Windowsowy interfejs wyszukiwarek
INFORMOWANIE	O zasobach własnych i innych bibliotek	O własnych, zdigitalizowanych i digitalizowanych zasobach	O zasobach wyspecjalizowanych z sieci ze względu na ich jakość, duża rola informacji faktograficznej

2. STRUKTURY

W tym rozdziale przedstawię struktury funkcjonujące na różnych poziomach GBC, czyli zagadnienia związane zgodnie z nomenklaturą metodologii 5S, ze Strumieniami, sekwencjami różnego rodzaju danych (bitów, bajtów, danych graficznych i sekwencji wideo) oraz Strukturami, czyli ustrukturyzowanymi obiektami cyfrowymi i ich zbiorami.

Pomimo tego, że każdy Strumień może posiadać inną naturę, rodzaj i przedmiot, wszystkie one mają zdolność do przechowywania i/lub dostarczania danych, a więc pośrednio także informacji, co znamionuje ich rolę w GBC. W zakresie Strumieni powstało wiele standardów technicznych, pozwalających na efektywne przekazywanie i prezentację danych¹. Największe znaczenie ma normalizacja w zakresie: kodowania znaków, formatów danych, języków kodowania i protokołów. W tej części książki rozpoczynam opis globalnej biblioteki cyfrowej od standardów stosunkowo niskiego poziomu, zapewniających reprezentację obiektów cyfrowych, a nawet kodowania poszczególnych znaków.

Standardy te uniezależniają wybór sposobu postępowania od konkretnego miejsca; procedury mogą być realizowane wszędzie, w sposób uniwersalny. Muszą one natomiast powstawać we właściwym momencie; nie mogą być tworzone za wcześnie, ani za późno w stosunku do ogólnego rozwoju technologii. Zbyt wczesne wprowadzenie standardu może spowodować ograniczenie rozwoju danej technologii, zbyt późne może doprowadzić do chaosu w technologii i rynkach aplikacji oraz powstania silnie wyizolowanych grup używających konkurencyjne standardy.

Zagadnienia te w kontekście biblioteki cyfrowej stanowią poważny problem ze względu na gwałtowny rozwój technologii, za którym próbuje nadążyć standaryzacja. Być może najłatwiej dostrzegalną cechą naukowego serwisu GBC, jest stałe napięcie pomiędzy dwoma ekstremalnymi zjawiskami: bardzo gwałtownymi zmianami technologicznymi, dokonywanymi w coraz krótszych okresach oraz potrzebą długotrwałego funkcjonowania serwisu. Niezbędne jest więc pogodzenie dążenia do wykorzystywania najnowszej, ciągle zmiennej technologii z ideą statycznej i długotrwałej, czy nawet „wieczystej” archiwizacji obiektów cyfrowych.

Nowe technologie są elementem łączącym Strumienie ze Strukturami. Temat ten nie dotyczy GBC bezpośrednio, stanowiąc część zagadnień związanych z ogólną infrastrukturą informacyjną – Internetem, znajdującą się na wyższym poziomie niż GBC, jednak w sposób oczywisty dokonania w tym zakresie wpływają na możliwości i kierunki jej rozwoju.

Na temat technologii stanowiących podstawę Internetu, a w związku z tym również GBC, można by napisać zapewne niejedną książkę, dlatego ograniczę się tylko do kilku nowych kierunków, które wydają się mieć najbardziej rewolucyjne znaczenie, znajdując zastosowanie w GBC. Są to:

- Komunikacja bezprzewodowa, zwana mobilną. Znajdują się tu różnego rodzaju urządzenia personalne do przesyłania informacji oraz przenośne urządzenia komputerowe,

¹ Standardy przedstawione w tym i następujących rozdziałach należy traktować jako wybór najważniejszych rozwiązań spośród setek oficjalnych i nieformalnych, stosowanych w cyfrowej komunikacji naukowej. O liczbie standardów stosowanych jedynie w związku z digitalizacją może dać wyobrażenie praca pod red. Grzegorza Płoszajskiego [Płoszajski, red. 2008].

które nadzwyczajnie rozpowszechniły się od połowy lat dziewięćdziesiątych XX wieku. Po laptopach przyszła kolej na kieszonkowe urządzenia przenośne, stanowiące połączenie komputera, telefonu, czytnika książek, odtwarzacza multimedialnego, fotoaparatu i innych [Gawrysiak 2008a, s. 277]. Pozwalają one na upowszechnienie dostępu „z każdego miejsca, o każdym czasie”, powodując powstawanie zupełnie nowych zachowań informacyjnych.

- Internet szerokopasmowy. Jego rozpowszechnienie oznacza możliwość przekazywania olbrzymich zasobów danych poprzez łącza telekomunikacyjne. Technologia ta dotyczy nie tylko możliwości konkretnych urządzeń, takich jak łącza światłowodowe. Chodzi o naprawdę wysoce funkcjonalną infrastrukturę, zdolną obsłużyć wielką liczbę aplikacji i serwisów, będących podstawą IIN, jak na przykład technologie Gridowe. Dotychczas technologie szerokopasmowe wykorzystywane były głównie w zakresie rozrywki i handlu elektronicznego. Następnym krokiem jest umożliwienie użytkowania Internetu jako przestrzeni publicznej (zob. nast. rozdz.). Serwisy GBC wykorzystują nowe możliwości w celu konwergencji wszelkiego typu danych (wideo, audio, tekst i in.), tworząc z nich jeden wielki strumień danych, do którego dostęp zapewniają sieci szerokopasmowe.
- Rozproszenie przechowywanie i przetwarzanie. Polega nie tyle na rozproszeniu w sensie lokalizacji, co uniezależnieniu danych od określonego komputera lub systemu. Technologia tego typu już działa dla dokumentów dźwiękowych, które można kupować i ściągać na dowolne komputery podłączone do Internetu. Oczekuje się rozwoju tej technologii dla wszystkich innych rodzajów danych i mediów; takie rozpowszechnienie nowych technologii jest ich ważną cechą. Inną jest łatwość stosowania, przy jednoczesnym zapewnieniu bezpieczeństwa danych. Rozwinięcie technologii tego rodzaju pozwoli skupić się twórcom serwisów GBC na archiwizacji danych.
- E-płatności. Chodzi tu głównie o systemy pozwalające na realizację mikropłatności za mikrotreści, o których jeszcze będzie mowa w dalszej części książki. Użytkownicy Internetu kupując treści dostępne w Internecie, płacą za małe części obiektów i usług (np. za artykuł, a nie za prenumeratę czasopisma), wydając każdorazowo niewielkie kwoty. W połączeniu z komunikacją mobilną, e-płatności wpływają na sposób realizacji procesu badawczego [Chien 2004, s. 20-21].

W pierwszej części rozdziału (2.1-2.4) przedstawię najważniejsze standardy używane w GBC oraz zasady ich wykorzystania. Opisują one podstawowy, fizyczny poziom struktur stosowany w tej bibliotece; Strumienie bowiem także nie są całkowicie nieustrukturyzowanymi, chaotycznymi zbiorami bitów i bajtów.

Struktury na wyższym, logicznym poziomie opisane zostaną w dalszej części rozdziału (2.5-2.8). Umożliwiają one, między innymi, tworzenie serwisów GBC, obsługujących różnicowane struktury logiczne, przedstawione w tej części rozdziału; wskazują one na sposób umieszczenia informacji w obiekcie cyfrowym, zdefiniowanym dzięki określeniu elementów informacyjnych oraz relacji między nimi. Struktury dotyczą zasad organizacji informacji oraz metainformacji, zawartych w publikacjach cyfrowych i ich opisach (metadanych tworzonych w sposób intelektualny oraz struktur indeksowych), pozwalających na ich wyszukiwanie.

2.1. Kodowanie znaków

Zasady kodowania znaków stanowią najniższy poziom standardów służących prezentacji tekstu. Stanowią podstawę, na której opierają się takie procesy, jak rozpoznawanie danych zapisanych w wielu językach, manipulacja nimi oraz ich wyświetlanie. Tworzenie znaków w formie cyfrowej wymaga sprzętu i oprogramowania odpowiedniego do ich wprowadzania, zapisywania, przetwarzania, sortowania, wyświetlania i drukowania.

Kodowanie polega na wyznaczeniu zbioru kodów znaków, służących reprezentacji otwartego tekstu (*plain text*), czyli takiego, w którym brak jest informacji dotyczących formatowania (na przykład rozmiaru czy kroju czcionki). Dokumenty tekstowe zbudowane z takich znaków mogą być rozpoznawalne przez komputery i urządzenia peryferyjne. Podstawą kodowania znaków tekstu otwartego przez wiele lat był standard ASCII², zawierający kody dla dużych i małych liter alfabetu łacińskiego, cyfry arabskie, znaki przestankowe oraz inne znaki specjalne, w sumie pozwalające na komputerowe zapisywanie, manipulacje i wyszukiwanie tekstu. Standard ten pozwalał na rozpoznawanie i obsługę tekstu cyfrowego przez procesory komputerowe, tworzenie ich przy pomocy klawiatury, wyświetlanie na ekranie i drukowanie bez względu na to, gdzie i przez kogo tekst ten był utworzony. Siedmiobitowy kod ASCII pozwalała na utworzenie jedynie 128 kombinacji, a więc zakodowanie takiej właśnie liczby znaków, co jest wystarczające dla zapisania większości tekstów w alfabecie łacińskim. Jednak dla zakodowania znaków na przykład alfabetu polskiego należało zubożyć ten kod o część znaków, zamieniając je na polskie litery. W dodatku każdy producent wybierał inne kody do zamiany, co spowodowało powstanie dziesiątków standardów ASCII z polskimi znakami diakrytycznymi. Istniał także problem stosowania różnych alfabetów jednocześnie. Nie rozwiązało go nawet wprowadzenie tzw. rozszerzonego, ośmiobitowego kodu ASCII, pozwalającego zapisać 256 znaków. Wybór niewłaściwego zestawu znaków powodował wyświetlanie różnego rodzaju znaków specjalnych zamiast potrzebnych znaków narodowych. Rozwiązaniem tego problemu jest 16-bitowy schemat kodowania znaków zwany Unicode, od 1993 r. funkcjonujący jako standard ISO³, w którym można zakodować ponad 96 tys. znaków. Prace nad Unicode rozpoczęły się z inicjatywy wydawców, potrzebujących do celów przetwarzania tekstów spójnego, międzynarodowego standardu kodowania znaków. Później przejęty został on także przez przemysł elektroniczny. Standard wciąż jest rozwijany, dodawane są nowe znaki dla coraz bardziej egzotycznych alfabetów, chociaż część alfabetów języków martwych i używanych przez niewielkie społeczności jest pomijana [Anderson 2003]. Unicode obsługiwany jest przez najważniejsze języki programowania, takie jak C, Perl i Java, a także najpopularniejsze przeglądarki internetowe, istnieją również zasady kodowania dla HTML i XML.

Unicode jest uniwersalny, co oznacza, że każdy dokument o określonym zestawie znaków może być skonwertowany do niego, a nawet powstały w ten sposób plik może być konwertowany z powrotem do poprzedniego zestawu znaków bez utraty informacji. Podczas projektowania Unicode cecha podwójnej kompatybilności była traktowana jako szczególnie pożądana. Jeżeli dana litera ze znakiem diakrytycznym w jakimś zestawie znaków reprezentowana jest jako pojedynczy znak, to jej odpowiednik musi być także umieszczony w Unicode, nawet, jeżeli istnieje lepszy sposób osiągnięcia tego samego efektu wizualnego [Witten, Bainbridge 2003, s. 138].

Kodowanie znaków w Unicode może być nieco skomplikowane, jednak ma duże znaczenie dla unifikacji zasobów rozproszonych w GBC i ich interpretacji. Każdy ciąg znaków może być zadeklarowany jako szereg 16-bitowych liczb całkowitych i wykorzystany do kodowania znaków w UTF-16⁴.

Podczas wyświetlania i drukowania znaków Unicode należy zachować ostrożność. Większość aplikacji stosujących ten kod uwzględnia cały arsenał fontów i stosuje wiele tablic mapujących każdy jego znak na literę możliwą do wyświetlenia w znanym kroju (foncie). Problemy sprawiają złożone znaki Unicode oraz kierunek wyświetlania sekwencji znaków. Na szczęście w nowych wersjach przeglądarek internetowych mankamenty te zostały odpowiednio rozwiązane.

² ASCII – American Standard Code for Information Interchange.

³ ISO/IEC 10646:2003 Information technology – Universal Multiple-Octet Coded Character Set.

⁴ UTF – UCS Transformation Format, gdzie UCS oznacza Unicode Character Set, format kodowania, do którego znaki Unicode są transformowane.

W przypadku braku możliwości kodowania znaków dla języków nie reprezentowanych na standardowej klawiaturze rozwiązaniem jest transliteracja, od dawna stosowana w bibliotekach. Polega ona na przydzieleniu znakom alfabetu niełacińskiego (najczęściej chodzi o język rosyjski, bułgarski lub języki blisko- i dalekowschodnie: japoński, arabski, chiński, koreański, perski, hebrajski i jidysz – zwanych w skrócie „JACKPHY”) znaków alfabetu łaćnińskiego. Jest ona niezbędna w przypadku, gdy sprzęt komputerowy nie obsługuje danego alfabetu.

2.2. Formaty danych

Formaty danych pozwalają na reprezentację obiektów cyfrowych. W zależności od rodzaju obiektu mogą one umożliwić tworzenie dokumentów tekstowych lub multimediów (zapisów grafiki, dźwięku lub obrazów ruchomych) oraz innych typów obiektów (arkuszy kalkulacyjnych, oprogramowania), a także połączeń tych obiektów w jednym dokumencie (obiekcie złożonym). Problem formatów danych, choć bardzo ważny z technicznego punktu widzenia, z perspektywy użytkownika jest mało dostrzegalny. Użytkownik żąda udostępnienia potrzebnych mu treści, bez względu na to, w jakim są formacie. To twórca tych treści i ich dostawca powinni zadbać o to, aby sposób prezentacji treści był najbardziej odpowiedni do ich zastosowań. Oznacza to także, że formaty powinny być tak implementowane, aby użytkownik nie doświadczał żadnych trudności związanych z ich stosowaniem.

Format jest podstawową cechą każdego zasobu cyfrowego, decydującą o możliwości jego efektywnego wykorzystania. Bez ściśle określonego formatu zasób cyfrowy jest jedynie zestawem nieodróżnialnych bitów. Informacja zawarta w zasobie w postaci bitów może być poprawnie zinterpretowana i przedstawiona w formie czytelnej dla człowieka wyłącznie pod warunkiem, że znany jest format zasobu. O ile więc można sobie wyobrazić archiwizację obiektów cyfrowych bez zwracania uwagi na format, jako zbioru bitów, to znaczenie tych bitów (a więc dostęp do kodowanych treści), może zostać zachowane wyłącznie pod warunkiem dbałości o obsługę formatu [Abrams 2007, s. 7]. Formaty danych określają więc strukturę fizyczną obiektu cyfrowego, która jest wynikiem przetworzenia informacji zawartych w tym obiekcie na dane w układzie bitowym, w odróżnieniu od struktury logicznej obiektów, którą zajmę się w następnym rozdziale.

Formaty używane w środowisku tak rozproszonym i zmiennym, jak globalna biblioteka cyfrowa, muszą umożliwiać wykorzystanie obiektów cyfrowych także w przyszłości, w związku z czym stosować należy formaty, co do których istnieje prawdopodobieństwo możliwie długiego funkcjonowania. Do określenia takich formatów mogą być przydatne następujące kryteria:

- Kryterium otwartości: formaty opisane w ogólnie dostępnych specyfikacjach nie są kodowane, funkcjonują na zasadzie licencjonowania GPL⁵ lub podobnej.
- Kryterium przenoszalności: format nie powinien być związany z określonym sprzętem lub oprogramowaniem.
- Kryteria jakościowe: format dobrej jakości powinien umożliwić oszczędności pamięci, w miarę możliwości zastępować inne formaty, być odporny na drobne uszkodzenia, ale jednocześnie prosty, szeroko stosowany, bezstratny. Ważna jest również możliwość tworzenia metadanych osadzonych (zob. p. 2.8), co uniezależnia od zasobów zewnętrznych i pozwala na tworzenie zasobów „samoopisujących się” [Clausen 2004, s. 11-12].

⁵ GNU General Public Licence.

Obecnie wciąż najczęściej używanym schematem obsługi formatów jest wykaz typów nośników (mediów) MIME⁶, zarządzany przez IANA⁷. Typy MIME są szeroko stosowane jako podstawowy sposób charakterystyki technicznej zasobów cyfrowych, w zastosowaniach związanych z globalną biblioteką cyfrową. Są one powszechnie wykorzystane, na przykład jako nagłówki typu treści w protokole HTTP, stosowanym w World Wide Web do informowania agentów Webu o typie zasobu. Stosuje się je w większości schematów metadanych, w tym Dublin Core.

MIME identyfikuje format dwupoziomowo: na wyższym poziomie podaje typ nośnika, na niższym jego podtyp. Na wyższym poziomie zdefiniowano następujące typy nośników (w nawiasie liczba podtypów): Application (90), Audio (46), Example, Image (12), Message (10), Model (3), Multipart (13), Text (18), Video(21). Na przykład format graficzny TIFF identyfikowany jest jako image/tiff. MIME staje się rozwiązaniem przestarzałym, coraz częściej odsłaniającym swe braki: niejednoznaczne określanie formatów powodowane nieodróżnianiem ich odmian, umieszczanie tego samego formatu w kilku miejscach (XML jest umieszczony zarówno jako text, jak i application). Kłopoty te powodują powstawanie nowych propozycji zestawów typów formatów.

Dla celów tej książki użyteczny jest inny podział formatów, pozwalający na wskazanie ich zastosowań w GBC. Można je podzielić następująco:

- Formaty dokumentów tekstowych (DOC, PS, DVI, HTML...);
- Formaty obiektów multimedialnych:
 - Formaty grafiki rastrowej (GIF, PNG, JPG, TIFF...) i wektorowej (SVG, CDR, SWF...).
 - Formaty dźwiękowe (WAV, WMA, MP3, MP4, OGG...).
 - Formaty obrazów ruchomych (wideo) (MPG, AVI...).
- Formaty hybrydowe: PDF, DjVu.
- Formaty danych (surowe dane z badań naukowych).
- Formaty grafiki ustrukturyzowanej (CAD, VSD, QXD...).
- Formaty arkuszy kalkulacyjnych (XLS...).
- Formaty baz danych (DBF, DDF...).
- Formaty kolekcji plików (ZIP, TAR...).
- Formaty konfiguracji i metadanych (CSS...).
- Formaty wspomagane programowo (TTF, zapisy stanu gry...).
- Formaty plików oprogramowania (Javascript, Java, SWF...) [Clausen 2004, s. 4-5].

Nazwy konkretnych formatów (w nawiasach) podane zostały jedynie przykładowo, a ich zastosowanie może się nawet zmienić. Jako przykład można podać format PDF, tworzony z myślą o dokumentach tekstowych, dla którego zostały później rozbudowane zastosowania w kierunku multimedii, arkuszy kalkulacyjnych, wypełnianych formularzy.

Równie ważną cechą formatów stosowanych w GBC jest możliwość indeksowania treści obiektów w tych formatach przez wyszukiwarki internetowe. Z tego powodu formaty graficzne mniej nadają się do tworzenia obiektów zawierających tekst. Format TIFF, stosowany często w bibliotekach cyfrowych, nadaje się do celów archiwizacji m.in. dzięki możliwości dodawania indeksowanych metadanych⁸. Tu przydatny jest jeszcze jeden podział formatów, ze względu na sposób ich wykorzystania w cyklu życia obiektu cyfrowego. Wówczas mowa jest o: 1) formatach pośrednich, służących do przygotowania materiału do digitalizacji czy publikacji, obróbki plików i ich porządkowania; 2) formatach archiwalnych, służących do długotrwałego przechowywania; oraz 3) forma-

⁶ MIME – Multipurpose Internet Mail Extensions.

⁷ IANA – Internet Assigned Name Authority (<http://www.iana.org/assignments/media-types/>).

⁸ W formacie TIFF, jak i wspomnianym już JPG możliwe jest automatyczne wstawianie metadanych, tworzonych przez urządzenia typu aparatów fotograficznych: w tym celu stosuje się standard przemysłowy Exif [Płoszajski, red. 2008, s. 40].

tach prezyencyjnych, przeznaczonych do udostępniania obiektów użytkownikowi, często ich indeksowania, a więc wyszukiwania i wykonywania innych czynności związanych z wykorzystaniem obiektu przez użytkownika końcowego.

Wspomniany już **format ASCII** jest prostą reprezentacją, przez co często dającą zbyt mało możliwości, jak na potrzeby biblioteki cyfrowej. Niemożliwe jest włączanie do niego metadanych (oprócz może nazwy pliku). Można stosować zautomatyzowane techniki indeksowania pełnotekstowego oraz pobierania takich elementów danych, jak tytuł, autor, daty i in. Metody wyróżniania tych danych polegają na wykorzystaniu nieformalnych zasad strukturyzowania dokumentów, przez co ich efektywność zależy od uporządkowania tej struktury. W wielkich zasobach danych cyfrowych nie należy jednak spodziewać się pełnej jednolitości ani nawet poprawności dostępnych struktur obiektów cyfrowych.

Obecnie powszechnie stosowane są bardziej rozbudowane formaty obiektów cyfrowych, pozwalające na łączenie tekstu i grafiki. Coraz większe zastosowanie znajdują języki opisu strony, takie jak **PostScript** i **PDF**. Języki takie pozwalają zapisywać strony dokumentów w sposób niezależny od urządzenia wyjściowego (drukarki, monitora) stosowanego do udostępniania dokumentu. Początkowo edytory tekstów oraz programy graficzne dołączały do tworzonych dokumentów kod określonego urządzenia wyjściowego, przez co nie mogły być stosowane z innymi urządzeniami. Zmiana nastąpiła po zastosowaniu języków opisu strony; tu programy generują dokumenty graficzne w formacie sprzętowo niezależnym, który umożliwia drukowanie na każdej drukarce zaopatrzonej w odpowiedni sterownik.

W zastosowaniach biblioteki cyfrowej dokumenty te traktowane mogą być w większości przypadków jak „czarne skrzynki”: dokumenty są tworzone w określonym języku, wyświetlane, przekazywane i drukowane bez potrzeby znajomości wewnętrznej struktury. Jednak dla tworzenia spójnych kolekcji tych dokumentów przydatna jest dodatkowa wiedza o ich formatach, gdyż daje to możliwość wykorzystania wybranych cech, takich jak na przykład indeksowanie tekstu, wstawianie zakładek i grafiki.

PostScript, pierwszy komercyjny język opisu strony, powstał w 1985 r. po czym szybko został zaimplementowany przez firmy softwarowe i producentów drukarek jako niezależny od platformy sposób na opis drukowania stron, na których można umieszczać zarówno tekst, jak i grafikę. Pojawiło się oprogramowanie, pozwalające drukować na drukarce dokładnie to, co wyświetlane było na ekranie, w tak zwanym trybie WYSIWYG (What You See Is What You Get).

PDF jest językiem opisu strony, który wyrósł z PostScriptu, jednak z pominięciem części jego wad. Generuje obraz dokładnie w taki sam sposób, jak poprzednik. Jest to format strony posługujący się zdefiniowanymi obiektami graficznymi, takimi jak linie proste, krzywe, tekst i obiekty graficzne, modyfikowanymi przez różnego rodzaju transformacje. PDF jest niezależny sprzętowo i wyrażany w ASCII. Ponieważ w PDF zawarty może być zarówno tekst, jak i grafika (rastrowa, wektorowa), a także elementy aktywne, format ten (podobnie jak opisany dalej DjVu) nazwać można formatem hybrydowym, posiadającym cechy zarówno formatów tekstowych, jak i graficznych.

Istnieją dwie główne różnice pomiędzy PDF a PS. Po pierwsze PDF nie jest pełnym językiem programowania, gdyż ograniczałoby to jego przenoszalność. Takie cechy jak kompresja czy szyfrowanie są wbudowane w format, nie ma więc potrzeby ich dodatkowego programowania. Po drugie PDF posiada nowe możliwości interaktywnego wyświetlania dokumentów. Struktura dokumentu jest narzucona, a nie wpisywana w zasady strukturyzacji dokumentu, jak w PS. Pozwala to na losowy dostęp do stron, hierarchiczną strukturę zawartości i nawigację w obrębie dokumentu. Możliwe jest tworzenie hiperlinków. Pliki PDF są zazwyczaj tworzone automatycznie przez odpowiednie aplikacje (np. edytory tekstów), a nie pisane ręcznie, jak się to zdarza w środowiskach specjalistów z zakresu nauk technicznych i ścisłych, chętnie stosujących PS.

Wiele cech formatu PDF dotyczy jego interaktywności. Możliwe jest dodawanie ad-

notacji, zawierających oprócz tekstu także dźwięk i wideo. Częścią dokumentu mogą być akcje, które uruchamiają wybrane aplikacje. Za pomocą tworzonych formularzy można zbierać ustrukturyzowane informacje. PDF jest stale rozwijany, powstają nowe wersje, które dawno już odeszły od pierwowzoru służącego drukowaniu dokumentów.

Kolejnym etapem rozwoju formatów strony jest format **DjVu**. Jest on nowatorską, wysoce efektywną metodą kompresji obrazu, utworzoną przez pracowników amerykańskiego koncernu AT&T do kompresji kolorowych, skanowanych dokumentów. Wkrótce okazało się, że świetnie nadaje się też do dokumentów generowanych elektronicznie i ma szansę zastąpić formaty PostScript czy PDF w publikowaniu i wymianie dokumentów elektronicznych. Format DjVu oparty jest na bardzo zaawansowanej, wciąż rozwijanej, metodzie segmentacji obrazu. Polega ona na rozdzieleniu każdego obrazu na odrębne warstwy i poddaniu ich odrębnej optymalizacji i kompresji. Jedną z warstw może być warstwa tekstowa, której zawartość może być indeksowana.

Dokumenty DjVu mają najmniejszy rozmiar ze wszystkich spotykanych w zastosowaniach profesjonalnych; są nawet do 1000 razy mniejsze niż pliki TIFF i zazwyczaj 5 do 100 razy mniejsze niż pliki JPEG czy PDF. Chociaż pliki PDF są obecnie powszechnie używane, były rozwijane głównie jako formaty stosowane do wydruków i jako takie nie nadają się najlepiej do zastosowań archiwizacyjnych czy publikacji sieciowych. Dokumenty DjVu są mniejsze, przez co szybciej przesyłane i otwierane, oraz daleko bardziej wiarygodne i niezawodne.

Format DjVu został zoptymalizowany do zastosowań w sieciach rozległych. Obsługuje odnośniki hiperlinkowe. Dzięki stosowaniu technik częściowego odczytu danych, treść dokumentu pojawia się w przeglądarce niemal natychmiast i wraz ze strumieniem danych przybiera swoją postać docelową. Format ten, jako jedyny może udostępnić wskazaną stronę dokumentu bez potrzeby wczytania wszystkich poprzednich stron. Pliki DjVu bez problemu mogą być osadzone na stronach HTML zamiast plików JPG czy GIF, jak i stanowią odrębne dokumenty będące uzupełnieniem serwisu WWW.

Część użytkowników Internetu publikuje swoje dokumenty w postaci, w której najczęściej są one wykonywane, a więc w formacie jednego z procesorów tekstu. Zarówno w Polsce, jak i na świecie, jednym z najbardziej popularnych programów jest **Microsoft Word**, więc posłuży on nam jako przykład. Zapisanie dokumentu wykonanego w edytorze tekstów powoduje utworzenie pliku w formacie wewnętrznym, specyficznym dla danego edytora. Microsoft Word pozwala zapisać dokument w różnych formatach, lecz najbardziej charakterystyczne są RTF (Rich Text Format) oraz format wewnętrzny, który można nazwać po prostu Word. Ten ostatni, jako format binarny pozwala tworzyć znacznie mniejsze pliki. Użyteczność formatu zwiększają takie opcje programu Word, jak obustronny import/eksport plików Word do HTML. Istnieje także możliwość łatwego przenoszenia plików do PDF. Jest to o tyle ważne, że w zastosowaniu do biblioteki cyfrowej mniej istotne są możliwości edycyjne formatu, a ważniejsza staje się możliwość zapisania dokumentu w wybranym formacie, ułatwiającym dostęp poprzez przeglądarkę internetową.

Poza formatami strony funkcjonują także formaty dokumentu, za pomocą których tworzy się strukturę całego dokumentu, a nie jego pojedynczych stron. Do formatów tych zaliczyć można języki kodowania, typu HTML czy XML, jednak zostały one odrębnie opisane (zob. p. 2.3) ze względu na ich znaczenie w globalnej bibliotece cyfrowej, wybiegające poza zwykłe tworzenie dokumentów (dotyczy to szczególnie XML).

W zakresie formatów dla multimediów można wyróżnić formaty dla obiektów graficznych (obrazów) oraz obiektów audio i wideo. Wszystkie one charakteryzują się dużym lub bardzo dużym zapotrzebowaniem na pamięci masowe.

Bardzo popularne są graficzne formaty rastrowe (bitmapowe). Obraz na ekranie lub wydruku tworzony jest z pojedynczych, czarno-białych lub kolorowych punkcików, zwa-

nych pikselami. Jakość obrazu determinowana jest liczbą punktów w jednostce powierzchni; zazwyczaj używa się cali. Wówczas powstaje jednostka: liczba punktów na cal, czyli w skrócie dpi (*dots per inch*). Każdy piksel zapisywany jest w określonej jednostce pamięci, zazwyczaj 1 bit dla obrazów czarno-białych, 8 do 32 bity dla obrazów kolorowych. Takie standardy powodują, że obrazy, szczególnie kolorowe, zajmują olbrzymią ilość pamięci. Kolorowy obrazek wielkości standardowej strony A4 zeskanowany w rozdzielczości 600 dpi (zapewniającej przeciętną jakość) zajmie ok. 100 Mb pamięci. Taki plik byłby trudny do wykorzystania, zarządzania i przechowywania.

Z tego powodu obrazy nie są przechowywane w tej formie na dyskach; większość stosowanych formatów zapewnia kompresję, pozwalającą zmniejszyć rozmiar pliku. Metoda kompresji stanowi ważną część każdego formatu, a jej stosowanie niekoniecznie oznacza utratę jakości. Istnieją dwa rodzaje kompresji: stratna i bezstratna. Kompresja bezstratna oznacza, że zdekompresowany plik jest identyczny, jak przed kompresją. W ten sposób zawsze dokonuje się kompresji tekstu, gdyż niewielkie różnice w bitach oznaczać mogą poważne zmiany w zawartości obiektu. Kompresja stratna nie gwarantuje, że zdekompresowany plik będzie identyczny z pierwotnym, jednak zmiany nie powinny być znaczące. Zazwyczaj stosuje się ją tam, gdzie drobne błędy mogą pozostać kompletnie niezauważalne.

Obecnie najczęściej używane są dwa bezstratne formaty: ogólnie znany GIF (Graphics Interchange Format) oraz nieco mniej znany, otwarty standard PNG (Portable Network Graphics). GIF pomyślany został jako sposób na wymianę obiektów graficznych, możliwych do stosowania na różnych platformach. Zapewniał rozsądny czas przesyłu danych po łączach modemowych. Problemy z prawami autorskimi do algorytmu kompresji stosowanego w GIF (który został opatentowany) spowodowały poszukiwania nowego, bezstratnego formatu. W ten sposób powstał PNG jako format otwarty. Podstawą kompresji jest tu schemat ogólnego stosowania zwany gzip. Dzięki lepiej funkcjonującemu algorytmowi kompresji PNG zazwyczaj osiąga o 10-30% lepszy poziom kompresji, niż GIF.

Najlepiej znanym stratnym formatem graficznym jest JPEG (Joint Photographic Experts Group) lub JPG. Jest on stosowany w różnych aplikacjach typu DTP, grafiki artystycznej, transmisji fotografii prasowych czy diagnostyce obrazowej w medycynie, jak również w powszechnie stosowanych urządzeniach typu fotograficznych aparatów cyfrowych. Wykorzystywany jest dla publikowania wszelkiego rodzaju fotografii w Web.

Oprócz formatów rastrowych stosuje się także formaty grafiki wektorowej (geometrycznej), w których obraz zapisany jest w postaci figur geometrycznych, opisanych wzorami matematycznymi. Dzięki temu pliki są mniejsze, a obrazy można skalować bez utraty jakości. Formaty wektorowe, takie jak **Flash** lub Scalable Vector Graphics (**SVG**), pozwalają na interaktywną pracę z użytkownikiem oraz wizualizację danych (zob. p. 3.6). Mogą one także funkcjonować jako interfejsy wyszukiwawcze.

Coraz częściej w bibliotece cyfrowej używane są nagrania dźwiękowe (audio) i wideo. Media tego typu wymagają kilkakrotnie więcej pamięci niż grafika, gdyż muszą opisywać zmiany sygnałów w czasie. Rozwój technologii pociągnął za sobą konieczność opracowania sposobu efektywnej kompresji, ponieważ np. obraz w standardzie PAL składa się 25 klatek na sekundę, 720 punktów w poziomie i 576 punktów w pionie, a kolor każdego z tych punktów opisany jest 24 bitami. Oznacza to, że każda sekunda nieskompresowanego obrazu w standardzie PAL ma wielkość prawie 30 MB, a przeciętny, półtgodzinny film zajmie ponad 156 GB, co stanowi pojemność 224 CD. Formaty dokonują kompresji maksymalizującej ilość informacji, którą można przesłać przez sieć o określonej przepustowości lub zapisać na dysku o danej pojemności.

Dokumenty są kodowane w celu ich transmisji (lub zapisu na nośnikach) oraz dekodowane podczas odtwarzania. Oprogramowanie kodera/dekodera (czyli kompresora/dekompresora) ogólnie nazywane jest kodekiem. **Kodeki** mogą być stosowane do informacji

audio lub wideo odrębnie, lub też łączone ze sobą pozwalając na stosowanie dokumentów audiowizualnych. Takimi kodekami są na przykład DivX, stosujące kompresję typu MPEG-4. Ogólnie zasada ich działania jest podobna, jak dla grafiki: możliwa jest kompresja wysokiego stopnia, ale stratna, lub bezstratna, jednak dająca gorsze rezultaty.

Wśród wielu stosowanych formatów wymienić można dla przykładu takie, jak formaty wideo: MPEG, AVI, QuickTime oraz formaty obiektów dźwiękowych WAV, AIFF i AU. Tak duża różnorodność jest nie tyle potrzebą techniczną, co wynikiem konkurencji między firmami komercyjnymi, popierającymi te rozwiązania (Microsoft, Sun, Apple i in.). Nie sprzyja ona niestety ujednoczeniu formatów zasobów GBC. Kłopoty sprawiają także metody wyszukiwania informacji multimedialnej, sposoby zapewnienia użytkownikom dostępu do różnych części obiektów, opis treści tych obiektów, ocena poprawności procesu digitalizacji.

Digitalizacja sygnałów fizycznych, analogowych w swojej naturze, jest trudnym zadaniem, do którego niezbędne jest doświadczenie i dbałość o szczegóły. Efekty zależą na przykład od właściwego oświetlenia obrazu oraz poziomu zapisu dźwięku. Od wyboru formatu wyjściowego zależy także stopień utraty danych, czyli liczba utraconych szczegółów. Zasadnicza decyzja dotyczy więc akceptowalnego poziomu strat generowanych przez format. Bywają zastosowania, w których nie można tolerować żadnych strat, nawet kosztem gwałtownie rosnącego zapotrzebowania na pojemność pamięci; często należą do nich aplikacje w nauce i komunikacji naukowej (archiwizacja).

W przypadku reprezentacji bezstratnej, czynności związane z edycją obiektów są proste i bardziej pewne. Istnieją co prawda narzędzia pozwalające na odtwarzanie i edycję obiektów w formatach skompresowanych, jednak na kolejnych dekompresjach i kompresjach bardzo traci jakość obiektu, który w efekcie może stać się nawet bezużyteczny. W związku z tym edycję wykonuje się na obiektach w formatach o niskiej kompresji, takich jak WAV czy AIFF (pliki dźwiękowe), dopiero końcowy rezultat zapisując w formacie typu MP3 (MPEG Audio Layer 3), zapewniającym wysoką kompresję przy wciąż niezłej jakości dźwięku.

Obecnie czymś normalnym stało się przesyłanie obiektów multimedialnych w sieci, bez względu na używane platformy sprzętowe. Zamiast zdawać się na jedno, stale stosowane rozwiązanie, użytkownicy mogą wybrać najbardziej odpowiadający im format, na który konwertowane są żądane obiekty, bez względu na format, w którym zostały wykonane. Pozwala to na dostęp użytkowników do materiałów w najlepszej dla nich formie, z jednoczesnym uwzględnieniem dostępnej przepustowości sieci.

W środowisku globalnej biblioteki cyfrowej funkcjonuje i ewoluuje coraz bardziej obszerny zestaw formatów, w których obiekty cyfrowe mogą być przechowywane i udostępniane, a także narzędzia służące do ich przetwarzania⁹ oraz zasady stosowania formatów, ich wyboru, tworzenia, wykorzystania i archiwizacji. Jednocześnie coraz szybciej rozwijają się nowe formy treści, charakteryzujące się powszechnością stosowania, zróżnicowaniem i coraz większym skomplikowaniem. Wraz z pojawianiem się nowych formatów lub nowych zastosowań istniejących, coraz bardziej widoczna jest ich istotna rola w przechowywaniu (archiwizacji) i udostępnianiu informacji. Format danych ma też wpływ na możliwości ich wyszukiwania.

⁹ Wymienić można narzędzia do walidacji, identyfikacji i charakteryzacji formatów typu JHOVE (JSTOR/ Harvard Object Validation Environment) czy repozytoria formatów, takie jak GDFR (Global Digital Format Repository), tworzony przez Harvard Univ. i OCLC.

2.3. Języki kodowania

Od początku istnienia usługi Web w Internecie, informacja udostępniana jest tam w formie stron Web, tworzonych za pomocą języków kodowania (*markup*). W praktyce na ekranie wyświetlane są plansze graficzne, które zazwyczaj zawierają tekst uzupełniony obiektami multimedialnymi. Od tradycyjnych różni się one tym, że są programowalne, tzn. zarówno ich treść, jak i forma może wynikać ze sposobu funkcjonowania programu komputerowego. W praktyce języki kodowania służą tworzeniu niewielkich i dość prostych programów sterujących wyświetlaniem obrazu na monitorze.

W najprostszych przypadkach kodowanie wykorzystuje się do określenia struktury dokumentów oraz do opisu ich wyglądu na ekranie użytkownika. Obejmuje takie elementy, jak podział tekstu na sekcje, tytuły (nagłówki), listy numerowane, cytowania (odnośniki), wspomniane multimedia. Jego zadaniem jest również identyfikacja metadanych związanych z kodowanym dokumentem. Oznacza to, że podczas kodowania posługujemy się językiem złożonym homologicznie [Wojciechowski 2006, s. 25].

Pierwszym językiem służącym kodowaniu dokumentów, z którego w jakimś stopniu wywodzą się wszystkie inne, jest **SGML**. Został on utworzony przez przemysł wydawniczy. Stosowane w nim etykiety pozwalają na tworzenie rozbudowanej struktury dokumentów; kodowanie obejmuje takie elementy struktury tekstu, jak tytuł, data, autor, główna część tekstu z podziałem na rozdziały i podrozdziały. Dostarczają także informacji o sposobie prezentacji dokumentu i zastosowanych elementach typograficznych, takich jak zmiany w wyglądzie typograficznym i przejście do nowego wiersza. Etykiety SGML zbudowane są ze znaków ASCII, do tworzenia plików SGML nie ma więc potrzeby używania specjalnego oprogramowania. Dzięki temu pliki te łatwo jest przesyłać w sieci, gdyż kody są z punktu widzenia zastosowanych znaków częścią tekstu dokumentu, nie muszą być specjalnie interpretowane, jak na przykład zmiany fontów w większości edytorów tekstów. Mimo tych cech SGML nie jest zbyt przydatny do stosowania w bibliotece cyfrowej ze względu na poziom skomplikowania samego kodu, jak i oprogramowania do jego przetwarzania, a także niedostateczne współdziałanie, utrudniające wymianę obiektów SGML pomiędzy różnymi systemami.

W zakresie języków kodowania istnieją obecnie trzy poziomy rozwiązania. Hypertext Markup Language (HTML) jest uproszczonym językiem kodowania sposobu wyświetlania dokumentu, który służy głównie jego prezentacji. Pozwala na tworzenie zarówno tekstu dokumentu, jak i etykiety opisujących jego format i strukturę przez formowanie nagłówków, akapitów, list, odnośników hipertekstowych i in. HTML jest uproszczoną wersją SGML, powszechnie stosowaną w Webie. HTML jest przykładem prostego zestawu etykiet, pozwalającego na rozpowszechnianie dokumentów, tworzonych za pomocą tych etykiet. Prostota tego rozwiązania może także stwarzać problemy, szczególnie, gdy chcemy za pomocą tych prostych struktur opisać złożoną informację.

Extensible Markup Language (XML), znajdujący się na drugim poziomie, jest kolejną odmianą SGML, specjalnie zaprojektowaną dla Webu jako struktura bardziej otwarta. Zawiera on elastyczny mechanizm opisywania struktury dokumentów oraz metadanych, a także służy przechowywaniu i przekazywaniu dokumentów. XML tworzony był z myślą o łatwiejszym przesyłaniu dokumentów niż SGML, ma także mniejsze wymagania co do oprogramowania aplikacyjnego. Nie zawiera on predefiniowanego zestawu etykiet, przez co pozwala na lepszą wymianę dokumentów. Język jest niezależny od platformy sprzętowej czy softwarowej, przez co jest bardzo odpowiedni do zastosowania w bibliotece cyfrowej.

Kolejnym językiem kodowania jest RDF (Resource Description Framework), który może być traktowany jako trzeci poziom struktury. Jest to format stanowiący aplikację XML, służący kodowaniu metadanych, czyli informacji ustrukturyzowanej.

W Internecie, nie tylko w zastosowaniach GBC, istnieje stała tendencja zwiększania kontroli nad sposobem wyświetlania i w ogóle wyglądem dokumentów. Strony prezentowane w Web muszą mieć „profesjonalny” wygląd. Jednak praktyka dyktowania przez autorów dokumentów sposobu ich prezentacji poważnie różni się od pierwotnej idei HTML, która polegała na rozdzieleniu struktury dokumentu od jego prezentacji i pozostawiała decyzję co do wizualizacji dokumentów przeglądarce internetowej. Powodowało to, że wygląd strony trudniejszy był do przewidzenia, gdyż sposób wyświetlania dokumentu przez różne platformy bywał odmienny (co prawda zazwyczaj w szczegółach). W takim zamieszaniu powstawał standard HTML, rozwijany w kolejnych wersjach. Dzięki temu, że HTML jest powszechnie znany i stosowany w Web, może być używany przez wszelkiego rodzaju interfejsy globalnej biblioteki cyfrowej. Również obiekty źródłowe, udostępniane w tej bibliotece często tworzone są jako dokumenty HTML, dzięki czemu eliminowanych jest większość kłopotów, związanych z reprezentowaniem czystego tekstu.

Współczesne języki kodowania stosują słowa zamknięte w łamane nawiasy jako etykiety opisujące dokument. Na przykład wyrażenie `<title>To jest moja strona</title>` określa tytuł dokumentu HTML, wyświetlany u góry ekranu. Tak jak w tym przykładzie, dla każdej etykiety otwierającej przewidziano etykietę zamykającą, o tej samej nazwie, ale poprzedzoną znakiem /. Wiele przeglądarek pozwala jednak na pomijanie etykiet zamykających, co jedni traktują jako bałaganiarstwo, a inni uważają za przydatne ułatwienie. W ten sposób zwykle jest traktowana etykieta `<p>`, oznaczająca początek akapitu i jednocześnie uznawana przez większość przeglądarek za koniec poprzedniego akapitu, w związku z czym stosowanie etykiety `</p>` nie jest niezbędne.

Bardzo istotną cechą HTML jest możliwość tworzenia odnośników hipertekstowych. W kodzie źródłowym są one oznaczane przez wskaźniki `<a>...`. Adres dokumentu, do którego tworzy się odnośnik (może to być inna strona Web lub inna część tego samego dokumentu) występuje w formie atrybutu. Odnośniki mogą odsyłać do obiektów multimedialnych (audio, video) oraz obiektów w różnych formatach, opisanych wcześniej. Jako odnośnik wyświetlany jest tekst pomiędzy etykietami `<a>...`. Kliknięcie na niego powoduje przejście do wywoływanego obiektu.

HTML pozwala na bezpośrednie wprowadzanie metadanych osadzonych w części nagłówkowej dokumentu, przez wykorzystanie etykiet `<meta>`, jednak rozwiązanie to ma swoje ograniczenia. Nie jest ono powszechnie stosowane, być może ze względu na niechęć do ingerencji w źródło obiektu za każdym razem, gdy należy uzupełnić lub zmodyfikować metadane. Zamiast tworzenia metadanych stanowiących część źródła dokumentu, a więc czasem trudno odróżnialnych od głównej jego treści, słuszną może być decyzja o wyodrębnieniu metadanych lub też utworzeniu specjalnej wersji dokumentu, zawierającej metadane, służącej wyłącznie jego udostępnianiu, podczas gdy oryginał przechowywany jest bez potrzeby ingerencji w jego strukturę.

XML również jest uproszczoną wersją SGML, zaprojektowaną w taki sposób, aby ułatwić współdziałanie globalnych zasobów biblioteki cyfrowej w Web. Współdziałanie ułatwiane jest między innymi przez fakt, że XML jest niezależny od platformy, co umożliwia łatwą wymianę dokumentów pomiędzy różnymi systemami. Z praktycznego punktu widzenia można go nazwać dialektem SGML, w odróżnieniu od HTML, który jest przykładem języka kodowania, który można opisać przy pomocy SGML. W efekcie każdy dokument XML jest jednocześnie dokumentem SGML. Standard pozwala na elastyczne tworzenie charakterystyk struktury dokumentów i ich metadanych, dzięki czemu mogą być wykorzystywane w bibliotece cyfrowej. Stąd też znajduje szybko coraz większe zastosowanie.

XML jest językiem kodowania, mechanizmem służącym definiowaniu etykiet i relacji strukturalnych istniejących pomiędzy elementami dokumentów. Sztywne reguły syntaktyczne XML utrudniają konwersję na ten standard dokumentów w starszych wersjach

HTML. Problemy dotyczą głównie części kodów, które były niedbale sformułowane, stwarzając kłopoty podczas przetwarzania tych dokumentów. Jednak stosując różne sztuczki możliwe jest stworzenie dokumentu XML możliwie blisko odpowiadającego oryginałowi w HTML.

Dokumenty XML powinny być poprawne składniowo i strukturalnie. Mówimy o dokumencie, że jest poprawny składniowo (*well-formed*), jeżeli jest utworzony zgodnie z regułami składni XML. Reguły te dotyczą między innymi konieczności domykania znaczników wszystkich etykiet. Dokument niepoprawny składniowo nie może być przetworzony przez parser XML. Inaczej więc niż w HTML, tutaj należy dbać o składnię: w nazwach etykiet rozróżniane są małe i duże litery, etykiety muszą być domykane, należy dbać o poprawne zagnieżdżanie elementów.

Dokument jest poprawny strukturalnie (*valid*), jeżeli jest zgodny z definicją dokumentu, tzn. dodatkowymi regułami określonymi przez użytkownika. Do precyzowania tych reguł służą specjalne języki, z których najczęściej używane są DTD¹⁰ i Schemat XML.

XML jest bardzo rozbudowanym językiem. Umożliwia tworzenie formatów danych w obrębie struktur organizacyjnych serwisów GBC. Każdemu opublikowanemu dokumentowi może towarzyszyć opis jego struktury (semantyki), w formie DTD¹¹ lub Schematu XML (plik XSD). Oba stanowią schemat danych, zawartych w dokumencie XML: jego dozwolone elementy, ich etykiety oraz zasady łącznego stosowania. Dzięki temu zarówno człowiek, jak maszyna może łatwo poznać strukturę i znaczenie danych bez względu na ich pochodzenie. Dokument XML może zawierać dane w pełni indeksowane przez wyszukiwarki internetowe oraz opis ich struktury. Podobnie w pliku XSD można definiować strukturę metadanych wraz z ich cechami (takimi jak powtarzalność, obowiązkowość).

Dokument XML nie precyzuje, jak należy wyświetlać przechowywane w nim dane, przez co większość przeglądarek internetowych przy próbie wyświetlenia, potraktuje go jak zwykły tekst. Istnieją jednak łatwe sposoby na prezentowanie dokumentów XML na stronach internetowych w przyjazny dla użytkownika sposób. Dodatkowe narzędzia pozwalają na definicję sposobu wyświetlania dokumentu podobnie, jak robi się to w HTML.

Wokół XML narosło wiele innych standardów, które zwiększają jego możliwości i zastosowania. Sam XML pozwala na wyrażanie informacji strukturalnej, a więc metadanych. Jednak jego zastosowania nie ograniczają się do metadanych; rodzaj danych nie odgrywa roli. Zastosowania XML rozszerzane są dzięki zastosowaniu innych standardów, takich jak CSS¹² lub XSL¹³. Pozwalają one na zmiany struktur dokumentów, obsługę zapytań, wydobywanie informacji oraz formatowanie. Warto wspomnieć także o XHTML, który jest przedstawieniem HTML w postaci XML. XHTML jest z kolei podstawą innych formatów, na przykład EPUB¹⁴. Istnieją także „dziedziczne” odmiany XML¹⁵.

Znaczenie XML nie polega tylko na ułatwieniu kodowania dokumentów Web. Ujednolicenie struktur danych przekazywanych w procesach komunikacji, w tym naukowej,

¹⁰ DTD – *Document Type Definition*.

¹¹ Najbardziej znanymi DTD są EAD (Encoded Archival Description) i TEI (Text Encoding Initiative). Również HTML można traktować jako DTD języka SGML.

¹² CSS (Cascading Style Sheet) – język służący do opisu sposobu wyświetlania (prezentacji) stron Web w HTML. Pozwala na odseparowanie treści dokumentu od jego formy.

¹³ XSL (Extensible Stylesheet Language) – język służący do opisu sposobu prezentacji i przekształceń dokumentów tworzonych w formacie XML.

¹⁴ EPUB to format dokumentu, opracowany przez International Digital Publishing Forum (<http://www.idpf.org/>), stworzony z myślą o stosowaniu w różnych urządzeniach, w tym mobilnych (czytniki ebooków, telefony komórkowe), gdyż sposób wyświetlania tekstu dostosowywany jest do możliwości technicznych (wielkości ekranu). W formacie zastosowano kilka innych standardów, takich jak XHTML, XML, Dublin Core, CSS.

¹⁵ Najbardziej znane to MathML (<http://www.w3.org/Math/>), stosowany w matematyce i SMIL (<http://www.w3.org/TR/SMIL2/>), służący prezentacji multimediiów.

pozwała na automatyzację tych procesów, a dane oraz ich kontekst nie są tracone. Dzięki zastosowaniu XML zanikają granice tego, co tradycyjnie uważane było za dokument. Jednostka komunikacji naukowej może być rozszerzana i zawężana [Rzepa, Murray-Rust 2001, s. 180]. Standardowy tekst może być rozbity na mniejsze jednostki lub powiększany aż do rozmiarów „docuverse” (*document + Universe*) powiązanych tekstów [Delany, Landow 2008, s. 94]. Ponieważ podstawą XML jest oddzielenie prezentacji informacji od jej treści, to sposób jej udostępniania także może być definiowany na różnych poziomach agregacji.

W efekcie tworzyć można samoopisujące się dokumenty o znanej strukturze, do których punktem dostępu może być każdy element treści (danych i metadanych). Dzięki temu wyszukiwarki stosujące XML mogą prowadzić na przykład wyszukiwania na zapytania typu „tytuł zawiera słowo *metadane* i data publikacji jest po 2006 roku”. Obecnie jest to trudne do wykonania w wyszukiwarkach, choć niektóre dają już takie możliwości [Lesk 2005, s. 148]. Wyszukiwanie takie wcześniej możliwe było w strukturach metadanych wyodrębnionych w stosunku do opisywanych obiektów (tradycyjnych i cyfrowych) i przechowywanych w OPAC lub bibliograficznych bazach danych (zob. też p. 4.3).

RDF utworzony został pod auspicjami World Wide Web Consortium (W3C) w celu ułatwienia współdziałania metadanych, ich tworzenia, wymiany i wielokrotnego stosowania w kolejnych projektach (*reuse*), szczególnie w obrębie World Wide Web. Współdziałanie jest zapewnione dzięki mechanizmom ułatwiającym używanie ogólnie przyjętych konwencji, dotyczących semantyki, syntaktyki i struktury [Miller 1998]. RDF można traktować jako następcę Warwick Framework, którego koncepcja powstała w latach dziewięćdziesiątych ubiegłego wieku podczas prac nad Dublin Core [Miller, Hillman 2002, s. 59].

Ponieważ metadane dotyczą zbyt zróżnicowanych informacji, aby można je było określić wyczerpująco i jednoznacznie, RDF przejął zasady XML: zamiast *a priori* wprowadzać zestaw wszystkich możliwych przypadków, dostarcza on środków dla opisu poprawnego systemu. Społeczności użytkowników powinny wspólnie utworzyć systemy RDF najbardziej odpowiadające ich wspólnym potrzebom. Będą oni musieli porozumieć się co do słownika, znaczenia stosowanej w nim terminologii oraz struktur wyrażen formowanych przy jego wykorzystaniu. W tym celu stosuje się Schemat RDF, na tej samej zasadzie, jak DTD i Schematy XML służą zarządzaniu słownikiem i strukturami XML.

Efektywne stosowanie metadanych w globalnej bibliotece cyfrowej wymaga uzgodnienia wspólnych zasad dotyczących ich semantyki, syntaktyki i struktur. RDF, infrastruktura służąca opracowaniu tych elementów, udostępnia niezbędne narzędzia dla zarządzania informacją w Web jako środowisku biblioteki globalnej oraz umożliwia transformację Web w kierunku bardziej efektywnego zasobu informacji [Lassila 1997].

Szczególnie ważną cechą RDF z punktu widzenia GBC, jest ułatwienie modularnego współdziałania pomiędzy odrębnymi zestawami elementów metadanych. Określone zostały zasady współdziałania, wśród których wymienić można standardowy mechanizm reprezentacji semantyki, stanowiący prosty, lecz silny model danych. Dodatkowo RDF pozwala na publikację słowników czytelnych zarówno dla człowieka, jak i maszyny. Słowniki traktowane są jako zestawy cech (atrybutów), czyli elementów metadanych, definiowanych przez społeczności, zajmujące się opisem zasobów. Dzięki standaryzacji przygotowania deklaracji słownika, ułatwione jest współużytkowanie i rozwijanie semantyki pomiędzy różnymi społecznościami informacyjnymi. Takimi społecznościami są między innymi użytkownicy poszczególnych schematów metadanych, na przykład Dublin Core¹⁶. RDF jest zaprojektowany do obsługi modułowości schematów metadanych, polegającej na tym, że pewne części ich semantyki przejmowane są przez nowe projekty i modyfiko-

¹⁶ Społeczność użytkowników Dublin Core osiągnęła porozumienie co do zasad stosowania schematu DC w RDF; są one przedstawione na stronie Web DCMI [Nilsson i in. 2008].

wane zgodnie z własnymi potrzebami. Przy pomocy RDF tworzona jest infrastruktura pozwalająca na łączenie rozproszonych rejestrów atrybutów, dzięki czemu nie ma potrzeby tworzenia centralnego rejestru. W efekcie poszczególne społeczności (biblioteki, archiwa, muzea, firmy zajmujące się e-handlem) mogą deklarować słowniki opisujące semantykę, udostępniać je innym, którzy je mogą je modyfikować dla zaspokajania specyficznych potrzeb danej dziedziny lub w danym zastosowaniu.

Niezbędne technologie zostały już przygotowane. XML stanowi elastyczną syntaktykę, a RDF zawiera model danych ogólnego stosowania, służący uruchamianiu mechanizmów reprezentacji i deklaracji, wykorzystywanych przez społeczności do tworzenia własnych semantyk. Żaden z tych standardów nie określa jednak, jakie to mają być semantyki. Te problemy mają rozstrzygać same społeczności, dostosowując standardy do własnych potrzeb za pomocą dostępnych technologii. Należy stworzyć właściwą praktykę organizacyjną i zasady opisu, a także sposoby stosowania słowników. W przygotowaniu wspólnej semantyki społeczność bibliotekarzy z pewnością jest najbardziej zaawansowana. W tej społeczności utworzone zostały schematy metadanych, słowniki cech obiektów, a także modele ich stosowania. Jest to jeden z czynników stymulujących rozwój globalnej biblioteki cyfrowej.

2.4. Protokoły

Podobnie jak w każdej bibliotece, również w bibliotekach cyfrowych, a szczególnie w globalnej bibliotece cyfrowej, niezbędne są standardy, służące reprezentacji zbiorów (w przypadku GBC: obiektów cyfrowych), metadanych i zapytań. O części takich standardów była już mowa. W tym rozdziale zajmę się jeszcze jednym elementem, bez którego tych wcześniej wymienionych nie udałooby się zebrać w całość i zapewnić sprawnej komunikacji między nimi w środowisku tak rozproszonym, jak współczesny Internet. Do tych celów służą różnego rodzaju protokoły.

Protokoły powstawały równoległe z rozwojem Internetu i aplikacji służących tworzeniu metadanych i publikowaniu elektronicznemu. Wymienia się zazwyczaj dwa podstawowe, stosowane w GBC: protokół Z39.50, rozwijany na potrzeby bibliotek i zarządzany przez Library of Congress¹⁷ i jego późniejsze modyfikacje oraz protokół Open Archives Initiative (OAI), tworzony przez przedstawicieli środowisk związanych z obsługą dokumentów elektronicznych. OAI korzystał z doświadczeń pochodzących z realizacji projektu badawczego w Cornell University o nazwie Dienst. Warto wspomnieć także o protokole OpenSearch.

Standard Z39.50¹⁸ definiuje protokół o szerokim zastosowaniu, służący wyszukiwaniu informacji przez klienta w bazie danych posiadzonej na serwerze. Prace nad nim rozpoczęły się w 1984 r., a w 2003 r. zatwierdzono jego najnowszą wersję.

Z39.50 jest przykładem standardu na poziomie aplikacji z Open Systems Interconnection¹⁹ (OSI) Reference Model, stanowiącym podstawę dla łączenia komputerów w środowisku sieciowym. Za pomocą Abstract Syntax Notation One (ASN.1) wyspecyfikowane

¹⁷ <http://www.loc.gov/z3950/agency>.

¹⁸ ANSI/NISO Z39.50-2003 Information Retrieval (Z39.50): Application Service Definition and Protocol Specification oraz ISO 23950-1998 Information and Documentation – Information retrieval (Z39.50) – Application service definition and protocol specification.

¹⁹ Open Systems Interconnection (OSI) jest standardem ISO (ISO IEC 7498-1 Information technology – Open Systems Interconnection – Basic Reference Model) definiującym strukturę komunikacji sieciowej. Struktura ta składa się z siedmiu warstw: fizycznej, łącza danych, sieciowej, transportowej, sesji, prezentacji i aplikacji. Ta ostatnia jest najwyższą warstwą, dotyczącą specyfikacji interfejsu, który wykorzystują aplikacje do przesyłania danych do sieci poprzez kolejne warstwy modelu ISO OSI.

zostały formaty komunikatów, serializowane dla transmisji przez poziom transportu OSI przy zastosowaniu Basic Encoding Rules (BER). Dla bieżącej komunikacji informacyjnej wykorzystywany jest standardowo TCP (Transmission Control Protocol), podstawowy standard Internetu.

Protokół Z39.50 stanowi sposób na zapewnienie współdziałania w zakresie wyszukiwania w rozproszonych zasobach GBC. Tego rodzaju wyszukiwanie, nazywane sfederowanym (zob. p. 4.3), jest prowadzone między różnymi systemami, dzięki temu, że lokalny system informacyjny akceptuje zapytania z zewnętrznych systemów wyszukiwania informacji. W tym celu ujednoczona (a przy okazji uproszczona) jest struktura metadanych, akceptowana przez wszystkie współpracujące systemy.

Zapewnienie dostępu do heterogenicznych źródeł rozproszonych i ich wyszukiwania przez protokół w sposób umożliwiający współdziałanie stanowi istotne wyzwanie. Wykorzystanie Z39.50 w różnych środowiskach i zastosowaniach, takich jak dane bibliograficzne, informacja muzealna i archiwalna czy metadane geoprzestrzenne, ułatwia zestaw klas, zwanych rejestrami, zawierających struktury i atrybuty osobno uzgodnione dla każdego zastosowania. Rejestry opisują syntaktykę zapytań, atrybuty pól, formaty wyszukiwania treści oraz komunikaty diagnostyczne. Jako formaty wyszukiwania treści stosuje się na przykład Simple Unstructured Text Record Syntax (SUTRS) oraz różne warianty formatu MARC.

Przeszukiwanie wymaga określenia jego atrybutów, zazwyczaj za pomocą standardu Bib-1²⁰, stosowanego do wyszukiwania w bibliograficznych bazach danych. Bib-1 przyporządkowuje najbardziej typowym kryteriom zapytań (na przykład elementom metadanych typu: Nazwa osobowa, Tytuł, Nazwa wydawnictwa) wartości numeryczne. Tytułowi przydzielono na przykład wartość 4, a autorowi wartość 1004.

Trudności sprawiać może określenie sposobu prezentacji wyników. Informacja bibliograficzna jest zapisywana w wielu strukturach. Obecnie najbardziej rozpowszechniony jest zapis w ISO 2709²¹ (rekordy MARC), ale coraz większą popularność zyskuje prezentowanie wyników w języku XML. Niezależnie od samego formatu opisu, konieczne jest uwzględnienie sposobu kodowania znaków diakrytycznych, o czym była mowa na początku rozdziału. Niekiedy sam serwer Z39.50 dokonuje przekodowania w locie na własną stronę kodową, co wiąże się zazwyczaj z utratą części informacji. W przypadku rozbudowanego opisu katalogowego, zawierającego oprócz opisu bibliograficznego również informację o egzemplarzach, zdarza się, że część bibliograficzna jest kodowana inaczej niż część dotycząca egzemplarzy. Istnieje także możliwość odrębnego kodowania wyrażenia wyszukiwawczego. Serwery Z39.50 działające w Polsce prezentują pełną gamę możliwych rozwiązań [Wolniewicz 2002].

Implementacja protokołu Z39.50 okazała się w wielu przypadkach kosztowna i skomplikowana, a przy tym nie udało się wyjść z tym protokołem poza środowisko biblioteczne. Do aplikacji komercyjnych. W związku z tym Z39.50 Maintenance Agency w Library of Congress utworzyła dwa kolejne protokoły wyszukiwawcze: Search and Retrieve Web Service (SRW) oraz Search and Retrieve URL (SRU), które łącznie nazywane są SRU/W. Są one obsługiwane przez tę samą agendę Library of Congress, która odpowiedzialna jest za Z39.50, więc nie powinno być obaw o jakość i ciągłość jego rozwoju.

SRU/W, pomimo tego, że różnią się one od Z39.50 w wielu aspektach, wykorzystują doświadczenie zdobyte przy jego stosowaniu, wykazując szczególną przydatność w środowisku Web. Podobnie jak Z39.50, pozwalają na wyszukiwanie zbiorów, złożone zapytania, umożliwiając też zadawanie serwerom pytań o udostępniane przez nie zasoby.

²⁰ Bib-1 jest zestawem atrybutów służących do formułowania zapytań przez klienta, stosowanym podczas posługiwania się protokołem Z39.50.

²¹ ISO 2709 Information and documentation – Format for information exchange. Najnowsza wersja pochodzi z 2008 r.

Jednakże zapytania wysyłane przez klientów SRU/W mogą być odczytywane przez ludzi, podobnie jak odpowiedzi serwerów, wysyłane w XML. SRU/W opiera się na powszechnie stosowanych protokołach oraz jest znacznie prostszy niż Z39.50.

Główna różnica pomiędzy SRW i SRU polega na tym, że SRW służy komunikacji pomiędzy klientem i serwerem z zastosowaniem komunikatów SOAP²², natomiast SRU wykorzystuje URL, czyli komendę HTTP GET, w celu wysłania zapytania i otrzymania odpowiedzi w XML. Wynika z tego, że SRW jest bardzo podobne do SRU, z tym wyjątkiem, że ten pierwszy stosuje komunikaty SOAP, a ten drugi URL do wyszukiwania danych.

SRU/W wciąż nie jest jeszcze zbyt rozpowszechniony, nie został szeroko zaimplementowany w oprogramowaniach bibliotecznych. Jednak, ponieważ jego implementacja jest znacznie prostsza niż Z39.50, a w dodatku oferuje on wiele dodatkowych i lepiej działających funkcji, należy spodziewać się szybkiego rozwoju jego zastosowań szczególnie w środowisku Web; wykorzystywany jest m.in. w wyszukiwaniu sfederowanym²³ [LeVan 2006, s. 152] (zob. p. 4.3).

Otwarty protokół **OpenSearch**²⁴ w swoim działaniu podobny jest do SRU, ponieważ stosowany jest do wyszukiwania materiałów, które odpowiadają konkretnym potrzebom informacyjnym, pozwala klientom wysłać zapytania do serwera i realizuje wyszukiwania przy użyciu URL. Udostępnia on stronom Web i wyszukiwarkom sposób na prezentację wyników wyszukiwania w standardowym i dostępnym formacie. Różni się on od SRU tym, że został wykonany przez prywatną firmę (związaną z Amazon), jest prostszy do zaimplementowania i przesyła wyniki w RSS²⁵, dzięki czemu mogą one być odbierane przez czytnik wiadomości (*newsreader*). Protokół obsługuje wszystkie rodzaje mediów, w tym zapisy obrazu, dźwięku i wideo. Pomimo niewielkiej liczby realizowanych funkcji jego zastosowanie wśród dostawców oprogramowania gwałtownie rośnie. OpenSearch wykorzystywany jest przez takie instytucje i przedsiębiorstwa jak Amazon, National Library of Medicine w USA, wiele redakcji gazet na całym świecie oraz dostawców informacji, realizujących usługi w Webie. Stosowany jest on także w wielu blogach, serwisach i przez mniejszych dostawców usług informacyjnych.

Celem organizacji Open Archives Initiative²⁶ jest utworzenie standardu współdziałania zasobów Web, służącego rozpowszechnianiu informacji, co ma zasadnicze znaczenie w GBC. Efektem prac jest protokół **OAI-PMH** (Protocol for Metadata Harvesting), służący gromadzeniu metadanych z wielu źródeł (rozproszonych repozytoriów) w jeden, scalony zasób. Prace nad protokołem OAI rozpoczęte zostały przez twórców publikacji elektronicznych, którzy poszukiwali sposobów szybkiego rozwoju repozytoriów cyfrowych prac naukowych, zapewniających dostępność archiwizowanych obiektów. Jednym ze sposobów osiągnięcia tych celów było stworzenie protokołu niezależnego od udostępnianych treści, co czyniło go użytecznym w wielu zastosowaniach, nie tylko dla komunikacji naukowej.

²² SOAP (Simple Object Access Protocol), protokół utworzony przez W3C, oparty na technologii XML. Służy wymianie ustrukturyzowanej informacji w środowisku zdecentralizowanym i rozproszonym.

²³ To podejście jest wykorzystywane przez NISO Metasearch Initiative (<http://www.niso.org/workrooms/mi>), która promuje standard MXG, oparty na SRU/W. Metasearch XML Gateway (MXG) jest metodą udostępniania treści do metawyszukiwania i efektywnego współdziałania z aplikacjami metawyszukiwarek. Protokół MXG definiuje prosty komunikat i odpowiedź pozwalające metawyszukiwarce na wysyłanie zapytań do bazy danych i otrzymywanie standaryzowanych odpowiedzi w XML.

²⁴ <http://www.opensearch.org>.

²⁵ RSS – RDF Site Summary, really simple syndication usługa służąca do informowania o zmianach zachodzących na często modyfikowanych stronach Web. Dokument RSS zawiera skrót lub całość strony Web, z którą jest związany. Dzięki niej użytkownicy nie muszą ręcznie odwiedzać wielu interesujących ich stron w poszukiwaniu nowych treści.

²⁶ <http://www.openarchives.org>.

System zgodny z protokołem OAI może udostępnić swoje metadane narzędziom indeksującym, co pozwala na włączenie tych metadanych do wielkich baz danych i ich wykorzystywanie przez zewnętrzne serwisy wyszukiwawcze. Metadane pobierane są z lokalnych baz danych zgodnych z OAI i gromadzone w jednym miejscu, wspólnej bazie danych, w której są wyszukiwane (wyszukiwanie odbywa się już poza OAI-PMH). Może także służyć do importu/eksportu metadanych pomiędzy repozytoriami [Bell, Lewis 2006, s. 270]. Protokół opisuje interakcje pomiędzy dostawcą danych a dostawcą usług, co w nieco inny sposób nazywa obie strony modelu klient-serwer (zob. rozdz. 3), co oznacza, że w tym protokole interakcja kierowana jest przez klienta, który ma prawo decydowania o usługach oferowanych użytkownikom. Natomiast dostawcy danych zajmować się mają obsługą repozytoriów. Nie muszą oni prowadzić wyszukiwań opartych na dostarczonych terminach wyszukiwawczych; oczekuje się od nich przekazania na żądanie metadanych w standardowej formie, bez dochodzenia sposobów wykorzystania tych danych. Wyszukiwanie jest realizowane przez dostawcę usług, a nie dostawcę danych. Naturalnie można zdecydować, że w tym samym miejscu posadowiony będzie zarówno dostawca usług, jak i danych. Może on także obsługiwać więcej niż jedno repozytorium. Wszystkie repozytoria umożliwiają stosowanie metadanych Dublin Core, chociaż protokół przewiduje także użycie rekordów formatu MARC, kodowanych w XML.

Oba protokoły – Z39.50 i OAI – najważniejsze z punktu widzenia zapewnienia współdziałania zasobów GBC (zob. p. 4.6), reprezentują odmienną politykę realizacji tego celu. O ile system zgodny z protokołem OAI udostępnia swoje metadane narzędziom indeksującym, pozwalając na włączenie ich do wielkich, zewnętrznych baz danych, to Z39.50 umożliwia wyszukiwanie sfederowane; metadane pozostają w zasobach źródłowych, ale lokalny system informacyjny akceptuje zapytania z oddalonych systemów wyszukiwania informacji. Dlatego też nie jest łatwo porównywać oba rozwiązania. OAI-PMH jest prostszy z technicznego punktu widzenia, co ułatwia jego stosowanie przez twórców repozytoriów. Protokół Z39.50 przewyższa OAI poziomem funkcjonalności. Zarządza on na przykład sesją oraz zestawem wyników wraz z możliwością filtrowania otrzymanych rekordów. Twórcy OAI nie zamierzali zastępować swoim rozwiązaniem innych produktów. Protokół jest raczej alternatywą, umożliwiającą łatwą implementację i szersze zastosowanie dla różnych celów, często nie realizowanych przez dotychczasowe, współdziałające systemy. Istotną zaletą w tym zakresie jest umożliwienie indeksowania przez wyszukiwarki internetowe metadanych gromadzonych przy pomocy OAI-PMH, dzięki czemu nawet w przypadku, gdy same dane pozostają z różnych powodów w ukrytym Webie, przynajmniej ich metadane są w nim widoczne.

2.5. Obiekty cyfrowe

W poprzednich rozdziałach pisałem o danych, dostępnych w wielu formach i formatach, z których każdy musi dawać możliwość wykonywania wielu operacji: tworzenia, przechowywania, wyszukiwania i manipulacji na obiektach cyfrowych, materializujących informację. Wiele prostych obiektów cyfrowych ma bezpośrednią analogię z obiektami materializowanymi na nośnikach fizycznych; cała informacja gromadzona jest w jednym obiekcie, który można traktować jako pliki komputerowe o nieskomplikowanej strukturze [Arms 2000, s. 225]. Obecnie, w zasobach globalnej biblioteki cyfrowej, znajdują się znacznie bardziej złożone obiekty cyfrowe. Dostępne są tam wszelkiego rodzaju treści, a jedynym ograniczeniem jest konieczność posiadania przez nie formy cyfrowej, przy czym taką formę przybierają wszystkie rodzaje materiałów, tradycyjnie gromadzonych przez biblioteki oraz nowe, które wcześniej nigdy w bibliotekach się nie pojawiały. Nośniki cyfrowe umożliwiają udostępnianie w bibliotece cyfrowej takich obiektów, jak opro-

gramowanie komputerowe, symulacje, animacje, filmy, prezentacje, nagrania dźwiękowe o odmiennych strukturach w stosunku do analogowych poprzedników. Gromadzone są też obiekty będące efektem przetwarzania komputerowego, takie jak arkusze kalkulacyjne, bazy danych, dokumenty hipertekstowe i wiele innych. W sieciach komputerowych udostępniane są niekończące się strumienie danych cyfrowych, głównie tekstowych, a także, w coraz większym stopniu, multimediów. Składają się one na publikacje elektroniczne, przedstawione w następnym rozdziale.

W globalnej bibliotece cyfrowej, stanowiącej część IIN, dokumenty²⁷ przyjmują postać obiektów cyfrowych, zazwyczaj złożonych, komponowanych z obiektów prostych. Oprócz tzw. DLO (Document-like object), których zasadniczą treść stanowi tekst, a więc najbardziej przypominających dokumenty tradycyjne, gromadzone są coraz częściej i w coraz większej liczbie obiekty zupełnie nowego rodzaju, cechujące się multimedialnością i hipertekstowością.

Twórcy części obiektów cyfrowych nadal stosują konwencje przyjęte dla dokumentów drukowanych, upodobiłając je do tradycyjnych książek czy artykułów z czasopism. Na ekranie komputera wyglądają one w znany sposób; wyświetlane są nagłówki, indeksy, odsyłacze, podział na rozdziały i akapity. Taka forma typowa jest dla okresu przejściowego, po którym obiekty te w pełni wykorzystywać będą możliwości sieciowego środowiska cyfrowego. Dokonuje się modernizacja, która przez pewien czas polegała na dodawaniu do obiektów o tradycyjnej formie nowych cech, takich jak aktywne odnośniki, możliwość skalowania ilustracji, przeszukiwanie treści. Oznaczało to jedynie modernizację stałych i przewidywalnych wzorców komunikacji (przykładem mogą być zmiany, jakie przeszły komunikaty, rozsyłane przez uczonych kolegom w formie listów, do czego obecnie stosuje się pocztę elektroniczną). Później zaczęły powstawać formy typowe wyłącznie dla środowiska cyfrowego, co wiąże się z włączaniem przez społeczność uczonych nowych form obiektów cyfrowych w sferę swojej aktywności. Za takie można na przykład uznać udostępnianie treści kustomizowanych według potrzeb poszczególnych użytkowników oraz osobiste strony domowe uczonych.

Do obiektów cyfrowych rzadko stosowana jest, powszechnie obowiązująca dla dokumentów tradycyjnych, zasada „jedno dzieło, wiele egzemplarzy”. Do rozpowszechnienia dzieła w formie cyfrowej wystarczy jego istnienie tylko w jednej formie i w jednym miejscu. Można powiedzieć, stosując terminologię FRBR [IFLA 1998], że jego materializacja jest tożsama z egzemplarzem. Jednocześnie jednak może ono posiadać wiele wersji (realizacji), zmateriałizowanych albo w jednej, albo w wielu lokalizacjach. Obiekty mogą istnieć w wielu kopiach w różnych bazach danych. Z każdego z nich może jednocześnie korzystać nieograniczona liczba użytkowników, również zlokalizowanych w niemal dowolnym miejscu. Ponadto obiekty cyfrowe, w szczególności funkcjonujące w komunikacji naukowej, które opublikowane zostały w Internecie, w zasadzie wymagają istnienia tylko jednego egzemplarza, który jest udostępniany globalnie; model opisujący tą sytuację przedstawiłem w jednej ze swoich poprzednich książek [Nahotko 2006b, s. 138]. Zmiana dokonana w danym obiekcie cyfrowym powoduje powstanie takiej samej zmiany w każdym miejscu, w którym będzie on materializowany.

Istnieje kilka technik służących obiektywizacji tej różnorodności²⁸. Wymienić można

²⁷ Dokument, zgodnie z definicją w Słowniku encyklopedycznym, rozumiany jest jako utrwalona informacja wraz z nośnikiem informacji; dokumenty elektroniczne cechują się utrwaleniem na nośniku elektronicznym [Bojar (oprac.) 2002, s. 50]. Istotne jest rozróżnienie między dokumentem a obiektem cyfrowym. W tym przypadku dokument (np. artykuł) jest rozumiany jako całość logiczna, która może składać się z wielu obiektów fizycznych, powiązanych relacjami. Można mówić o architekturze dokumentu, określającej jego strukturę, tzn. podział na takie elementy, jak rozdziały, akapity.

²⁸ Jak twierdzi Alan Burk, przejście pomiędzy kolejnymi nośnikami informacji – od rękopisów do wersji drukowanych i od druków do wersji elektronicznych powoduje znaczny wzrost liczby różnych wersji dzieła dostępnych jednocześnie dla użytkownika. Efekt wywołany w tym zakresie digitalizacją publikacji można więc

identyfikatory obiektów cyfrowych, dane oraz metadane strukturalne, opisujące relacje pomiędzy złożonymi obiektami cyfrowymi i ich częściami. Łącznie techniki te pozwalają na stworzenie modelu obiektu, stanowiącego metainformacyjny opis, umożliwiający przechowywanie i udostępnianie złożonej informacji w systemach komputerowych. Omawiając te zagadnienia należy odnieść się do problemów współdziałania i długotrwałej archiwizacji obiektów cyfrowych. Informacja, obecnie funkcjonująca w globalnych zasobach, powinna być również możliwa do wykorzystania za wiele lat, za pomocą systemów komputerowych, które obecnie trudno sobie nawet wyobrazić. Tymczasem usunięcie obiektu z sieci, gdzie występuje w jednym egzemplarzu, może spowodować całkowitą i nieodwracalną utratę dostępu.

Wiele obiektów cyfrowych, stanowiących elementy zasobów globalnej biblioteki cyfrowej, nie może być reprezentowanych jako statyczne pliki zawierające dane. Są one charakteryzowane jako:

- Obiekty dynamiczne. Są to programy komputerowe, w tym gry komputerowe, aplety Javy, symulacje czy dane pochodzące z sensorów stosowanych w badaniach naukowych. Użytkownik, po uzyskaniu dostępu do obiektu, za każdym razem otrzymuje różne treści, więc podstawowego znaczenia nabiera wersjonowanie takich obiektów cyfrowych.
- Obiekty złożone. Obiekty cyfrowe w bibliotece globalnej mogą składać się z wielu elementów wzajemnie powiązanych licznymi i różnorodnymi relacjami. Występują komplementarne elementy treści (np. obraz graficzny i towarzyszący mu tekst), bądź alternatywne materializacje (np. ta sama treść obiektu graficznego w różnych parametrach technicznych, takich, jak rozdzielczość, nasycenie barwami), lub surogaty, jak dane i metadane. W praktyce różnice te są często nieostre. W zakresie obiektów multimedialnych problemy narastają, gdy na przykład trzeba decydować, czy miniaturka większej fotografii jest jej odrębną materializacją, czy raczej surogatem, rodzajem metadanych.
- Obiekty o wzmożonej identyfikacji zawartości. Wiele obiektów cyfrowych wciąż naśladuje swoich drukowanych przodków, czego następstwem jest traktowanie tych obiektów tak, jakby nie posiadały wewnętrznej struktury, były zamkniętymi jednostkami. W globalnej bibliotece cyfrowej każde słowo i identyfikowalne związki słów, stają się odrębnymi obiektami dyskretnymi. Prowadzi to do zwiększenia liczby punktów dostępu do obiektów cyfrowych przynajmniej o dwa rzędy wielkości.
- Obiekty cyfrowe o alternatywnych sposobach rozpowszechniania. Mogą oferować użytkownikom możliwość wyboru metody dostępu.
- Bazy danych. Bazy danych składają się z wielu rekordów, wybieranych pojedynczo i w kombinacjach za każdym razem, gdy użytkownik uzyskuje dostęp do jednej z nich. Stanowią bardzo zróżnicowane środowisko. Część z nich może być traktowana jako zasoby wielu obiektów cyfrowych. Inne, takie jak informatory, można traktować jako pojedyncze obiekty.

Jak wynika z powyższego wyliczenia, obiekty cyfrowe są dokumentami dynamicznymi. Ta cecha może być uznana za fundamentalną różnicę pomiędzy nimi a dokumentami drukowanymi. Dzięki oddzieleniu treści od nośnika, zmiany treści są łatwe do przeprowadzenia i nie powodują istotnych modyfikacji fizycznych lub technicznych charakterystyk obiektu. Poprawki wprowadzane do treści nie wpływają na nośnik, na którym informacja jest zapisana, przez co nie mogą być wykryte bez odniesienia do zewnętrznego wzorca (certyfikowanej kopii, sumy kontrolnej czy innych sposobów weryfikacji autentyczności – zob. p. 4.4).

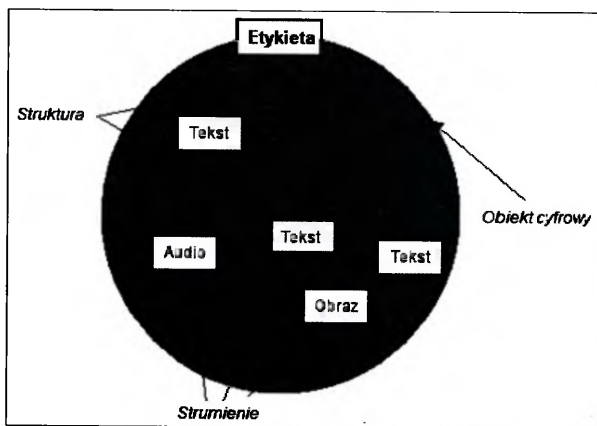
Dynamiczność obiektów cyfrowych nie oznacza, że stała zmienność jest ich konieczną cechą charakterystyczną. Chodzi raczej o dynamikę potencjalną. Jak twierdzi John Mackenzie Owen, obiekty cyfrowe są zmienne jedynie w takim sensie, że mogą przyjmować dowolny stan pomiędzy stałością i zmiennością, a stan ten w każdym momencie może

porównać do efektów spowodowanych rozwojem druku [Burk, Kerr, Pope 2000].

ulegać zmianie [Owen 2007, s. 139]. W rzeczywistości komunikacji naukowej większość obiektów cyfrowych pozostaje niezmienna. Chodzi jednak o to, że obiekty te z zasady, niejako potencjalnie są zmienne, w związku z czym można zastanowić się nad sposobami wykorzystania tej cechy w komunikacji naukowej, o czym będzie jeszcze mowa.

Jak już wspomniałem, we współczesnej komunikacji naukowej funkcjonują obiekty złożone, będące złożonymi jednostkami komunikacji naukowej. Są one agregacjami prostych jednostek informacyjnych, które po połączeniu tworzą logiczną całość [Sompel, Lagoze 2007, s. 32]. Agregacje mogą upodabniać obiekt złożony do dokumentu tradycyjnego, czego przykładem może być zdigitalizowana książka, która składa się z agregacji rozdziałów, z których każdy jest agregacją zeskanowanych stron. Obiekty złożone mogą wykorzystywać specyficzne cechy formy elektronicznej, tak jak w przypadku publikacji naukowej, która jest agregacją tekstu (artykułu) i materiałów pomocniczych, takich jak zestawy danych, narzędzia programistyczne lub zapisy wideo z realizacji eksperymentów. W efekcie obiekt złożony zawiera rezultaty wszystkich etapów procesu badawczego, w formie prostszych obiektów, wraz z opisującymi je metadanymi, co znacznie ułatwia m.in. ocenę, a także rozpowszechnianie i archiwizację obiektów cyfrowych [Hunter 2006, s. 34]. Obiekt taki przed opublikowaniem w Sieci stanowi odpowiednik realizacji według terminologii FRBR.

Nowe jednostki komunikacji naukowej, będące rezultatem funkcjonowania współczesnego, sieciowego środowiska badawczego, posiadają złożoną naturę, odmienną od natury materiałów tradycyjnych i ich bezpośrednich następców, prostych wersji cyfrowych, tworzonych na przykład w formatach PDF i LaTeX. Są one agregacjami części odmiennych pod względem rodzaju treści (jak artykuł, symulacja, nagranie wideo, zestaw danych, oprogramowanie), typu medium (tekst, obraz, dźwięk, wideo, mieszany), formatu (PDF, XML, MP3 i in.) oraz lokalizacji w sieci (różne części udostępniane przez różne repozytoria). Poza tym każdy agregat posiada związany z nim identyfikator. Powiązanie identyfikatora z obiektem złożonym następuje poprzez system informacyjny, w którym skomponowano agregat, tworząc logiczną jednostkę komunikacji naukowej. Takie agregaty nazywać można obiektami złożonymi.



Rys. 5. Złożony obiekt cyfrowy

Na rys. 5 przedstawiłem obiekty proste organizowane w złożone obiekty cyfrowe. Te ostatnie składają się ze zbioru ustrukturyzowanych elementów, oznaczonych wspólną nazwą – identyfikatorem (etykietą). Obiekt może składać się z pojedynczego elementu lub też z kilku, połączonych w jedną strukturę, tworząc obiekt złożony.

Obiekty złożone są podstawowymi elementami składowymi e-nauki, stanowiąc jeden z zasadniczych aspektów IIN [Cyberinfrastructure 2007, s. 33]. Model złożonego obiektu cyfrowego zaproponowany został przez Gregory Crane'a, który pisał o tzw. dokumentach rekombinowanych [Crane 2006]. Różnią się one od tradycyjnych i ich prostych kopii cyfrowych wieloma cechami. Obiekty umieszczone w bibliotece cyfrowej odczytują nawzajem swoją zawartość, porównując ją przed utworzeniem nowego obiektu, będącego syntezą wcześniej istniejących, który następnie przedstawiany jest użytkownikowi. Agregują nową informację z informacji istniejącej, której struktura znana jest na poziomie bitów i bajtów. Agregacje mogą być tworzone przez człowieka, autora obiektu złożonego, który na przykład otrzymał takie zadanie w ramach podziału zadań w naukowych grupach roboczych, lub też generowane maszynowo, na przykład w wyniku stosowania technik nauczania wspomaganego komputerowo lub surfowania w Webie. Agregacja powstałych wcześniej jednostek informacji w obiekty złożone (czyli ich ponowne wykorzystanie, *reuse*) nie jest wynikiem inherentnej natury jednostki zagregowanej, lecz działania algorytmu lub intencji człowieka, który skomponował obiekt złożony. Dodatkowo obiekty te mogą być agregowane i rozwijać się w czasie, na podstawie wzorców ich stosowania oraz działań społeczności użytkowników, dostarczającej dodatkowych kontekstów dla informacji zawartej w obiektach.

Złożony obiekt cyfrowy posiada wewnętrzną strukturę, odzwierciedlającą relacje, zachodzące pomiędzy elementami obiektu, opublikowanymi w Webie jako źródła. Relacje wyrażane są przez powiązania pomiędzy źródłami, realizowane w postaci odnośników. Są sekwencjami elementów określonego typu (bitów, znaków, obrazów). W tym znaczeniu mogą modelować zarówno treści statyczne, jak i dynamiczne. Przykładem tych pierwszych są dokumenty tekstowe, tych drugich prezentacja cyfrowego filmu wideo lub też sekwencja danych opisujących czas i pozycję w przestrzeni (np. z GPS), dotyczące obiektu ruchomego.

Źródła mogą być połączone między sobą odnośnikami albo wychodzącymi z elementu obiektu na zewnątrz, do innych źródeł lub łączącymi źródła opublikowane w Webie. Część elementów obiektu złożonego, na przykład wybrane pliki, może nie zostać opublikowana. Odnośniki stanowią podstawę rozbudowanego środowiska informacyjnego, jakim jest Web. Jednak standardowe hiperlinki, nie definiują relacji, zachodzących pomiędzy źródłami, będącymi elementami obiektu złożonego. Logiczna całość, którą jest obiekt złożony, jest dezintegrowana na zestaw odrębnych źródeł, rozpoznawalnych z pozycji innego źródła w Webie, a struktura tych obiektów pozostaje ukryta.

Twórcy systemów informacyjnych starają się ujawnić tę strukturę poprzez udostępnianie użytkownikowi specjalnego miejsca (strony Web), gdzie wyszczególniane są wszystkie elementy obiektu złożonego, wraz z odnośnikami do źródeł, z którymi związane są relacjami. Dzięki temu ukryta struktura obiektu złożonego, wyrażana przez wewnętrzne odnośniki, staje się jawna, bo opisana jest na tej stronie zbiorczej. Przedstawia ona strukturę całego obiektu złożonego, udostępnianego w Webie, więc, zgodnie z przyjętymi zasadami, URI tej strony należy stosować jako URI obiektu złożonego. Rozwiązanie to, przydatne z punktu widzenia użytkowników-ludzi, może być jednak problematyczne z punktu widzenia przetwarzania maszynowego.

Oprócz identyfikacji obiektu złożonego i definicji jego struktury, należy też rozwiązać problem braku metody w architekturze Webu na bezpośrednie odesłanie do źródła w obrębie obiektu złożonego. Ta funkcjonalność jest ważna dla komunikacji naukowej, ponieważ każde źródło może być wielokrotnie wykorzystane jako składnik dowolnej licz-

by obiektów złożonych. Możliwość odesłania do takiego źródła jest istotna ze względu na potrzebę rejestracji historii obiektu (proweniencji) i umożliwienie jego cytowania, gdyż konkretne znaczenie źródła lub odesłania do niego, może zależeć od kontekstu narzucającego przez ten obiekt.

Rozwiązaniem tego problemu może być infrastruktura zaproponowana przez Herberta Sompela i Carla Lagoze, zbudowana ponad architekturą Webu, pozwalająca na wielokrotne wykorzystywanie złożonych obiektów informacyjnych i odsyłanie do nich, z zachowaniem współdzielenia i możliwości zastosowania w różnych aplikacjach sieciowych [Sompel, Lagoze 2007, s. 36], w tym w standardzie OAI-ORE²⁹. Prace te czerpią swoją inspirację z wyników działań twórców i ideologów Semantycznego Webu (zob. p. 4.9), co nie oznacza jednak potrzeby stosowania jego technologii, takich jak RDF/XML czy OWL. Ich podstawą są tak zwane mapy źródeł (*resource map*). Opisują one obiekt złożony (zagregowany), w tym rodzaj zagregowanych źródeł i relacje między nimi, na przykład typu 'isVersionOf' lub 'isPartOf'. Każda mapa źródeł określa tylko jedną agregację, jednak element obiektu zagregowanego może należeć do więcej niż jednej agregacji, a wszystkie opisywane są przez odpowiednie mapy źródeł.

2.6. Publikacje elektroniczne

Zasób informacyjny GBC – publikacje elektroniczne powstają w wyniku procesów cyfryzacji i digitalizacji, omówionych w p. 4.1. W tym rozdziale przedstawiamy podstawowe rodzaje publikacji oraz ich pochodzenie. Istotne także wydają się zagadnienia związane z nowymi cechami publikacji elektronicznych i ich przemianą od publikacji naśladujących wydawnictwa tradycyjne do nowych, nie spotykanych wcześniej, rodzajów materiałów. Jak to wyjaśniliśmy w poprzednim rozdziale, publikacja elektroniczna powstaje po umieszczeniu (opublikowaniu) wszystkich jej agregatów w Webie i powiązaniu ich relacjami³⁰. Obiekty niepublikowane stanowią rodzaj realizacji według terminologii FRBR. Można więc stwierdzić, że publikacja powstaje jako agregat obiektów prostych w złożony obiekt cyfrowy³¹, umieszczony w Sieci, co stanowi odpowiednik materializacji w terminologii FRBR. W środowisku GBC publikacje nadal mogą być agregowane w zestawy publikacji; najbardziej popularne to czasopisma elektroniczne, biblioteki cyfrowe i różnego rodzaju repozytoria. Może też zachodzić zjawisko odwrotne; na przykład nowe sposoby organizacji dostępu do treści naukowych powodują oddzielenie artykułów naukowych od agregatów (czasopism naukowych), w których były tradycyjnie zamieszczane.

Sieciowe obiekty cyfrowe, opisane w poprzednim podrozdziale i publikacje elektroniczne są to dwa aspekty tego samego bytu; mówiąc o obiekcie mamy zazwyczaj na myśli jego cechy techniczne, a mówiąc o publikacji – jej cechy funkcjonalne (użyteczność). W tym podrozdziale zajmę się tym drugim aspektem: opublikowanymi obiektami cyfrowymi (zazwyczaj złożonymi), służącymi zaspokajaniu informacyjnych potrzeb użytkowników.

Publikacje elektroniczne różnią się znacznie od drukowanych wieloma, istotnymi cechami. Decyduje o tym omówiona w poprzednim rozdziale specyfika tworzących je

²⁹ Open Archives Initiative-Object Reuse and Exchange (OAI-ORE) pozwala na opis i wymianę złożonych obiektów cyfrowych pomiędzy różnymi repozytoriami (nawet o odmiennej architekturze). Może on pełnić rolę interfejsu pomiędzy repozytorium i innymi serwisami, działającymi na przykład w celu ułatwienia umieszczenia, wymiany, wizualizacji czy archiwizacji obiektów cyfrowych (<http://www.openarchives.org/ore/>). Standard ten wbrew podobieństwu nazwy nie jest częścią protokołu OAI-PMH, jego rozwinięciem, ani go nie zastępuje.

³⁰ Pomijam tu publikowanie obiektów cyfrowych na nośnikach o dostępie lokalnym, typu CD-ROM, gdyż są one zazwyczaj prostym przeniesieniem publikacji tradycyjnych na nośnik cyfrowy.

³¹ W szczególnym przypadku publikacja może składać się tylko z jednego obiektu cyfrowego.

obiektów cyfrowych. Publikacje te są znacznie bardziej zmienne, niż druki; do ich rozpowszechniania wystarczy tylko jeden egzemplarz umieszczony w sieci, a dokonywane w nim zmiany powodują utratę wersji oryginalnej. Możliwe jest także wykonywanie idealnie dokładnych kopii, które mogą być szybko i szeroko rozpowszechniane. Łatwość ta może nawet sugerować traktowanie publikacji elektronicznych jako towaru o zerowym koszcie wytworzenia [Gawrysiak 2008c, s. 62]. Obiekty mogą być modyfikowane bez pozostawienia jakichkolwiek śladów tych czynności. Można także na bieżąco uzupełniać oryginalne treści o przydatne dodatkowe materiały (na przykład dane z badań empirycznych, suplementy, załączniki, uaktualnienia, komentarze, erraty), nawet po jego opublikowaniu i rozpowszechnieniu.

Zalety publikacji elektronicznych wynikają zazwyczaj z zastosowania nośnika cyfrowego oraz udostępniania sieciowego. Są to:

- Łatwość modyfikacji i dostosowania do indywidualnych potrzeb i upodobań takich elementów technicznych, jak kolor, jakość i rozdzielczość grafiki, wysokość i krój czcionki.
- Większa dostępność, nie ograniczana przez zdarzenia takie, jak zagubienie egzemplarza czy odesłanie go do oprawy, a także brak potrzeby dostosowywania się do godzin pracy jakiegokolwiek instytucji; każdy serwis GBC czynny jest w trybie 24/7, materiały dostarczane są wprost do domu (czy innego miejsca użytkowania sprzętu komputerowego).
- Zastosowanie skomplikowanych technik wyszukiwania informacji, na przykład łączenia wielu kryteriów wyszukiwawczych.
- Stosowanie technik multimedialnych, łączenie mediów, takich jak obraz i dźwięk, obiekty trójwymiarowe, w jednej prezentacji, co znajduje coraz szersze zastosowanie w tekstach naukowych.
- Zapewnienie bezpośredniego dostępu do materiałów, pozostających w relacjach (na przykład dokumenty, w których następują wzajemne cytowania), bez względu na format i lokalizację.
- Ułatwiony kontakt między autorem a odbiorcą jego dzieła.
- Wzrost wskaźników dotyczących poziomu wykorzystania dokumentów elektronicznych (w tym cytowań), szczególnie dostępnych w trybie OA [Gargouri i in. 2010].
- Stosowane mogą być ułatwienia dla osób niedowidzących i niewidomych, poprzez znaczne powiększenie tekstu lub zastosowanie wersji dźwiękowych, dzięki czemu dokument może być użytkownikowi przeczytany.
- Publikacje cyfrowe mogą być tańsze niż drukowane odpowiedniki.
- Nie ulegają one uszkodzeniom w trakcie użytkowania (uszkodzeniu może ulec sprzęt).

Rozproszone zasoby danych i informacji GBC są bardzo zróżnicowane, gdyż uzależnione od takich czynników, jak potrzeby użytkowników i polityka instytucji, które je tworzą. Z punktu widzenia użytkownika jej zasoby obejmują zarówno materiały tworzone w bardzo odległych miejscach (często wymagające spełnienia pewnych warunków dostępu, na przykład rejestracji), jak i te, które powstały lokalnie w macierzystej instytucji, w lokalnym języku, zaspokajające lokalne potrzeby. Część z tych zasobów jest dostępna bezpłatnie dla każdego, podczas gdy inne ze względu na opłaty czy licencjonowany dostęp będą służyły tylko określonym, uprawnionym użytkownikom. Mogą stanowić także zasób o zastrzeżonym dostępie. Stanowią one zasoby bardzo zróżnicowane, zarówno pod względem zakresu tematycznego, przeznaczenia czytelniczego oraz formatów i języków, chociaż w zasobach internetowych dominujący jest język angielski. Różnorodność jest jedną z głównych przyczyn problemów technicznych, występujących w GBC, a także działań służących ich rozwiązywaniu, takich jak standaryzacja.

W podrozdziale tym opiszę wybrane rodzaje publikacji elektronicznych, źródeł informacji dostępnych w internetowych zasobach globalnej biblioteki cyfrowej i sposobów

ich udostępniania w tym środowisku. W globalnej bibliotece cyfrowej udostępniane są wszelkie publikacje, znane z bibliotek tradycyjnych oraz nowe rodzaje prezentacji treści. Zasoby te podzieliłem na kilka grup:

- Dokumenty tekstowe, takie jak czasopisma i książki elektroniczne oraz tradycyjnie zaliczane do szarej literatury, jak materiały konferencyjne, raporty i dysertacje, które często agregują treści na wzór drukowanych pierwowzorów.
- Zasoby Webu 2.0, nie znajdujące bezpośrednich poprzedników w świecie druku.
- Archiwa list dyskusyjnych.
- Nowe sposoby agregacji treści, głównie repozytoria i biblioteki cyfrowe.
- Strony Web, które mogą być traktowane jako odrębne publikacje lub agregaty publikacji (na przykład strona Web wydawcy czasopism, zawierająca ich treść).
- Materiały multimedialne.
- Zbiory danych (*data sets*).

Oprócz wymienionych zasobów danych istnieją obszerne zasoby metadanych, w szczególności tworzone przez bibliotekarzy, opisujące zbiory bibliotek, zarówno tradycyjnych, jak i cyfrowych. Zbiory te liczą dziesiątki milionów rekordów o standardowych strukturach. Obok nich funkcjonują zasoby indeksowe wyszukiwarek.

Na rolę publikacji cyfrowych można patrzeć nie tylko z punktu widzenia problemów technicznych, ale także pod kątem znaczenia tych publikacji dla członków społeczności, którzy je tworzą i używają. Z tego punktu widzenia publikacje funkcjonują w „życiu społecznym” [Borgman 2003a, s. 94]. Publikacja naukowa powstaje jako wynik negocjacji, prowadzonych w zespole przygotowujących ją autorów; dotyczą one treści i sposobu ich prezentacji, wyboru formatu publikacji i ewentualnych załączonych danych, interpretacji wyników oraz autorstwa. Użytkownicy publikacji cyfrowej nie interpretują jej w sposób wyizolowany, lecz porównują z innymi, w stosunku do których jest ona kontynuacją lub z którymi stoi w sprzeczności. Na interpretację tę wpływają także czynniki zewnętrzne, takie jak reputacja wydawnictwa, autora, instytucji go zatrudniającej oraz finansującej badania.

Elektroniczne, naukowe **dokumenty tekstowe** wciąż przypominają swoich drukowanych poprzedników. Są to głównie czasopisma oraz książki elektroniczne, dla udostępniania których tworzone są repozytoria i archiwa. Na ogół, oprócz zwykłego umieszczenia tekstu dokumentu (obiektu cyfrowego) w Internecie, co umożliwia dostęp do niego, wydawca oferuje dodatkowe usługi, polegające na tworzeniu i udostępnianiu metadanych, własnych interfejsów obsługujących wybrane formaty, indeksowaniu treści (stwarza to warunki do przeszukiwania zasobów). W ten sposób powstaje serwis GBC, wyodrębnione miejsce w Internecie służące udostępnianiu treści, jednocześnie jednak zintegrowane z innymi serwisami GBC.

W środowisku cyfrowym zacieraniu ulegają różnice pomiędzy podziałami, stworzonymi na podstawie tradycyjnego kryterium zamiaru kontynuacji publikacji³². Zasoby GBC oparte są na obiektach informacyjnych, które mogą przybierać różne formy (często zmienne i dynamiczne) i formaty [Owen 2000]. Różnice pomiędzy nimi ulegają zatarciu. W tym sensie rację mają ci, którzy twierdzą, że wszystkie zasoby internetowe zaliczyć można do szarej literatury [Carvalho 2001]. W efekcie zanikają różnice pomiędzy obszernym artykułem, do którego dołączono zestawy danych z eksperymentów, raportem technicznym lub naukowym i książką elektroniczną. Zacierają się także różnice między komunikacją naukową formalną i nieformalną, gdy na przykład ten sam tekst publikowany jest w czasopiśmie, repozytorium i na stronie Web uczonego, a film rejestrujący jego odczytanie, ilustrowane slajdami, na YouTube.

³² Efektem zastosowania tego kryterium jest podział na książki i czasopisma, czy raczej, używając terminologii fachowej, na wydawnictwa zwarte i ciągłe.

Według szacunków istnieje ok. 25 tys. tytułów naukowych czasopism recenzowanych, w których publikuje się ok. 1,4 mln artykułów rocznie [Björk, Roos, Lauri 2008, s. 187]. Obecnie można powiedzieć, że praktycznie każde czasopismo naukowe, tworzone w najbardziej rozwiniętych krajach świata dostępne jest online. Ma to tym większe znaczenie, że jak wynika z badań prowadzonych w USA, uczeni, w przypadku każdej podejmowanej działalności naukowej, korzystają z artykułów w czasopismach (w 94% przypadków, w porównaniu z 47% dla książek, 34% dla stron Web i 24% dla materiałów konferencyjnych). W 62% przypadków jest to artykuł z czasopisma elektronicznego [Tenopir i in. 2009, s. 17]. Większość naukowych **czasopism elektronicznych** to nadal kopie drukowanych poprzedników, chociaż zawartość obu wersji może się różnić, na przykład wersja elektroniczna może zawierać materiały trudne do wykorzystania w formie drukowanej ze względu na format (multimedia) albo zbyt dla niej obszerne. Bardzo powolna zmiana formy czasopism elektronicznych, a co za tym idzie również niepełne wykorzystywanie możliwości nowego medium, może wynikać nie tylko z braku pomysłowości wydawców czasopism, ale także z obawy przed uznaniem publikacji w nowej formie za mniej wartościowe. Według Heleny Francke upodabnianie czasopism elektronicznych do ich drukowanych poprzedników (w czym pomaga stosowanie takich formatów, jak PDF) wynika w dużej mierze z chęci przejęcia części ustalonej reputacji czasopism tradycyjnych przez ich elektronicznych następców [Francke 2008, s. 319].

Dostęp do treści publikacji komercyjnych możliwy jest albo bezpośrednio u wydawcy publikacji, stosującego model własnościowy: pobiera bezpośrednio opłaty za odczyt pojedynczych artykułów (bez potrzeby założenia prenumeraty), albo przez bibliotekę prenumerującą dany tytuł (dostęp dla zarejestrowanych użytkowników biblioteki, identyfikowanych IP komputera lub hasłem). Biblioteka zwykle nie posiada czasopisma elektronicznego na własność, tak jak egzemplarza drukowanego, lecz płaci jedynie za licencję na korzystanie z dostępu do platformy wydawcy (czy agregatora) w określonym czasie. Po tym okresie dostęp do archiwalnych zeszytów czasopisma może zostać przerwany; obawa o utratę dostępu do archiwum jest głównym powodem utrzymywania prenumeraty na czasopisma drukowane [Miller, Harris 2009, s. 19]. Według Chrisa Armbrustera takie praktyki są w konflikcie z logiką funkcjonowania komunikacji naukowej i wymogami globalnie zintegrowanej nauki [Armbruster 2008c]. Pomimo tego oraz innych zagrożeń, związanych na przykład z tzw. zjawiskiem Big Deal³³ coraz więcej bibliotek rezygnuje całkowicie z zakupu wersji drukowanej, poprzestając na publikacji elektronicznej³⁴.

Próbą przeciwstawienia się komercjalizacji komunikacji naukowej są inicjatywy środowiska naukowego, oparte na rozwiązaniach niewłasnościowych, stanowiące nawrót do wcześniejszego okresu wydawania czasopism przez grupy współpracujących uczonych. Takim projektem jest wspomniany już Public Library of Science (PLoS), stanowiący pró-

³³ Big Deal polega na oferowaniu bibliotekom i ich konsorcjom, przez wielkich wydawców, całych pakietów czasopism i baz danych (czasem także innych materiałów), w wersji elektronicznej, na podstawie wieloletnich umów. Za zakup pakietu zwykle oferowane są znaczne zniżki. W efekcie biblioteki rezygnują z oferty innych, małych wydawców [Frazier 2001].

³⁴ Jednym z wielu przykładów rozwoju tej tendencji może być Hanford Technical Library, część Departament of Energy (DoE) w USA. Biblioteka prenumeruje ponad 8 tys. tytułów czasopism online i 30 czasopism drukowanych. Na prenumeratę czasopism elektronicznych wydawane jest 70% budżetu przeznaczanego na zbiory; dalsze 26% przeznaczane jest na inne źródła elektroniczne. Jak z tego wynika, na publikacje drukowane przeznaczają się 4% funduszy przeznaczonych na zakup zbiorów [Noonan, McBurney 2007, s. 151]. Podobne procesy realizowane są w bibliotece Drexel University, gdzie zrezygnowano z drukowanych wersji wszystkich czasopism mających odpowiedniki elektroniczne, zarówno w postaci pełnotekstowych baz danych, jak i e-czasopism. Biblioteka ta udostępniła ok. 9 tys. tytułów czasopism, w tym 8600 tytułów elektronicznych i 370 drukowanych [King, Montgomery 2002]. Od początku XXI w. podobne tendencje widoczne są też w polskich bibliotekach uczelnianych, gdzie gwałtownie spadają wydatki na czasopisma drukowane na rzecz źródeł elektronicznych, które od 2006 r. stanowią największą grupę wydatków na zbiory [Derfert-Wolf, Górski, Jazdon 2009].

bę publikowania wysokiej jakości, recenzowanych czasopism naukowych Open Access bez udziału wydawców komercyjnych (zob. p. 4.1). Jak dotychczas inicjatywa ta rozwija się pomyślnie, służąc wydawaniu kilku tytułów z zakresu nauk biologicznych i medycznych. Liczbę recenzowanych czasopism naukowych OA szacuje się na 3-4 tys. tytułów, w zależności od źródła [Ware, Mabe 2009, s. 20].

Czasopismo funkcjonujące w otwartym środowisku elektronicznym przestało być podstawową jednostką komunikacji naukowej. Stał się nią artykuł naukowy, który rozpoznać indywidualnie życie, często w oderwaniu od jakiegokolwiek czasopisma. Tendencja ta powoduje rozwój repozytoriów i archiwów, w których umieszczane i przechowywane są głównie artykuły (opublikowane w czasopismach lub nie), chociaż często zdarzają się także inne rodzaje dokumentów.

Od lat siedemdziesiątych ubiegłego wieku obserwować można rozwój elektronicznego publikowania, którego efektem były pierwsze **książki elektroniczne** (e-książki). Wykorzystanie e-książek rozwija się znacznie wolniej niż czasopism elektronicznych. Największą trudnością, tylko częściowo występującą w przypadku czasopism elektronicznych, jest stworzenie rozwiązań technicznych umożliwiających długotrwałe czytanie obszernych tekstów na ekranie monitora; pierwsze komputery osobiste zupełnie nie dawały takich możliwości. Na przełomie XX i XXI wieku powstało wiele przedsiębiorstw udostępniających e-książki, oferowano różne modele czytników, pozwalających wyświetlać strony tekstów na poręcznych urządzeniach. Po początkowych kłopotach z rozpowszechnieniem tych czytników, gdy powrócono do stosowania bardziej uniwersalnych urządzeń, korzystając z gwałtownie rosnącej popularności komputerów przenośnych, obecnie następuje bardzo szybki rozwój tego sprzętu³⁵. Książki elektroniczne oferowane są w coraz większym wyborze przez wydawnictwa naukowe w ten sam sposób, jak czasopisma, gdyż problem dedykowanych czytników dla publikacji naukowych nie jest tak istotny jak dla innej literatury. Inną możliwością jest druk e-książki na życzenie; część bibliotek publicznych świadczy takie usługi.

Innym problemem, występującym jeszcze wyraźniej, niż dla czasopism elektronicznych, są zagadnienia prawa autorskiego (zob.p. 5.3). Wydaje się także, że rola książki jako publikacji naukowej w środowisku elektronicznym maleje na rzecz innych form publikowania, które umożliwiają szybsze i sprawniejsze rozpowszechnianie wyników najnowszych badań naukowych. Dotyczy to nawet nauk społecznych i humanistycznych, w których książka najdłużej przetrwała jako „złoty standard”, służący uzyskiwaniu stopni i promocji naukowych [Estabrook, Warner 2003].

Dla rozpowszechnienia e-książek istotne jest przekroczenie „masy krytycznej”, zapewniającej wystarczający wybór. Książki elektroniczne stanowią poważny element biblioteki cyfrowej w zakresie udostępniania dzieł literackich. Część tych zasobów dostępna jest bezpłatnie. Jednym z najstarszych i najlepiej znanych jest projekt Gutenberg³⁶, udostępniający w wersji elektronicznej ponad 20 tys. książek z podziałem na literaturę popularną, poważną i wydawnictwa informacyjne (encyklopedie, słowniki). Funkcjonowanie projektu opiera się na wolontariacie – ochotnicy przygotowują elektroniczne wersje wybranych dzieł. Dla wszystkich utworów, udostępnianych w tym projekcie, wygasły prawa autorskie, przez co mogą być użytkowane bezpłatnie. Na stronach projektu propo-

³⁵ Ostatnio nadzieje wiąże się z czytnikami rozpowszechnianymi przez Amazon (Kindle), Apple (iPhone) i Google (Android GI). Postęp w tym zakresie jest bardzo duży; nowe modele urządzeń pojawiają się co pół roku. Sukces zapewnić może wszechstronność nowych urządzeń; są to nie tylko czytniki, ale również odtwarzacze multimedialne i urządzenia telekomunikacyjne. Według niektórych autorów rozpowszechnienie czytników zapewniłoby ich rozdawanie (sprzedaż za symboliczną kwotę) i uzyskiwanie zysków z udostępniania treści, co jest częstą praktyką na rynku mediów i urządzeń elektronicznych (telefon komórkowy za darmo – opłaty za rozmowy; drukarka bardzo tania – opłaty za toner, czyli wydruki) [Gołębiowski 2009, s. 28].

³⁶ <http://www.gutenberg.org>.

nuje się dostęp do publikacji nie tylko na ekranach komputerów, ale także przy pomocy smartfonów i palmtopów (PDA). Inny sposób digitalizacji przyjęty został w projekcie Google Books³⁷, gdzie publikacje drukowane, znajdujące się w bibliotekach są skanowane. Umowa z organizacjami wydawców pozwala na skanowanie publikacji objętych prawem autorskim. Teksty mają być udostępniane na czytnikach firm Apple i Google. To gigantyczne przedsięwzięcie skłoniło innych do realizacji wielu mniejszych projektów³⁸, z różnym efektem.

E-książki nie są odpowiednie dla każdego rodzaju materiałów i dla wszystkich użytkowników. Część autorów, jak na przykład Pam Saunders uważa, że wartość użytkową mają głównie publikacje przeznaczone do wykorzystania na standardowych PC [Saunders 2002]. Píše ona, że e-książki najlepiej nadają się dla treści informacyjnych i nie literackich, jak na przykład materiały dydaktyczne, które ze swojej natury nie są przeznaczone do czytania „od deski od deski”. Książki elektroniczne rzeczywiście często stosowane są w dydaktycznej działalności uczelnianej. W ten sposób udostępniane bywają podręczniki, publikowane wyłącznie w formie elektronicznej lub obok wersji tradycyjnej. Forma cyfrowa przydatna jest dla wszelkiego rodzaju publikacji informacyjnych, takich jak encyklopedie, słowniki, almanachy.

Wraz z rozpowszechnieniem się książki elektronicznej produkcja wydawnicza publikacji tradycyjnych nie ulega zmniejszeniu, w wielu krajach nawet rośnie. Być może oba rynki są w jakiś sposób komplementarne. Przemysł książek elektronicznych zorganizowany jest wokół International Digital Publishing Forum³⁹ (wcześniej znane jako Open eBook Forum), międzynarodowej organizacji handlu i normalizacji. Jest ono między innymi twórcą standardów, opartych na XML: Open Publication Structure (OPS), Open Packaging Format (OPF) oraz Open Container Format (OCF). Pozwalają one wydawcom produkować i udostępniać publikacje cyfrowe współdziałające z wieloma platformami programowymi i sprzętowymi.

W rozwoju e-książek można znaleźć wiele analogii do historii czasopism elektronicznych, które również początkowo zdobywały rynek powoli, z tych samych powodów, które dotąd hamują rozpowszechnienie się książek elektronicznych. Po likwidacji barier i zmianie przyzwyczajęń czytelnicych dojdzie zapewne do znacznego przyspieszenia tego rozwoju [Delany, Landow 2008, s. 85]. W bliskiej przyszłości, po upowszechnieniu e-książek, można spodziewać się bardziej ścisłego powiązania pomiędzy książkami i czasopismami elektronicznymi, nie tylko poprzez odsyłacze bibliograficzne, ale także pomiędzy publikacjami bliskimi tematycznie. Również wydawcy coraz częściej przedstawiają ofertę łączoną. Zapewne możliwy będzie zakup książki z usługą regularnej aktualizacji treści (wydawnictwa informacyjne, dydaktyczne).

Szara literatura, jako grupa publikacji, ciągle nie doczekała się jednoznacznej definicji. Najczęściej przyjmuje się, że jest to literatura tworzona przez wszystkie poziomy instytucji rządowych, naukowych, przemysłu i biznesu w dowolnej formie (drukowanej i elektronicznej), nie publikowana przez wydawnictwa komercyjne (czyli takie, dla których działalność wydawnicza jest podstawą funkcjonowania). Zdarza się jeszcze inne rozumienie tego terminu, na przykład jako literatury niedostępnej dla wyszukiwarek, czyli ukrytego Webu (definicja ze strony Web National Library of Australia⁴⁰). W zakresie komunikacji naukowej szara literatura zaliczana bywa do medium służącego komunikacji nieformalnej, nie jest to jednak regułą. Tritne Søndergaard zaliczyła szarą literaturę do dokumentów przekazywanych kanałami komunikacji formalnej (wraz z czasopismami

³⁷ W Polsce dostępny jako Google Książki: <http://books.google.pl>.

³⁸ Najbardziej znane, to: Światowa Biblioteka Cyfrowa (<http://www.wdl.org/>), Million Book Project (<http://www.ulib.org/>), Europeana (<http://www.europeana.eu/>), Gallica (<http://gallica.bnf.fr/>).

³⁹ <http://www.idpf.org>.

⁴⁰ <http://www.nla.gov.au/padi/topics/372.html>.

elektronicznymi i preprintami), w odróżnieniu od takich kanałów komunikacji nieformalnej, jak poczta elektroniczna, grupy dyskusyjne i konferencje internetowe [Søndergaard, Andersen, Hjørland 2003, s. 295]. Wynika to z tego, że wraz ze wzrostem zastosowań komputerów w nauce, nastąpiło przesunięcie Internetu z medium traktowanego jako nieformalny kanał informacji, do medium będącego ważnym kanałem komunikacji formalnej. Na przykładzie **list dyskusyjnych** można jednak stwierdzić, że dla mediów elektronicznych tego typu podziały nie są jednoznaczne. Komunikat wystany na naukową listę dyskusyjną może być przekazywany do wielu innych grup, a także dostępny w archiwum danego forum przez dowolnie długi czas, jednak niewielu specjalistów uzna go za medium komunikacji formalnej. Mimo tego, zapis na przykład nieformalnej dyskusji prowadzonej w ramach kolaboratorium może mieć duże znaczenie dla nauki. Możliwość zachowania takiej wymiany myśli wywołuje pytania o jej status w komunikacji naukowej; czy jej wartość uzasadnia inwestycje w archiwizację i długotrwałe udostępnianie.

Do szarej literatury zalicza się na ogół wszelkiego rodzaju raporty (z prac naukowych, techniczne, z odbytych rozmów), materiały robocze, dokumenty biznesowe, doktoraty i materiały konferencyjne (w szczególności niepublikowane). Wraz z coraz częstszym korzystaniem przez młodych pracowników nauki z wszelkiego rodzaju zasobów Internetu (zob. p. 5.1), jako środki zaliczane do nieformalnej komunikacji naukowej pojawiły się także blogi⁴¹, osobiste strony Web, podcasty⁴², vodcasty⁴³, filmy na YouTube oraz strony typu Wiki i Web konferencje [Kajewski 2007]. Większość z nich zaliczyć należy do zasobów **Webu społecznościowego**. Cechą charakterystyczną tych serwisów jest powstawanie dzięki inicjatywom oddolnym badaczy (niskie zinstytucjonalizowanie) oraz wysoka interakcyjność – treści powstają w wyniku współpracy zainteresowanych osób. W odróżnieniu od wcześniejszych form publikowania elektronicznego, które były odmianą publikacji tradycyjnych (na przykład poczta elektroniczna zastępowała pocztę tradycyjną), nowe sposoby naukowej komunikacji nieformalnej są specyficzne dla środowiska cyfrowego. Z komunikacji nieformalnej mogą one trafiać do oficjalnego obiegu, tak, jak blog biblioteczny lub Wiki projektu naukowego, realizowanego przez instytucje naukowe.

Jedną z najczęściej wymienianych cech szarej literatury jest trudny dostęp do tego typu dokumentów. W porównaniu z drukowanymi dokumentami szarej literatury, dostęp do materiałów elektronicznych, umieszczonych w Internecie jest znacznie łatwiejszy. Ten rodzaj literatury w Sieci może realizować funkcje komunikacji naukowej bardziej efektywnie, niż publikacje tradycyjnie, rozpowszechniane kanałami komunikacji formalnej. Preprint w repozytorium elektronicznym można uznać, ze względu na tempo jego udostępnienia, za bardziej efektywny środek komunikacji naukowej niż czasopismo naukowe. Przykład ten dodatkowo wskazuje na tendencję przejawianą przez część autorów (jak choćby cytowana już Elizabeth Carvalho i P. Rajendiran [Rajendiran 2006]), którzy dużą część zasobów Internetu uważają za zbiory szarej literatury.

Uczeni samorzutnie tworzą nieformalne kanały komunikacji w sytuacji, gdy kanały komunikacji formalnej nie potrafią zaspokoić ich potrzeb informacyjnych. Powstałe w ten sposób nowe media ewoluują w czasie, w wyniku czego następuje modyfikacja zasad obowiązujących w nauce [Meyer, Schroeder 2008]. Na ogół kanały informacji for-

⁴¹ Blogi traktowane mogą być jako rozwinięcie koncepcji listy dyskusyjnej, której moderator z osoby anonimowej, funkcjonującej za kulisami stał się centralną postacią blogu, wokół której, jej opublikowanych poglądów organizowana jest ta forma wypowiedzi. Być może jest to jeden z przejawów tendencji do likwidacji anonimowości w komunikacji między użytkownikami sieci społecznych.

⁴² Podcast, podkast – forma internetowej publikacji dźwiękowej lub filmowej, najczęściej w postaci regularnie dodawanych części, rozpowszechniana z użyciem technologii RSS. Nazwa wzięła się z połączenia słów iPod – odtwarzacz muzyczny firmy Apple i *broadcast* (z ang. transmisja, przekaz). Przekazywane są treści wszelkiego rodzaju: podcast może mieć formę nieformalnego bloga, profesjonalnej audycji radiowej, kursu językowego, szkoleń dla hobbystów albo dzieła literackiego czytanego w odcinkach przez lektora.

⁴³ Vodcast – dostarczanie online filmów wideo na życzenie.

malnej i nieformalnej są w stosunku do siebie komplementarne, w takim znaczeniu, że uzyskanie znaczącego wyniku badawczego wymaga posłużenia się oboma typami komunikacji. Wiele artykułów publikowanych w czasopismach naukowych rozpoczynało swój żywot jako publikacje przekazywane poprzez kanały komunikacji nieformalnej, takie jak raporty i materiały konferencyjne. Najprostszym wytłumaczeniem potrzeby korzystania z szarej literatury jest zbyt długi proces obróbki redakcyjnej i recenzowania artykułów w periodykach naukowych. Dodatkowo publikowanie formalne w czasopiśmie drukowanym lub hybrydowym często nie dostarcza wystarczającej ilości informacji, aby opisywany eksperyment naukowy mógł zostać powtórzony. Brak miejsca powoduje niedostępność wielu informacji, jak specyfikacja użytego sprzętu, rozbudowane wnioski, a także opis popełnionych błędów. Przyjmuje się, że komunikacja nieformalna jest bardziej efektywna w zakresie przekazywania wiedzy ukrytej (*know-how*), natomiast komunikacja formalna służy do przekazywania faktów i opisów (*know-what*).

Najlepsze rezultaty daje łączenie obu (formalnych i nieformalnych) form publikacji. Taka propozycja pojawiła się m.in. w pracach prowadzonych przez NISO, w których badano możliwości wersjonowania publikacji naukowych [NISO/ALPSP JAV 2008, s. 5]. Oprócz kolejnych wersji artykułu naukowego (oryginał autorski, wersja dostarczona, zaakceptowana, recenzowana, opublikowana, poprawiona) pojawia się tam możliwość powstawania równoległych wersji tej samej treści konceptualnej, zaliczanych do szarej literatury, jak wersja robocza, raport, blog, prezentacja. Pomiędzy wszystkimi wersjami zachodzą częste relacje, wynikające z wymiany między nimi treści naukowych. Efektem jest zacieranie różnic pomiędzy odmiennymi do niedawna formami publikowania naukowego.

W dalszej części przedstawię dwie formy publikowania elektronicznego: biblioteki cyfrowe i repozytoria. Zazwyczaj są one częścią publikowania Open Access (zob. p. 4.1). W nich następuje agregacja publikacji elektronicznych, często wcześniej wyjętych z innych całości (na przykład eprint artykułu wyjęty z czasopisma i umieszczony w repozytorium). Jak już pisałem, mogą one być uważane za szarą literaturę. Oprócz wymienionych, istnieją inne formy publikowania, takie jak na przykład strony Web, na których mogą być udostępniane zarówno pojedyncze publikacje (na przykład rozprawa doktorska), jak i całe agregaty publikacji (na przykład strony Web wydawców czasopism).

Biblioteki cyfrowe są formą publikowania, realizowaną przez biblioteki tradycyjne, które umieszczają w Sieci zdigitalizowane wersje publikacji drukowanych, zazwyczaj posiadanych w swoich zasobach⁴⁴. Gwoli ścisłości należałoby więc odróżnić elementy biblioteki cyfrowej służące publikowaniu od ich zasobów (publikacji) oraz towarzyszących im zasobów metadanych. Możliwe są też inne formy organizacji, których przykładem jest wspomniany już projekt Gutenberg. Najczęściej digitalizowane są zbiory, dla których upłynął okres ochrony praw autorskich (*copyright*), stąd kolekcje te mają często charakter historyczny. Wydaje się, że tworzenie bibliotek cyfrowych (digitalizacja tradycyjnych zbiorów bibliotecznych), obecnie wspomagane przez takie giganty światowe, jak Google oraz realizacja projektów digitalizacji na skalę masową, wskazuje na rozpoczęcie etapu, znanego wcześniej z historii rozwoju druku. W czasach rozpowszechniania się wynalazku Gutenberga starano się uwiecznić w druku wszystkie wcześniej istniejące materiały rękopiśmienne. Obecnie na tej samej zasadzie digitalizowane są masowo wcześniej powstałe druki.

⁴⁴ Należy jednak zauważyć, że biblioteki cyfrowe w tym znaczeniu traktowane są jak wszystkie inne biblioteki. Według zasad digitalizacji przedstawionych przez ALA, biblioteki cyfrowe są nadal bibliotekami, których rolą jest zapewnienie łatwego dostępu, przestrzeganie zasad prywatności, traktowanie zasobu, tworzonego zgodnie z przyjętą polityką jako dobra publicznego. Profesjonalni bibliotekarze tworzący biblioteki cyfrowe powinni zadbać o długotrwałą archiwizację i zgodność ze standardami. Z drugiej strony stwierdza się, że zasoby te są częścią globalnej infrastruktury informacyjnej [Johnson 2007, s. 7].

Biblioteki cyfrowe, jako serwisy GBC, zawierają pełne teksty zdigitalizowanych materiałów tradycyjnych. Zasoby te udostępniane są w Webie, w różnych formatach, dostosowanych do potrzeb dystrybucji skanowanych obrazów stron. Różne też mogą być prawa dostępu; zwykle jest on bezpłatny, może jednak być ograniczony do uprawnionych osób. Początkowo wiele bibliotek cyfrowych stosowało jedynie listy posiadanych i udostępnianych zbiorów, obecnie przeważają bazy danych, w których biblioteki przechowują metadane. Czasem tworzone są one na podstawie metadanych wcześniej sporządzonych dla publikacji tradycyjnych (na przykład przez konwersję). Twórcy bibliotek cyfrowych dbają też zwykle o indeksowanie posiadanych treści przez wyszukiwarki internetowe.

Repozytoria, kolejny rodzaj serwisów GBC, powstają często jako reakcja na powszechną komercjalizację publikowania naukowego, która oprócz wzrostu cen czasopism, stała się przyczyną żądań wydawców przekazywania praw własności intelektualnej przez autorów, często bez żadnej rekompensaty (zob. p. 1.2). Wiele instytucji naukowych rozpoczęło publikowanie prac naukowych we własnych repozytoriach, z pominięciem wydawnictw komercyjnych lub jako uzupełnienie oferty komercyjnej. Repozytorium jest to zbiór zasobów cyfrowych, służący gromadzeniu i zachowaniu efektów twórczości intelektualnej określonej grupy lub społeczności (specjalistów z danej dyscypliny lub pracowników określonej instytucji naukowej). Udostępniane obiekty opatrywane są zazwyczaj metadanymi. Do repozytoriów trafiają tak zwane eprinty, które można podzielić na preprinty, udostępniane przed opublikowaniem w publikacjach tradycyjnych oraz postprinty, wersje artykułów opublikowane w czasopismach, zazwyczaj po przejściu procesu kontroli jakości (recenzowania).

Repozytoria spełniają podobne funkcje, co czasopisma naukowe. Mark Ware wymienił następujące funkcje czasopisma: rejestrację, rozpowszechnianie, kontrolę jakości (certyfikację) i archiwizację, sugerując dodatkowo nawigację [Ware 2006]. Herbert van de Sompel dla repozytoriów praktycznie wskazał te same funkcje: rejestracja, kontrola jakości, uświadamianie, archiwizacja i usatysfakcjonowanie [Sompel i in. 2004]. Należy więc oczekiwać, że w środowisku cyfrowym różnice pomiędzy tymi dwoma sposobami publikowania (czasopismem i bazą danych artykułów) będą coraz bardziej zanikać. Według Davida Weinbergera cyfryzacja komunikacji pozwala zauważyć, że agregacja mniejszych obiektów w większe (artykułów w czasopisma) w świecie druku nie była powodowana względami merytorycznymi, lecz ekonomią wytwarzania i dystrybucji artefaktów fizycznych, jakimi są publikacje drukowane; posiadanie wersji elektronicznej pozwala na o wiele bardziej zróżnicowaną i użyteczną aranżację treści w większe zasoby [Weinberger 2007, s. 10].

Ze względu na sposób organizacji archiwum elektronicznego wyróżnić można:

- Repozytoria dziedziczne, powstające najwcześniej, gromadzące materiały z określonej dziedziny lub kilku dziedzin pokrewnych. Takim repozytorium jest ArXiv⁴⁵, pierwsze repozytorium z zakresu nauk ścisłych i informatyki. Zamiarem Paula Ginsparga, twórcy tego repozytorium, nie było pominięcie komercyjnych wydawców, a jedynie przyspieszenie obiegu informacji naukowej przez archiwizowanie zrecenzowanych artykułów. Repozytorium to zyskało tak wielką popularność, że obecnie wielu autorów umieszcza w nim swoje publikacje bez przekazywania ich do jakiegokolwiek czasopisma [Gentil-Beccot, Mele, Brookes 2009].
- Repozytoria instytucjonalne, tworzone przez uczelnie i inne instytucje naukowe dla archiwizacji prac swoich pracowników oraz studentów. Archiwizowane prace mają bardziej zróżnicowany charakter, gdyż oprócz artykułów udostępniane są wszelkiego rodzaju publikacje i materiały dydaktyczne z zakresu reprezentowanego przez instytucję. Zdarza się umieszczanie także materiałów administracyjnych (na przykład rozporządzeń rektora). Bywa, że przekazywanie materiałów do repozytorium jest obligatoryjnym nakazem władz instytucji naukowej.

⁴⁵ <http://arxiv.org>.

- Repozytoria gromadzące dokumenty określonego rodzaju, na przykład materiały tradycyjnie niepublikowane, takie jak dysertacje i prace magisterskie.

Rozróżnienie biblioteki cyfrowej od repozytorium komplikuje niejednolita terminologia. Repozytorium jeszcze kilka lat temu definiowane było jako część biblioteki cyfrowej (w znaczeniu lokalnej, nie globalnej biblioteki), w której przechowuje się obiekty cyfrowe [Arms 2000, s. 274]; obecnie repozytoriami nazywa się każdy zasób danych cyfrowych, w znaczeniu takim, jak podano to powyżej. Różnica pomiędzy lokalną biblioteką cyfrową a repozytorium stała się funkcjonalna, tzn. dotyczy sposobu realizowania funkcji. Biblioteka cyfrowa jest to bowiem zasób tworzony (zazwyczaj planowo) przez bibliotekarza czy inne osoby odpowiedzialne, poprzez digitalizację (skanowanie) zasobów danej biblioteki. Bibliotekarz dysponuje wówczas metadanymi publikacji tradycyjnej (w OPAC biblioteki), które często stosuje także dla kopii elektronicznej. Repozytorium natomiast to miejsce, w którym składowane są nowe, na bieżąco tworzone publikacje i inne materiały, które od razu powstają w formie elektronicznej, nie wymagają więc zazwyczaj digitalizacji, lecz ewentualnego przeformatowania. W repozytorium gromadzenie danych jest mniej planowe, gdyż wynika z bieżącej produkcji naukowej. Metadane również muszą być tworzone od nowa. Często repozytorium (szczególnie instytucjonalne) trudno odróżnić od biblioteki cyfrowej; może być ono jej częścią. Efektem zróżnicowania funkcji jest powstawanie kolejnych odmienności, na przykład biblioteki cyfrowe częściej niż repozytoria postrzegane są jako „miejsca” realizujące funkcje archiwizacji obiektów cyfrowych. Obecnie jednak również repozytoria uznaje się za odpowiednie do realizacji długotrwałej archiwizacji [Day, Penneck, Allison 2007].

W repozytorium, znacznie częściej niż w bibliotece cyfrowej, dokumenty umieszczane są przez ich autorów, a nie wyspecjalizowany personel. Powoduje to problemy, dotyczące jakości metadanych, poziomu obróbki redakcyjnej materiałów (szczególnie w przypadku preprintów) oraz zagadnień prawnych (w repozytorium umieszcza się tekst, który jednocześnie lub nieco później trafia do czasopisma). Spotyka się opinie, że mimo tego należy dążyć do napełniania repozytorium obiektami cyfrowymi w maksymalnym stopniu, gdyż dostęp do pełnego tekstu jest ważniejszy od poprawnej formy i posiadania profesjonalnych metadanych [Adams 2008].

Oprócz wcześniej wymienionych serwisów, również **strony Web** stanowią bardzo liczne i zróżnicowane źródło informacji⁴⁶. Szacuje się, że liczba indeksowalnych stron Web na początku 2009 r. przekroczyła co najmniej 25 miliardów (w 2005 r. szacowano tę ilość na 11,5 mld), z czego największa wyszukiwarka, Google indeksuje 18 miliardów⁴⁷ (w 2005 r. 8 mld) [Gulli, Signorini 2005]. Pomimo indeksowania większości dostępnych stron przez wyszukiwarki internetowe, duża część informacji znajduje się poza ich zasięgiem. Pozostają jeszcze strony Web, które w ogóle nie są indeksowane⁴⁸; ta jego część nazywana jest ukrytym, niewidzialnym lub głębokim Webem, którego rozmiary mogą przekraczać Web indeksowalny ponad 500-krotnie⁴⁹ [Pamuła-Cieślak 2006, s. 153]. Z punktu widzenia GBC bardzo ważne jest, aby strony Web i obiekty, które się na nich znajdują, mogły być indeksowane przez wyszukiwarki, a więc aby nie stanowiły ukrytego Webu [Landow 2006, s. 39]. Dotyczy to na przykład stron bibliotek tradycyjnych i zawartości ich OPAC a także bibliograficznych baz danych, wciąż często niewidocznych dla wyszukiwarek.

⁴⁶ Wszystkie serwisy również posiadają swoje strony Web, jednak służą one przede wszystkim do udostępniania innych treści, poprzez stronę Web. W takim sensie strona Web stanowi raczej identyfikowalne miejsce w Sieci, służące publikowaniu obiektów cyfrowych, niż jest nośnikiem konkretnej treści. Nośnikiem w takim przypadku jest artykuł lub książka.

⁴⁷ Dane szacunkowe, podają za stronę: <http://www.worldwidewebsitesize.com/>.

⁴⁸ Strony te mogą nie być indeksowane z różnych powodów: braku odnośników do/ze strony, brak tekstu na stronie (zawartość multimedialna), pobierania opłat za dostęp, nieodpowiedniego formatu (np. bazy danych, w tym katalogi biblioteczne), zmienności i nietrwałości (np. strony dynamiczne Web).

⁴⁹ Z szacunkami tymi nie zgadzają się inni autorzy [Cisek, Sapa 2007], nie zmienia to jednak faktu, że rozmiar nieindeksowanego Webu są ogromne; same zasoby naukowe oblicza się na 20 do 100 mld dokumentów.

Ze względu na to, że każdy może opublikować w Webie dowolne treści, jakość zawartej w nim informacji jest bardzo różna. Oprócz wymienionych stron bibliotek i innych wiarygodnych instytucji istnieją strony prywatne, o nieokreślonej jakości. Tworzone są specjalne poradniki instruujące w zakresie określania jakości stron Web [Bednarek-Michalska 2002]. Kryteria jakościowe dla stron Web nie różnią się wiele od kryteriów dla jakichkolwiek publikacji, również drukowanych. Więcej na temat jakości serwisów GBC znajduje się w p. 4.8.

Strony Web posiadają strukturę hierarchiczną. Zazwyczaj ze strony głównej można przemieszczać się do innych stron, umieszczonych na niższych poziomach hierarchii. Przejście w obu kierunkach możliwe jest dzięki mechanizmowi hipertekstowych odnośników⁵⁰, których rolę może odgrywać dowolny element strony Web (część tekstu, element graficzny). Przeglądarki internetowe, dzięki zapamiętywaniu historii poszczególnych kroków nawigacji, pozwalają na cofanie się wstecz do poprzednich etapów przeglądania dokumentów Web.

Z punktu widzenia typów treści i formatów, strony Web stanowią zwykle niejednorodne, złożone obiekty Internetu. Oprócz tekstu, który jest zasadniczą częścią treści tych obiektów (co pozwala na ich indeksowanie przez wyszukiwarki), składają się na nie także wszelkiego rodzaju obiekty multimedialne. Mogą one być ze sobą łączone w skomplikowany sposób, tworząc sieć relacji.

Zasoby multimedialne mogą stanowić ważne źródło informacji w wielu dziedzinach. Dynamiczna natura dokumentów elektronicznych pozwala na wykorzystywanie tego rodzaju informacji niestacycznych. Multimedia uważane są za zasoby stwarzające wartość dodaną dokumentów elektronicznych, niemożliwą do osiągnięcia w formie drukowanej. Zasoby te powstają często w wyniku digitalizacji zbiorów muzealnych, galerii sztuki, prywatnych kolekcji, w celu udostępnienia szerszej publiczności. Obrazy, w tym ruchome, tworzone są w formie cyfrowej lub digitalizowane przez instytucje edukacyjne dla wykorzystania w celach dydaktycznych, w tym w e-learningu. Dla budowy i przechowywania tych obiektów używane są specjalne technologie, odmienne od stosowanych dla dokumentów tekstowych. Źródła informacji cyfrowej, zawierające nagrania dźwiękowe mowy i muzyki, wzbudzają zainteresowanie w wielu instytucjach edukacyjnych, gromadzone są także przez biblioteki publiczne. Cyfrowe książki mówione stanowią odpowiednik tradycyjnych książek mówionych, nagrywanych na taśmie magnetycznej, które miały duże znaczenie dla użytkowników niewidomych.

Problem zaopatrywania obiektów multimedialnych w odpowiednie metadane jest nawet poważniejszy niż dla dokumentów tekstowych, gdyż standardowe indeksowanie pełnotekstowe nie wchodzi tu na razie w rachubę, chociaż realizowane są eksperymenty w tym zakresie; pewne nadzieje wiązać można z rozwojem zastosowań tzw. wyszukiwania uniwersalnego (zob. p. 4.3). Jak pisze Christie Borgman, ze względu na możliwość różnorodnych interpretacji, opis obiektów dźwiękowych i graficznych jest znacznie trudniejszy, niż obiektów tekstowych [Borgman 2003a, s. 148]. Fotografia jest często interesująca ze względu na to, kto został sfotografowany, a czasem ze względu na to, kto ją wykonał. W obu wypadkach zazwyczaj nie da się wygenerować tej informacji automatycznie. Obiekty graficzne mogą być opisywane za pomocą takich kryteriów, jak zawartość (ludzie, drzewa, budynki...), kolor, rozmiar, data czy twórca. Możliwe jest automatyczne odróżnianie takich cech, jak występowanie twarzy ludzkich; polega to na porównywaniu kształtów z wzorcem przedstawionym przez użytkownika. Najlepszym sposobem opisu zapisów dźwięku: muzyki, mowy, efektów dźwiękowych, jest zaopatrywanie ich w opis tekstowy. Istnieją również systemy automatycznego rozpoznawania mowy, słów i zdań, które mogą być konwertowane do terminów wyszukiwawczych.

⁵⁰ Ścisłej mówiąc, odnośniki we współczesnym Webie są jednokierunkowe, więc dla umożliwienia przejść w obu kierunkach hierarchii konieczne jest utworzenie odpowiednich mechanizmów na wszystkich poziomach dokumentu (strony Web).

Zarówno elektroniczne, jak i drukowane publikacje naukowe, w których za podstawową jednostkę uważany jest artykuł w czasopiśmie naukowym, powstają na podstawie zbiorów **danych empirycznych** (*data sets*), tworzonych w trakcie realizacji badań naukowych. Obecnie same dane stają się równie ważne, jak artykuły naukowe, traktowane w większości przypadków jako interpretacja danych. Projektowanie procesu gromadzenia danych, ich pozyskiwanie oraz interpretacja stanowią bardzo ważną część procesu przygotowania publikacji.

Współczesna nauka bezpośrednio uzależniona jest od procesów tworzenia, rozpoznań i analizy danych. Z kolei te procesy charakteryzują się masową skalą produkcji danych i globalnym rozprzestrzenieniem ich zasobów. Uczni potrzebują pomocy w identyfikacji i selekcji danych, przydatnych w pracach badawczych oraz ich archiwizacji i ochrony [Borgman, Wallis, Enyedy 2007, s. 20].

Ważną tendencją w publikowaniu naukowym jest zacieranie się różnic pomiędzy prymarnymi zasobami danych, głównie zawierającymi dane nieprzetworzone lub nieprzeanalizowane, a źródłami informacji, umiejscawiającymi dane we właściwym kontekście, takimi jak artykuły, referaty i książki. Zestawy danych są wyszczególniane w spisach publikacji naukowych oraz cytowane wraz z innymi źródłami bibliograficznymi w artykułach naukowych. Publikacje naukowe coraz częściej zawierają dołączone zestawy danych, modele, ruchome ilustracje, nagrania dźwiękowe i tym podobne obiekty. Dane są materiałem wyjściowym, powstającym w trakcie badań oraz źródłem do publikacji naukowych, a także do prowadzenia kolejnych badań i kształcenia. Są więc podstawą działalności naukowej.

Przykładami dobrze zarządzanych zasobów danych, mających znaczny wpływ na rozwój badań naukowych, mogą być faktograficzne bazy danych, takie jak GenBank⁵¹ czy Protein Information Resource⁵², jak również modyfikowane zasoby powstające w współpracy wielu zainteresowanych stron, takie jak strony Web dotyczące określonych zagadnień czy dziedzin, które łączą dostęp do źródeł prymarnych, analiz i komentarzy. Wymienić można także systemy informacji geograficznej (GIS) i geoprzestrzennej, pozwalające na gromadzenie danych, służących przewidywaniu pogody, zmian demograficznych a nawet badaniu starożytnych miast. Bardziej złożonym przykładem są koloaboratoria, łączące wykorzystanie wideokonferencji, synchronicznego i asynchronicznego przesyłania komunikatów tekstowych, rozproszonej obsługi instrumentów naukowych, dostępu do „surowych” danych (z czujników urządzeń badawczych, symulacji i archiwów danych), analiz i narzędzi służących wizualizacji, literaturowych baz danych i narzędzi umożliwiających publikowanie. Za pomocą mieszania, miksowania danych, z angielska zwanego *mashup 'em*⁵³, analogicznego do miksowania muzyki i tworzenia kolaży, łączy się ze sobą dane z różnych zasobów lub scala różnorodne dane w spójną całość. W ten sposób mogą być także prowadzone analizy, niemożliwe lub bardzo trudne do uzyskania w innych warunkach. Istnieje pewna sprzeczność pomiędzy potrzebami związanymi z wyszukiwaniem (indeksowaniem) danych a ich prezentacją ułatwiającą prowadzenie analiz. W celu wyszukiwania za pomocą wyszukiwarki dane muszą być zindeksowane, do czego najlepiej nadają się dane tekstowe. Z drugiej strony olbrzymia ilość gromadzonych danych wymaga zastosowania specjalnych technik ich prezentacji, do których zalicza się wspomnianą wizualizację. Wynika z tego, że postać danych najlepiej nadająca się do wy-

⁵¹ GenBank jest bazą danych sekwencji DNA tworzoną w NIH (National Institute of Health, USA). Zob. <http://www.ncbi.nlm.nih.gov/Genbank/>.

⁵² Protein Information Resource (PIR), zasoby danych o sekwencjach proteinowych i ich funkcjach, prowadzi UniProt. Zob. <http://pir.georgetown.edu/>.

⁵³ W informatyce *mashup* jest to aplikacja Web scalająca dane z różnych źródeł w jedno, zintegrowane źródło. Najczęściej podaje się przykład wykorzystania danych kartograficznych z Google Maps na stronach opisujących nieruchomości, przestępczość i in. zjawiska występujące na danym terenie.

szukiwania (tekstowa) mniej nadaje się do ich analizy i interpretacji (tu przydatna jest postać graficzna) i na odwrót.

Obecnie zagadnienia dotyczące zbiorów danych wiąże się z badaniami prowadzonymi w naukach ścisłych, jednak jest bardzo prawdopodobne, że wkrótce powstawać będą także duże zasoby danych, przygotowywane przez specjalistów z zakresu nauk społecznych [Davis, Vickery 2007, s. 26]. Dzięki nowym narzędziom, pozwalającym na badanie, analizę i porównywanie danych w formie cyfrowej, mogą oni na przykład uzyskać lepszy obraz tekstów literackich i historycznych.

Dane mogą pochodzić z różnych źródeł. Wymienić można:

- Dane pochodzące z obserwacji, takie jak elementy pogody, związane z określonym miejscem i czasem.
- Dane obliczeniowe, powstające w wyniku wykonania modelowania lub symulacji komputerowej, zarówno dla rzeczywistości fizycznej jak i wirtualnej.
- Dane eksperymentalne, powstające podczas badań laboratoryjnych, takich jak wykonywanie reakcji chemicznych czy kontrolowane eksperymenty behawioralne.
- Dokumenty rządowe, biznesowe oraz życia społecznego także stanowią źródła przydatnych danych wykorzystywanych w badaniach nauk ścisłych i społecznych [Borgman 2007, s. 120].

„Czyste” dane, uzyskiwane w trakcie badań naukowych oraz publikacje powstające na ich podstawie, przez długi czas były traktowane oddzielnie, a nawet jako przeciwieństwa, dla których prowadzono odrębne badania. W wyniku tworzenia globalnych zasobów danych sytuacja ta uległa zmianie. Według Henry Rzepy i Petera Murray-Rust, w wielu dyscyplinach dane z badań są niezbędne dla pełnego wykorzystania artykułu naukowego [Rzepa, Murray-Rust 2001, s. 178]. W związku z tym dane, związane z artykułem, muszą być reprezentowane w sposób bardziej precyzyjny niż dotychczas, tak aby mogły one być nie tylko rozpoznawane przez ludzi, ale także przetwarzane maszynowo. W ten sposób dane zbliżają się do koncepcji Semantycznego Webu (zob. p. 4.9), który ze swej natury zawiera cały proces publikacji. Centralnym pojęciem Semantycznego Webu jest samodefiniowanie danych, dzięki czemu decyzje dotyczące ich treści i warunków funkcjonowania, w tym możliwych transformacji, mogą być podejmowane nie tylko bezpośrednio przez ludzi, ale także przez agentów (oprogramowanie). Granice tego, co tradycyjnie nazywane jest „artykułem” mogą być rozszerzane w górę i w dół; w dół aż do pojedynczych elementów kodowanych w XML, w górę agregowane do postaci czasopisma, kolekcji czasopism a nawet całego Semantycznego Webu. Nie oznacza to utraty odrębności i identyfikowalności, gdyż każda jednostka informacji może być łączona z metadanymi zawierającymi niezbędne dane. Powstaje więc globalna baza wiedzy (biblioteka cyfrowa, docuverse – zob. p. 2.3), składająca się ze ściśle zagregowanych jednostek treści, „artykułów”, pomiędzy którymi zacierają się granice.

Prawdziwa korzyść ze stosowania publikowania elektronicznego może wynikać z zastosowania **nowych form publikacji**, wykorzystujących specyficzne cechy środowiska cyfrowego [Borgman 2003a, s. 90]. Nowe rodzaje publikacji powinny umożliwiać lepsze wyszukiwanie, sortowanie, wyświetlanie informacji, tworzenie relacji (hiperlinki) między obiektami oraz połączeń obiektów z metadanymi.

Publikowanie elektroniczne jest dziedziną, w której, podobnie jak w wielu innych elementach GBC, zachodzą od kilku lat szybkie i poważne zmiany. W ich efekcie w Sieci dochodzi do zanikania tradycyjnych form, takich jak książka, czasopismo, artykuł, raport – na rzecz bardziej ujednoliconej formy, którą można nazwać wypowiedzią na określony temat, o długości i strukturze związanej raczej z tematem i funkcją wypowiedzi. Nie ma potrzeby kontynuowania tradycyjnych form wypowiedzi. Podstawową formą publikacji internetowej staje się raport naukowo-techniczny, zajmując miejsce między krótszym artykułem a bardziej sformalizowaną książką elektroniczną. Forma raportu nie jest tak

ograniczona jak artykuł, w szczególności co do objętości i możliwości zamieszczania materiałów uzupełniających, takich jak zestawy danych uzyskanych w trakcie badań. Jednocześnie nie wymaga tak starannego przygotowania i opracowania redakcyjnego jak książka, dzięki czemu może powstawać znacznie szybciej, zawierając treści równie aktualne, jak artykuł. Podobnie zacierają się różnice pomiędzy czasopiśmie elektronicznym, udostępnianym na stronie Web, a pełnotekstową bazą danych. Zmiany te wynikają zarówno z personalizacji, jak i z socjalizacji procesów komunikacji naukowej [Hull, Pettifer, Kell 2008].

Zachowaniem czystej formy tradycyjnego przekazu nie są zainteresowani zarówno twórcy publikacji elektronicznych, jak również ich odbiorcy; nie jest dla nich ważne czy informacje zawarte są w książce, czasopiśmie, bazie danych, blogu, czy na stronie Web [OCLC 2004]. Dostawcy treści, szczególnie komercyjni, w coraz większym stopniu zwracają się do tych, lubiących eksperymentowanie użytkowników informacji, dostarczając im dane w zróżnicowanej formie, z czym często powiązane są różne modele i struktury cen, bardziej odpowiadające indywidualnej konsumpcji informacji. Przykładowo część tekstu może być zapisana w formacie HTML lub PDF i umieszczona na stronie Web w wolnym dostępie, natomiast zainteresowani mogą otrzymać cały tekst po rejestracji i/lub dokonaniu zapłaty. Metodę tę stosuje na dużą skalę księgarnia internetowa Amazon, gdzie za pomocą opcji „search inside the book”, użytkownik może przeczytać pewną liczbę stron (w tym zwykle wstęp, spis treści i zakończenie + indeks) dla upewnienia się w trafności wyboru, jednak aby otrzymać pełną treść, należy ją kupić. Podobnie serwis Google Books oferuje dostęp do obszernych fragmentów zdigitalizowanych książek.

Wciąż rośnie ilość informacji przesyłanej od twórcy do odbiorcy za pomocą zupełnie innych kanałów niż te, do których byli oni od dawna przyzwyczajeni (głównie bibliotek). Przekazywanie treści tymi kanałami nie jest już uzależnione od formy publikacji, przez co użytkownik także uniezależnia się od tradycyjnych sposobów dystrybucji informacji. Tendencja ta dotyczy komunikacji naukowej w nie mniejszym stopniu niż rozrywki. Dla bibliotek oraz dostawców publikacji oznacza to konieczność zmian w procesach gromadzenia, zarządzania i dostarczania informacji.

Badania kierunków zmian, prowadzone w USA, wskazują na wyhamowywanie impetu publikacji tradycyjnych. Wzrostowi oferty towarzyszy spadek popytu [OCLC 2004, s. 9]. Treści wysokiej jakości wyszły poza ograniczenia związane z formą publikacji i przeniosły się do otwartego Webu; możliwe jest to dzięki rozpowszechnieniu narzędzi wspomagających autopublikowanie i autoarchiwizację oraz dostępności usług telekomunikacyjnych. Pracownicy naukowcy umieszczają własne artykuły naukowe w repozytoriach, na stronach Web i blogach, dzięki czemu wyszukiwarki internetowe mogą wyeksponować ich własność intelektualną na globalnym rynku idei, a oprócz tego mogą oni opublikować je w bardziej tradycyjnej formie drukowanej lub w czasopiśmie elektronicznym.

Również czasopisma elektroniczne eksperymentują ze sposobami dostarczania treści do użytkowników, tworząc różne formy transmisji:

- Czasopisma wirtualne. Są to zdecentralizowane czasopisma, które nie gromadzą centralnie tekstów, lecz agregują je z różnych miejsc w sieci, co upodabnia je do portalu. Teksty mogą znajdować się na przykład na stronach autorów lub serwerach uczelnianych. Na stronie czasopisma tworzona jest wartość dodana przez grupowanie tekstów w zeszyty, ułatwianie dostępu poprzez spisy treści, dodanie możliwości wyszukiwania i powiadomiania. W podobny sposób publikowana jest część materiałów konferencyjnych.
- Czasopisma typu *living review*. Jest to sposób na uporanie się z szybkością, w wielu naukach i rodzajach publikacji (na przykład opisujących stan bieżący), dezaktualizacją materiałów. Artykuły są aktualizowane w razie potrzeby przy wykorzystaniu odpowiedniego oprogramowania, pozwalającego na dostęp do kolejnych wersji, wskazującego na miejsca zmian. Dzięki temu artykuły stale pozostają aktualne [Nahotko 2007b, s. 58].
- Multimedialne czasopisma interaktywne. Podobnie jak wszelkie publikacje elektroniczne

ne, również czasopisma uzupełniane są sekwencjami wideo, ścieżkami dźwiękowymi czy grafiką animowaną. Interaktywność reprezentowana jest przez otwarte recenzowanie (czasem po opublikowaniu tekstu), komentarze, punktowanie przez użytkowników.

- Idea *skywriting*. Część naukowych list dyskusyjnych może być uważana za nowy rodzaj publikacji. W tym celu muszą one spełniać dwa warunki: nadsyłane materiały muszą być czymś więcej niż notatką czy zapytaniem (na przykład rozbudowaną odpowiedzią lub komentarzem) oraz muszą być archiwizowane. Publikacje naukowe wyłaniają się w takich warunkach ze swobodnej dyskusji na listach dyskusyjnych, prowadzonych za pomocą poczty elektronicznej.
- Publikacje przekraczające tradycyjne granice form (książki, artykułu), o czym była już mowa.
- Czasopisma ewoluujące w kierunku baz danych. Dołączanie do artykułów naukowych zestawów danych gromadzonych podczas opisywanych badań naukowych powoduje, że powstają publikacje na pograniczu czasopism i baz danych. Uczni publikują zasoby danych tak, jak wcześniej publikowali artykuły, co powoduje, że wiele czasopism przyjmuje wkrótce formę faktograficznych baz danych. Dołączanie odnośników do bibliografii załącznikowych powoduje natomiast powstawanie indeksów cytowań.

Istotną cechą nowych publikacji elektronicznych jest coraz śmielsze wykorzystywanie jednej z cech najwyraźniej odróżniających je od publikacji tradycyjnych – potencjalnej możliwości i łatwości dokonywania modyfikacji, czemu sprzyja oddzielenie treści od nośnika fizycznego. Publikacje elektroniczne nie tylko funkcjonują w wielu wersjach, ale mogą być do tego stopnia łatwo modyfikowane, a hipertekst tak ogranicza autonomię tekstu, że zanika dawne pojęcie nie tylko ostatecznej, autorytarnej wersji, ale nawet autorstwa; są efektem ciągłej pracy społeczności sieciowej. [Landy 2006, s. 131]. Powoduje to między innymi problemy z tradycyjnym cytowaniem takich prac.

Główne kierunki zmian w zawartości obiektów cyfrowych nie dotyczą wyłącznie modernizacji technologii; to także inne sposoby tworzenia, gromadzenia, wykorzystania i archiwizacji treści. Podobnie jak w innych kierunkach rozwoju infosfery, wiele najbardziej ewidentnych zmian przebiega poza tradycyjnymi miejscami związanymi z zarządzaniem informacją. Są realizowane m.in. przez producentów mobilnych, zminiaturyzowanych urządzeń komunikacyjnych, które umożliwiają uzupełnianie lub zastępowanie tradycyjnych kanałów przekazywania treści, takich jak na przykład usługi pocztowe czy wypożyczenia międzybiblioteczne. Duże znaczenie ma stałe obniżanie kosztów przesyłania treści nowymi kanałami. Konwergencja technologii czyli łączenie technologii w wysocze zintegrowany system [Castells 2008, s. 80], powoduje, że uczeni, podczas dostarczania i odbioru treści, nie są ograniczeni wyłącznie do jednej technologii, a zasadnicza zmiana społeczna polega na tym, że technologie te są nierozdzielnie włączone w sposób funkcjonowania nauki.

Dzięki tym zmianom pracownicy nauki mogą w coraz większym stopniu kontaktować się między sobą, wymieniać informację i docierać do niej bez względu na miejsce i czas. Zmieniają się też ich przyzwyczajenia i zachowania – coraz więcej użytkowników korzysta z tekstów na ekranie różnych, zwykle mobilnych urządzeń elektronicznych. Dokumenty elektroniczne są coraz rzadziej drukowane dla dokładnego zapoznania się z ich treścią, co było częstą praktyką w końcu XX wieku.

Interesującym rezultatem zmian jest kilka, często kolidujących ze sobą, zjawisk, takich jak zapotrzebowanie na indywidualne tworzenie i dystrybucję treści, wzrost akceptacji informacji ze źródeł „nieautorytatywnych” i w rezultacie tego utowarowienie informacji oraz rozwój technologii służących rozwojowi publikowania osobistego i społecznego.

Serwisy Webu 2.0, takie jak Wiki i blogi, jako formy publikowania społecznego, wskazują kierunki zmian w systemach informacyjnych, mogą nawet doprowadzić do zmiany paradygmatu publikowania. Formy te, uzupełnione mechanizmami syndykalizacji treści

typu RSS, umożliwiają każdemu użytkownikowi Internetu publikowanie własnych treści w ścisłym kontakcie z ich odbiorcami. Te same formy coraz częściej są wykorzystywane jako naturalny sposób komunikowania się obsługi serwisów GBC (bibliotekarzy) ze społecznością użytkowników i co równie ważne, użytkowników z bibliotekarzami. Paradoksalnie medium globalne, jakim jest Internet, jest wykorzystywane przez uczestników publikowania społecznego nie do masowego, globalnego rozpowszechniania treści, ale raczej do odszukania w wielkiej masie użytkowników sieci zległej tej niewielkiej grupy odbiorców, z którymi blog ułatwi utrzymanie stałego kontaktu.

Połączenie nowych technologii, stosowanych zarówno w Webie 2.0, jak i Semantycznym Webie staje się podstawą publikowania semantycznego [Shotton 2009; Shotton i in. 2009]. Jego podstawą jest kodowanie treści w XML, związana z nim standardowa strukturyzacja artykułu i jego abstraktu, uzupełnianie publikacji bogatym materiałem dodatkowym (zbiory danych, multimedia), powiązania z innymi artykułami m.in. poprzez odnośniki w bibliografiach załącznikowych (indeksy cytowań) oraz wzbogacanie semantyki tekstu poprzez łączenie go z zawartością ontologii, tworzonych dla danej dziedziny, co pozwala m.in. na kontrolę terminologii. W ten sposób powstaje semantyczne środowisko udostępniania nie tyle tekstów artykułów, co ustrukturyzowanej wiedzy. Powoduje ono zmiany w sposobie korzystania z tekstów naukowych; publikacje stanowią raczej sieć powiązanych informacji, niż zamknięte, dobrze zdefiniowane jednostki, takie jak artykuły z drukowanych czasopism naukowych. Według Allena Reneara i Carole Palmer, uczeni tylko przeglądają czasopisma dla zdobywania wartościowych informacji, nie czytając ich w całości. „Skanują” oni teksty w poszukiwaniu przydatnej sobie terminologii, akapitów, diagramów i streszczeń ich interesujących, co nazywane jest czytelnictwem strategicznym [Renear, Palmer 2009, s. 830]. Czytanie materiałów drukowanych było wyraźnie oddzielone od innych czynności, takich jak pisanie lub rozmowa telefoniczna. Obecnie coraz częściej oznacza ono nie tylko zapoznanie się z tekstem na ekranie, ale także na przykład odpowiadanie na pocztę elektroniczną, gdzie czytanie i pisanie, w odczuciu użytkownika, wykonywane jest jednocześnie [Marshall 2005, s. 130; Górska 2009, s. 125].

Wraz ze wzrostem stopnia integracji informacji wokół wykonywanych zadań, różnica pomiędzy środkiem przekazu a przekazem staje się jeszcze mniej wyraźna. Do tej sytuacji muszą dostosować się także bibliotekarze. Dotychczasowi eksperci w zakresie udostępniania, gromadzonych w magazynach, dokumentów tradycyjnych muszą znaleźć sposób na zaistnienie w świecie, w którym powszechna jest integracja treści i kanałów ich dystrybucji.

2.7. Organizacja informacji i wiedzy

Może się wydawać, że organizowanie informacji naukowej nie różni się od organizowania czegokolwiek innego. Takie założenie prowadzi do interpretacji organizacji informacji (wiedzy)⁵⁴ jako rutynowego zastosowania technik modelowania baz danych, powszechnie stosowanych do organizacji takich jednostek (encji), jak pracownicy, oddziały i projekty, w jakimkolwiek przedsiębiorstwie. Organizacja informacji stwarza jednak dodatkowe problemy. Najbardziej istotny z nich, leżący u podstaw wielu złożonych układów specyficznych dla organizacji informacji, jest fakt, że organizując informację musimy brać pod uwagę organizację dwóch jednostek łącznie, ale z uwzględnieniem ich odrębności: abstrakcyjnych dzieł i materialnych egzemplarzy [Svenonius 2000, s. 10]. W środowisku

⁵⁴ Według Jadwigi Woźniak-Kaspepek terminy organizacja informacji (system organizacji informacji) oraz organizacja wiedzy (system organizacji wiedzy) mogą być traktowane synonimicznie [Woźniak-Kaspepek 2008, s. 117].

elektronicznym rozróżnienie to może przyjmować formę wspomnianego już przeze mnie odróżnienia obiektów cyfrowych od publikacji elektronicznych, o którym mowa była w poprzednim rozdziale.

Zasadniczym celem systemu służącego organizacji informacji, a więc także zasobów globalnej biblioteki cyfrowej, jest, najogólniej mówiąc, zgrupowanie tych obiektów, które według przyjętych kryteriów zawierają podobną informację i odrzucenie pozostałych. Projektowanie systemów realizujących to zadanie musi uwzględniać różnego rodzaju ograniczenia: system musi być ekonomicznie efektywny, stanowić kontynuację wcześniejszych rozwiązań (w szczególności uwzględniać istnienie milionów obiektów informacyjnych zorganizowanych wcześniej), powinien wykorzystywać najnowsze osiągnięcia technologiczne. Projektując system należy stosować zasady, stanowiące ogólne specyfikacje lub dyrektywy dla decyzji podejmowanych podczas projektowania [Nahotko 2006b, s. 66].

Organizacja informacji jest ogólną dyscypliną, która obejmuje wiele teorii i działań praktycznych, dotyczących opisu i reprezentacji informacji. Szczególne trudności może sprawiać organizowanie informacji rozproszonej i bardzo dynamicznej, a z taką mamy do czynienia w globalnej bibliotece cyfrowej. Nowego spojrzenia wymaga także od bibliotekarzy, od dawna przyzwyczajonych do organizowania dokumentów w zasobach swoich bibliotek, w sposób pozwalający na możliwie łatwy, a jednocześnie efektywny (również finansowo) dostęp do zawartych w nich informacji. Nie wszystkie, nawet stosunkowo nowe (jak FRBR), osiągnięcia w tym zakresie możliwe są do łatwego i bezpośredniego przeniesienia do globalnej biblioteki cyfrowej, gdyż obciążone bywają potrzebą uwzględnienia fizycznego nośnika informacji i równie fizycznego jego rozmieszczenia w linearnym porządku półek magazynu bibliotecznego. Mimo tego można i należy korzystać z tych doświadczeń, a ich ignorowanie byłoby błędem [Witten, Bainbridge 2003, s. 47]. Rozwiązania w zakresie organizacji informacji, stosowane w bibliotekach, dostarczają bowiem znacznie większych możliwości wyszukiwawczych, niż standardowo używają ich użytkownicy (por. p. 4.8).

Gdy projektowanie zasad organizacji informacji należało przede wszystkim do specjalistów z zakresu informacji naukowej, a jej praktyka polegała głównie na profesjonalnym tworzeniu zasobów metadanych, stosunkowo łatwo było zachować jednolitość, a nawet standaryzację rozwiązań. Obecnie społeczności użytkowników tworzą własne zasoby pełnotekstowe, często nie uwzględniając opinii specjalistów w zakresie organizacji informacji. Organizacja informacji przestaje opierać się na obiektach fizycznych przenosząc się w przestrzeń online, pozwalające na współuczestnictwo członków tych społeczności [Haider, Sundin 2010]. Powoduje to potrzebę stworzenia takich sposobów organizacji informacji, które mogą być stosowane przez specjalistów każdej dyscypliny, odpowiadających za publikowane treści, ale bez przygotowania w zakresie zarządzania informacją. W tym kierunku idą rozwiązania proponowane pod wspólną nazwą Web 2.0 (zob. p. 4.9).

Z organizacją informacji związane jest pojęcie **architektury informacji**, które jest jedną z metafor, związanych z GBC. W ten sposób określa się zazwyczaj działalność, mającą na celu zapewnienie funkcjonalności serwisów GBC, co wymusza uwzględnianie potrzeb użytkowników. Zajmują się nią osoby różnych profesji, tacy jak graficy komputerowi, bibliotekarze, dziennikarze, inżynierowie, marketingowcy oraz specjaliści od projektowania interfejsów i ogólnie interakcji człowiek-komputer. Głównym jednak zadaniem architekta informacji jest pełnienie funkcji osoby zarządzającej projektami (*project manager*), mającej za zadanie organizację współpracy wszystkich twórców serwisu GBC, jako środowiska informacyjnego w taki sposób, aby jednocześnie zapewnione były przyjazność dla użytkownika oraz wymagania projektu. Architekt informacji powinien również stworzyć całościową wizję architektury serwisu. Ten aspekt jego działań związany jest z organizowaniem, etykietowaniem (nazywaniem) i tworzeniem schematów nawigacji w obrębie systemów informacyjnych.

Brak jednoznacznej definicji powoduje, że architekturę informacji określać można wielorako, na przykład jako:

1. Połączenie organizacji informacji, etykietowania (nadawania nazw) i schematów wyszukiwania w systemie informacyjno-wyszukiwawczym.

2. Strukturalne projektowanie przestrzeni informacyjnej, ułatwiającej wyszukiwanie i udostępnianie informacji.

3. Sztukę i naukę strukturyzacji i klasyfikacji serwisów internetowych i intranetowych, ułatwiającą wyszukiwanie i wykorzystanie informacji.

4. Nową dziedzinę badawczą i praktyczną, zajmującą się tworzeniem zasad projektowania i budowy konstrukcji w krajobrazie wirtualnym [Rosenfeld, Morville 2003, s. 21].

Architektura informacji posługuje się narzędziami służącymi do organizacji informacji w komunikacji tradycyjnej; są nimi na przykład spisy treści czy indeksy umieszczone na końcu książki, a także tradycyjne JIW. Współcześnie znajdują one swoją kontynuację w SOW [Sosińska-Kalata 2008, s. 103]. Należą do nich: listy terminów, klasyfikacje i kategorie oraz listy relacyjne. Do tych ostatnich należą tezaury, sieci semantyczne i ontologie [Sosińska-Kalata 2005, s. 156-158]. Ze względu na znaczenie dla budowy Semantycznego Webu (zob. p. 4.9), w dalszej części rozdziału przedstawię ontologie, które stanowią jeden z najbardziej złożonych modeli organizacyjnych [Woźniak-Kasperek 2008, s. 120].

Termin **ontologia**, choć pochodzi z języka greckiego, utworzony został dopiero w XVII w.⁵⁵ z połączenia słów *ontos* (być) oraz *logos* (słowo). W filozofii odnosi się on do istnienia i egzystencji⁵⁶, czyli bytu jako takiego. Dokładniej, jest to nauka o kategoriach rzeczy istniejących lub mogących istnieć w danej dziedzinie⁵⁷ [Sowa 2000, s. 51]. Ontologia dziedziny zawiera rodzaje rzeczy (kategorie), które funkcjonują w tej dziedzinie. Istnieją dwa źródła kategorii ontologicznych: obserwacja i wnioskowanie. Obserwacja dostarcza wiedzy o świecie fizycznym, natomiast wnioskowanie nadaje sens obserwacji poprzez tworzenie bytów abstrakcyjnych.

Wybór kategorii ontologicznych jest pierwszym etapem w projektowaniu bazy danych, bazy wiedzy lub systemu obiektowego. W każdym z tych zastosowań funkcjonuje inna terminologia: w teorii baz danych kategorie nazywane są dziedzinami (*domain*), w sztucznej inteligencji tykami (*type*), a w systemach obiektowych klasami (*class*). Bez względu na stosowaną terminologię, wybór kategorii determinuje zakres możliwej reprezentacji, dostępnej w danej aplikacji komputerowej, a nawet w całej grupie aplikacji. Jakikolwiek braki kompletności, zniekształcenia lub ograniczenia w strukturze kategorii, muszą nieuchronnie ograniczać funkcjonowanie każdego oprogramowania i bazy danych, wykorzystujących te kategorie. Dla odróżnienia zastosowania terminu „ontologia” w filozofii i reprezentacji wiedzy często w tym pierwszym stosuje się liczbę pojedynczą (ontologia), a w drugim – mnogą (ontologie).

Definiując ontologie w sposób niesformalizowany można stwierdzić, że dotyczą one terminologii danej dziedziny (jej słownictwa), wszystkich podstawowych pojęć, stosowanych klasyfikacji, taksonomii, relacji (w tym wszystkich hierarchii i ograniczeń⁵⁸) oraz aksjomatów. Używając bardziej formalnego zapisu można stwierdzić, że dla danej dziedziny **D**, opisywanej za pomocą języka **L** ontologie dostarczają katalogu rodzajów rzeczy, co do których zakłada się, że istnieją w **D**; rodzaje w ontologiach są reprezentowane za

⁵⁵ Wcześniej nazywano ją filozofią pierwszą (Arystoteles) lub metafizyką [Stróżewski 2003, s. 16].

⁵⁶ Przedmiotem ontologii jako dziedziny filozofii jest także nie-byt i nieistnienie, nic.

⁵⁷ Dziedzina ontologii rozumiana jest jako zestaw kategorii ontologicznych, których wybór determinuje możliwości reprezentacji obiektów i ich cech w aplikacji.

⁵⁸ Mowa o ograniczeniach określających opisywane relacje, ich rodzaj i inne cechy. Na przykład opisując relację pomiędzy pojęciami „ojciec” i „syn” należy stworzyć ograniczenie, stwierdzające, że syn może mieć tylko jednego ojca.

pomocą pojęć, relacji i predykatów języka L [Gašević, Djurić, Devedžić 2006, s. 45]. Stosowane definicje ontologii przytoczone zostały w mojej wcześniejszej pracy [Nahotko 2006b, s. 38-43]. Do dalszych rozważań przydatna może być definicja, zaproponowana przez Yannis Kalfoglou, mówiąca, że „ontologie to jawne reprezentacje wspólnego rozumienia ważnych pojęć w określonej dziedzinie”, gdzie słowo „wspólny” oznacza, że ontologie zawierają zasób wiedzy, uzgodnionej przez grupę specjalistów albo całą społeczność [Kalfoglou 2001]. Dzięki uzgodnieniom wiedza zawarta w ontologiach jest obiektywna, w odróżnieniu od subiektywnej wiedzy jednostki. Ontologie wspierają rozprzestrzenianie wiedzy w obrębie grup agentów (ludzi i/lub oprogramowania). Agenci mogą komunikować się między sobą, wymieniając jednoznacznie zdefiniowane pojęcia, za pomocą nośników elektronicznych, w sytuacjach, gdy zwykle nie ma możliwości uzyskania szczegółowych wyjaśnień co do znaczenia pojęć. Ontologie są więc jednym z narzędzi zamiany wiedzy ukrytej w zakresie pojęć i relacji między nimi, w wiedzę jawną.

W reprezentacji wiedzy termin „ontologie” stosowany jest do określania:

- Słownika reprezentacji, często wyspecjalizowanego w zakresie obsługi określonej dziedziny lub tematyki.
 - Obszaru wiedzy opisującej określoną dyscyplinę przy użyciu słownika reprezentacji.
- Oba elementy są ze sobą ściśle powiązane. Zawsze funkcjonuje także podstawowa struktura danych reprezentująca ontologie.

Ontologie są wykorzystywane do reprezentacji wiedzy i w systemach inteligentnych, a także w procesach przetwarzania wiedzy. Ich przydatność przejawia się w zakresie:

- Kontroli słownictwa. Ontologie dostarczają słownictwa (czyli nazw), odpowiadającego terminom danej dziedziny, stąd w ten sposób nazywane są różnego rodzaju słowniki kontrolowane: kartoteki wzorcowe, słowniki haseł przedmiotowych czy tezaury. Różnica pomiędzy takimi słownikami, tworzonymi głównie dla człowieka, a ontologiami polega na tym, że te ostatnie zawierają wyrażenia logiczne opisujące terminy, ich relacje, dozwolone i zakazane, z innymi terminami. Określają także zasady łączenia terminów oraz ich znaczenie, które jest niezależne od użytkowników i kontekstu (na przykład używanego języka) [Hartley 2009, s. 155].
- Hierarchizacji. Ontologie zawierają kategoryzację hierarchiczną lub klasyfikację jednostek w danej dziedzinie. Często tworzone są także grupy (*clusters*) jednostek w oparciu o wspólne cechy ontologiczne. Hierarchie, zwane taksonomiami, przygotowywane są w formie czytelnej i przetwarzalnej maszynowo. Jednak ontologia i taksonomia to nie to samo – ta pierwsza zawiera pełną specyfikację dziedziny.
- Teorii treści. Ontologie można traktować jako teorie treści, identyfikujące klasy obiektów i relacje między nimi (w tym strukturę hierarchiczną), funkcjonujące w danej dziedzinie. Ontologie nie tylko identyfikują te elementy, ale także wyszczególniają je w sposób opisowy, z wykorzystaniem określonych języków reprezentacji ontologii. Specyfikacja klas zawiera ich cechy, wartości cech oraz ograniczenia wartości. Jako teorie treści, ontologie wyjaśniają strukturę wiedzy danej dziedziny. Ich tworzenie powinno być poprzedzone analizą ontologiczną, która ma ujawnić pojęcia stosowane w danej dziedzinie, ich strukturę hierarchiczną i organizacyjną.
- Rozprzestrzeniania i wykorzystywania wiedzy. Głównym zadaniem ontologii nie jest funkcjonowanie jako słownik lub taksonomia; jest to wiedza, która może być udostępniana i wykorzystywana przez aplikacje. Chodzi więc o stworzenie modelu wiedzy danej dziedziny. Pojęcia i relacje, opisywane w ontologiach, są stosowane przez inteligentnych agentów, którzy powinni móc komunikować tego typu wiedzę ontologiczną. Współużytkowane ontologie pozwalają na tworzenie szczegółowych baz wiedzy, opisujących konkretne sytuacje, oparte na wspólnych podstawowych strukturach i organizacji wiedzy. Ontologie umożliwiają więc lub ułatwiają współdziałanie systemów i aplikacji.

Wymienione wyżej cechy powodują, że ontologie bywają uważane za rozwinięcie i następcę tezaurusów (a także innych tradycyjnych JIW). Jednak gdy tradycyjny tezaurus służy do identyfikacji znaczenia i relacji w zestawie terminów, ontologie funkcjonują w odwrotnym kierunku: służą do analizy ram konceptualnych funkcjonujących w ludzkich umysłach oraz zakodowania ich przy pomocy formalnego języka w trakcie procesu, który można zaliczyć do działalności inżynierskiej. Tezaurusy o dobrze zdefiniowanej strukturze logicznej bywają uznawane za ontologie; wydaje się, że takie podejście w niektórych zastosowaniach może być usprawiedliwione [Doerr 2008, s. 8].

Języki reprezentacji ontologii bywają stosowane jako języki metadanych. Ontologie mogą służyć definiowaniu atrybutów, opisowi relacji między nimi i określaniu znaczenia danych. Tego rodzaju ontologie mogą zawierać struktury słowników dla metadanych umożliwiających zarządzanie, wyszukiwanie i udostępnianie zasobów w Web. W takim przypadku należy zapewnić określony poziom jednolitego stosowania terminów, zdefiniowanych w ontologii jako metadane. Częściową jednolitość zapewnia działalność standaryzacyjna organizacji i instytucji zarządzających strukturami metadanych, takich jak Dublin Core Metadata Initiative, która wspomaga rozwój i promocję współdziałających schematów metadanych oraz tworzenie wyspecjalizowanych słowników, służących opisowi różnego rodzaju zasobów (publikacji) elektronicznych.

Schematy metadanych nie są ani ontologiami, ani w ogóle słownikami. Element metadanych jest częścią schematu, który wyraża po pierwsze relacje, a po drugie ograniczenia co do możliwych wartości, które element może przyjmować. Ontologie służące integracji schematów metadanych jawnie definiują dziedzinę, relacje i zakresy wartości, stanowiąc deklarację znaczenia dziedziny i zakresu niezależnie od kontekstu, w postaci metadanych danego obiektu cyfrowego.

Struktury metadanych zbliżone są do zagadnień związanych z metamodelowaniem. Metamodel jest to zbiór składników i zasad niezbędnych do tworzenia modeli szczegółowych w obrębie danej dziedziny. Taki punkt widzenia sytuacji metamodel jako ontologię, gdyż wspomniane składniki i zasady reprezentują jednostki określone w danej dziedzinie oraz relacje między nimi, czyli zbiór składników stosowanych do budowy modeli dziedziny. Innymi słowy metamodel jest ontologią wykorzystywaną przez budowniczych modeli [Gašević, Djurić, Devedžić 2006, s. 73]. Takie podejście wykorzystane zostało w jednej z moich poprzednich prac [Nahotko 2006b].

Pomimo wątpliwości, wyrażanych nie bez podstaw [Bojar 2009, s. 22], ontologie będą zapewne odgrywać coraz ważniejszą rolę w funkcjonowaniu globalnej biblioteki cyfrowej. Umożliwiają one maszynom przetwarzanie informacji dostarczonej przez ludzi, w tym automatyczne wnioskowanie, w sposób odpowiedni do założonego znaczenia danych. W związku z tym odgrywają istotną rolę mediacyjną zarówno pomiędzy użytkownikami i systemami informacyjnymi (serwisami GBC), jak i między ekspertami w zakresie danej dyscypliny i specjalistami technologii informacyjnych. Wspomagają też proces decyzyjny w trakcie projektowania systemów.

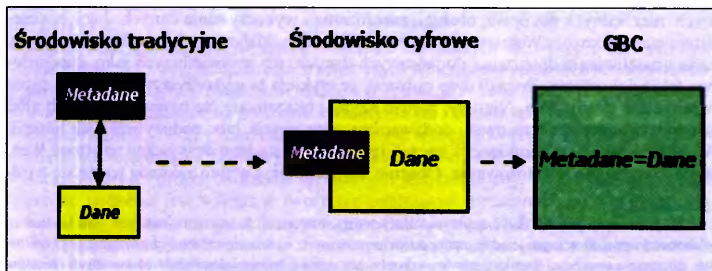
2.8. Metadane

Globalna biblioteka cyfrowa jest przydatna tylko w takim zakresie, w jakim jej poszczególne zasoby i serwisy mogą być odszukane, a treści, które one zawierają, zidentyfikowane. W zasobach globalnej biblioteki cyfrowej, oprócz różnego rodzaju obiektów cyfrowych, o których mowa była w poprzednim rozdziale, funkcjonują także informacje dotyczące tych obiektów, czyli dane o danych, zwane metadanymi. Metadane są to dane ustrukturyzowane na różnym poziomie, dotyczące obiektu cyfrowego, kolekcji obiektów lub części składowych obiektu złożonego, dodane do niego lub z nim powiązane, uła-

twijące jego lokalizację i identyfikację, wyszukanie, wykorzystanie lub zarządzanie nim przez użytkowników – ludzi lub oprogramowanie. W GBC dużą rolę odgrywają także metadane na poziomie serwisu GBC, opisujące go oraz jego kolekcje i umożliwiające jego lokalizację.

Są one częścią struktur opisujących i organizujących informację w taki sposób, aby umożliwić i ułatwić dostęp do niej. Terminowi temu przyznaje się tak wiele znaczeń w tak wielu zastosowaniach, że czasem wydaje się, iż można nim określić wszystko, czyli w sumie nic konkretnie [Day 2005]. Jedna z najprostszych typologii metadanych dzieli je na: metadane opisowe, służące wyszukiwaniu i identyfikacji obiektów cyfrowych; metadane strukturalne, ułatwiające wyświetlanie i nawigację w obiektach oraz metadane administracyjne obejmujące dane przydatne w zarządzaniu obiektem, związane z jego tworzeniem, przechowywaniem, udostępnianiem, pochodzeniem i własnościami.

Metadane mogą być tworzone manualnie przez człowieka (na przykład przez bibliotekarza-katalogera) lub automatycznie przez maszynę (na przykład podczas indeksowania pełnotekstowego) – zob. p. 4.2. W tradycyjnej bibliotece sens istnienia metadanych, w postaci opisów katalogowych, jest oczywisty. Stanowią one surogat dokumentów, gdy bezpośredni dostęp do tych ostatnich nie jest możliwy i musi odbywać się poprzez katalogi. Potrzeby posiadania metadanych i ich zasobów (katalogów) nie usunął nawet wolny dostęp do półek. W środowisku globalnej biblioteki cyfrowej udostępnianie informacji wymaga identyfikacji obiektów istniejących zarówno tam (w Sieci), jak i poza nią, ich ścisłej lokalizacji i określenia formy. Użytkownicy GBC muszą mieć możliwość odszukania zasobów, którymi są zarówno serwisy GBC, serwery, pliki, obiekty cyfrowe, w tym DLO, jak i metadane funkcjonujące jako surogaty dla artefaktów fizycznych. W globalnej bibliotece cyfrowej zawartość tekstowych obiektów cyfrowych jest źródłem wszystkich, niezbędnych metadanych opisowych (obiekt można opisać dzięki indeksowaniu pełnotekstowemu). Można więc wskazać kierunek rozwoju metadanych, spowodowany przejściem od środowiska w pełni analogowego (publikacji drukowanych) do środowiska w pełni cyfrowego GBC. Rozwój ten przedstawiłem na rys. 6. W środowisku tradycyjnym metadane funkcjonują odrębnie od dokumentów fizycznych: istnieją dwa wyraźnie wyodrębnione, chociaż powiązane, zbiory danych i metadanych. Po powstaniu Internetu i rozpowszechnieniu obiektów cyfrowych różnice te zaczęły zanikać, a metadane stawały się częścią danych. Sytuacja ta zmierza w kierunku całkowitego zrównania danych z metadanymi: tekstem służącym wyszukiwaniu jest nie zawartość rekordu metadanych, ale cały tekst dokumentu. W środowisku cyfrowym jedyna różnica pomiędzy metadanymi a danymi polega na tym, że metadanymi nazywa się znany użytkownikowi, poszukiwany element obiektu cyfrowego (jak nazwisko autora), a danymi element tego samego obiektu, który jeszcze nie jest znany (na przykład dowolny element treści) [Weinberger 2007, s. 104].



Rys. 6. Przejście od metadanych do danych

Istnienie danych i metadanych w formie cyfrowej umożliwia tworzenia dowolnych relacji między treściami, realizującymi potrzeby autorów, użytkowników, wydawców i innych. Często powstają one w sposób niejawni, na przykład w trakcie automatycznej obserwacji sposobów posługiwania się przez uczzonego narzędziami wyszukiwawczymi podczas poszukiwania rozwiązania problemu. W wyniku aktywności użytkowników powstaje gęsta sieć odnośników i zbiory etykiet. Potrzeba manualnego (jawnego) tworzenia metadanych opisowych, będących odpowiednikiem tradycyjnego opisu formalnego i rzeczowego może wydawać się więc problematyczna⁵⁹.

Metadane tworzone przez człowieka, zarówno jawnie, jak i niejawnie, są przydatne do realizacji takich funkcji, jak:

- Ułatwienie komunikacji i wymiany danych pomiędzy serwisami globalnej biblioteki cyfrowej.
- Rozbudowa opisu obiektu, przydatna z różnych względów: słowa i terminy użyte przez autora są niepoprawne, albo też dokument w ogóle jest nietekstowy (grafika, obiekt trójwymiarowy, nagranie dźwiękowe, film); również dane z badań naukowych bez określenia kontekstu (opis badań, w tym metoda, narzędzia) stanowią tylko nic nieznaczące tablice liczb lub ciekawe obrazki [Borgman 2008, s. 34].
- Dostarczanie danych uzupełniających o obiekcie, takich jak jego lokalizacja lub sposób wykorzystania; w ten sposób rozszerza się pojęcie opisu jego „treści”.
- Dostarczanie informacji o przeszłości obiektu, takiej jak proveniencja obiektu, zmiany w jego rozmiarach i inne modyfikacje.
- Zabezpieczenie opisywanego obiektu, fizycznego lub cyfrowego, za pomocą cyfrowych surogatów szeroko rozpowszechnianych w Sieci, które pozostają w niej nawet po usunięciu opisywanego obiektu.
- Wprowadzanie czynnika intelektualnej organizacji do zasobu obiektów cyfrowych, przez ich grupowanie na zasadzie wykazywania podobieństw i odrębności, na podstawie określonych kryteriów i zasad, na przykład takich, jak wskazane w FRBR.
- Przedstawianie (meta)danych o określonych, istotnych cechach obiektu.
- Standaryzacja opisu obiektu.
- Ułatwienie administrowania i obsługi technicznej poszczególnych obiektów, zasobów i serwisów GBC.

Metadane te traktowane mogą być przez narzędzia indeksujące tak, jak każdy tekst, specyficzne w tym, że są surogatami-suplementami danych, i także poddawane indeksowaniu pełnotekstowemu, którego efekty wspomagają indeksowanie danych. Paradoksalnie więc okazać się może, że rekordy metadanych w serwisach GBC tworzone są bardziej dla wyszukiwarek internetowych niż żywych użytkowników.

Termin „metadane” po raz pierwszy zastosowany został w kontekście systemów zarządzania bazą danych jako ogólna nazwa dla wszystkich rodzajów danych uzupełniających, niezbędnych dla opisu, obsługi, zarządzania i wykorzystania danych. Już we wczesnym etapie rozwoju Webu, w syntaktyce HTTP, przewidziana została etykieta <meta> w celu umożliwienia dołączania dodatkowych danych, nie wyświetlanych jako standardowy tekst dokumentu. Początkowo sądzono, że etykieta ta wykorzystywana będzie przez właściciela strony Web. Niestety, bardzo szybko okazało się, że twórcy stron Web albo nie byli w ogóle zainteresowani dodawaniem metadanych, albo nadużywali tych funkcji. W efekcie twórcy wyszukiwarek zaczęli ignorować metadane wpisywane na stronę Web, więc zaprzestano ich stosowania. Obecnie nowe nadzieje w tym zakresie wiąże się z mi-

⁵⁹ W pierwszych pracach dotyczących metadanych często tworzone kategorie metadanych jako kontinuum głębokości opisu, szeregując je od najmniej ustrukturyzowanych, za które uważano indeksowanie pełnotekstowe, po ręcznie tworzone, ustrukturyzowane rekordy, zawierające bogatą interpretację opisywanych obiektów [Weibel, Iannella, Cathro 1997]. Obecnie techniki indeksowania pełnotekstowego, szczególnie rozwijane przez wyszukiwarki globalne zostały na tyle rozbudowane, że takie proste podziały straciły rację bytu.

kroformatami (zob. p. 4.2). Jak twierdzi Terrence Brooks, właściwe funkcjonowanie wyszukiwarek typu Google wymaga od ich użytkowników pewnego rodzaju „ignorancji”, czyli braku pełnej wiedzy o sposobie ich działania [Brooks 2004]. Wyszukiwarki muszą uwzględniać przede wszystkim teksty bezpośrednio udostępniane użytkownikowi. Stosowanie metadanych przeniosło się więc ze środowiska stron prywatnych i komercyjnych do obszaru publikacji naukowych o kontrolowanych treściach (w tym ocena jakości), takich, jak portale dziedzinowe, biblioteki cyfrowe, repozytoria, czasopisma elektroniczne i inne. Umieszczane tam bazy metadanych mogą być indeksowane przez wyszukiwarki tak, jak każdy inny tekst (o ile spełniają warunki techniczne, na przykład nie są częścią ukrytego Webu). Tworzenie metadanych w repozytoriach staje się ich wąskim gardłem, stąd propozycje zamieszczania obiektów nawet z minimalnymi metadanymi, które później, w razie potrzeby zostaną uzupełnione. Najważniejsza bowiem wydaje się możliwość dotarcia do elektronicznej wersji dokumentu, nawet pozbawionej rozbudowanych metadanych [Adams 2008].

Gwałtowny rozwój zastosowań sieciowych miał dwojaki wpływ na stosowanie metadanych. Z jednej strony, doprowadził do powstania wielkiej liczby nowych standardów metadanych, wyspecjalizowanych w obsłudze określonych dziedzin wiedzy i działalności praktycznej, treści, funkcji i zastosowań. Z drugiej strony, tak wielka różnorodność rozwiązań spowodowała stosowanie metadanych do zapewnienia zarówno technicznego, jak i semantycznego współdziałania między systemami, co ma zasadnicze znaczenie dla funkcjonowania globalnej biblioteki cyfrowej. Metadane wspomagają w szczególności uzyskanie współdziałania semantycznego, osiąganego głównie za pomocą tworzenia tablic konwersji (mapowania) pomiędzy schematami metadanych oraz wykorzystania schematów powszechnie znanych i stosowanych, takich jak Dublin Core, jako swego „przełącznika”. Dla osiągnięcia wyższego poziomu współdziałania semantycznego systemy muszą posiadać zdolność do wymiany danych o precyzyjnie określonym znaczeniu oraz do tłumaczenia znaczenia danych do formy zrozumiałej dla systemu. W tym zakresie stosowane są ontologie.

Dalszy wzrost roli metadanych wiązany jest z zastosowaniami naukowymi, charakteryzującymi się produkowaniem olbrzymich zasobów danych. Istnienie metadanych o właściwej jakości umożliwia optymalne wykorzystanie zgromadzonych danych. Metadane stają się niezbędnym warunkiem tworzenia narzędzi badawczych następnej generacji, takich jak wyspecjalizowane wyszukiwarki, narzędzia służące eksploracji danych i ich wizualizacji, stanowiących podstawę tworzenia wyspecjalizowanych serwisów globalnej biblioteki cyfrowej. Są także zasadniczym elementem IIN, w zakresie obsługi i ochrony danych.

Więcej informacji na temat metadanych tworzonych przez wyszukiwarki przedstawie w dalszej części rozdziału. Tutaj poprzestane na najważniejszych cechach schematów metadanych, tworzonych przez człowieka (nazywanych metadanymi „ustrukturyzowanymi”). Są to:

- **Modułowość:** pozwala projektantowi schematu metadanych na tworzenie nowych twórców w oparciu o wcześniej istniejące schematy. Elementy danych z różnych schematów, słowników czy innych elementów składowych, można ze sobą łączyć w sposób umożliwiający współdziałanie na poziomie syntaktycznym i semantycznym.
- **Rozszerzalność:** twórcy schematów metadanych muszą stworzyć możliwość ich przyszłego rozszerzania i rozbudowy w przypadku zaistnienia nowych potrzeb.
- **Szczegółowość:** proces tworzenia metadanych jest kosztowny, więc z ekonomicznego punktu widzenia jest wskazane tworzenie schematów metadanych na tyle szczegółowych, aby spełniały stawiane im wymagania funkcjonalne i tylko te.
- **Wielojęzyczność:** jeżeli schematy metadanych mają być stosowane w tak zróżnicowanym środowisku jak Internet, powinny one brać pod uwagę wielość stosowanych ję-

zyków i odmienność kulturową. Wielojęzyczność można traktować jako element szerszych zjawisk, ogólnie określanych mianem wielokulturowości [Duval i in. 2002].

Zastosowanie przedstawionych powyżej zasad kształtuje praktykę funkcjonowania schematów metadanych. Jej elementem są:

- Profile aplikacyjne. Trudno oczekiwać, aby jakiegokolwiek schemat metadanych spełniał potrzeby funkcjonalne we wszystkich aplikacjach. Istotne znaczenie ma jednak zachowanie standaryzacji rozwiązań. Ułatwiają to profile aplikacyjne: zestawy elementów metadanych, pochodzących z różnych schematów i tworzących łącznie nowy schemat metadanych, odpowiedni do lokalnych potrzeb. Jednocześnie zapewnione jest współdziałanie pomiędzy ogólnymi i lokalnymi schematami. Jest to praktyczna realizacja zasady modułowości i rozszerzalności.
- Syntaktyka i semantyka. Syntaktyka dotyczy formy metadanych, natomiast semantyka ich znaczenia. Współpraca w zakresie wymiany metadanych wymaga współdziałania w obu zakresach. Niezbędne jest jak najdalej idące oddzielenie, niezależnienie semantyki od syntaktyki, dzięki czemu elementy metadanych, tworzone w różnych okresach, mogą być wyrażane w różnych schematach syntaktycznych (HTML, XML lub RDF i ich odmianach), zgodnie z potrzebami i bez zmiany ich znaczenia. Dzięki zastosowaniu właściwej, standardowej i ogólnie używanej syntaktyki, takiej jak XML, powstaje możliwość indeksowania metadanych, a więc także ich semantycznego wyszukiwania za pomocą powszechnie stosowanych narzędzi, jakimi są wyszukiwarki internetowe. Stosowanie syntaktyki przeznaczonej do wybranych zastosowań, jak ISO 2709, powoduje przesunięcie danych do ukrytego Webu.
- Sposoby łączenia metadanych z opisywanymi obiektami. W tym zakresie wyróżnia się zazwyczaj meta dane osadzone, wyodrębnione i tzw. trzeciej strony” [Nahotko 2004, s. 29]. Każdy obiekt cyfrowy można opisać w wielu rekordach metadanych, tworzonych do różnych celów i z różnych punktów widzenia (na przykład opisy tworzone przez autora, wydawcę i bibliotekarza).
- Identyfikacja i nazewnictwo elementów metadanych. Globalny zasięg przestrzeni nazw Web URI oznacza, że każdy element danych w schemacie metadanych może być reprezentowany przez globalną nazwę (swoje URI). Zastosowanie niezmiennych identyfikatorów ułatwia maszynowe przetwarzanie metadanych, jednak może powodować kłopoty przez swoją nienaturalną formę, daleką od języka naturalnego. Identyfikatory oparte na URI są trudne w użyciu dla ludzi, którzy przyzwyczajeni są raczej do nazw w języku naturalnym. W związku z tym wiele systemów stosuje bardziej przyjazne etykiety, lepiej zrozumiałe i przydatne w danym kontekście językowym, kulturowym lub dyscypliny.
- Rejestry metadanych. Wraz ze wzrostem liczby schematów metadanych i profili aplikacyjnych, rośnie także znaczenie rejestrów opisujących te zasoby. Oczekuje się, że rejestry metadanych dostarczą narzędzi dla właściwej identyfikacji istniejących schematów i profili aplikacyjnych, pozwalając także na łatwą konwersję metadanych pomiędzy poszczególnymi schematami.
- Odpowiednia kompletność opisu. Nie zawsze konieczne jest wykorzystywanie wszystkich elementów schematu metadanych; użytkownik powinien być wspomagany w dokonywaniu właściwych wyborów. Szczegółowość opisów metadanych uzależniona jest od polityki i praktyki agend tworzących meta dane, a te z kolei wynikają z wymagań funkcjonalnych oferowanych usług i aplikacji.
- Obowiązkowość/opcjonalność elementów. Projektowanie standardów metadanych dla środowiska globalnego wymaga wysokiego poziomu elastyczności. Element niezbędny w jakimś zastosowaniu może w ogóle nie mieć zastosowania w innym. Dlatego tylko niewiele elementów metadanych w schemacie ogólnym powinno mieć status obowiązkowych. Z drugiej strony, w obrębie poszczególnych zastosowań, wymagania co do me-

tanym powinny być mocniej ujednolicone, co powinno pociągnąć za sobą potrzebę większej standaryzacji i uniformizacji, na przykład w formie profili aplikacyjnych.

- Elementy metadanych formalne i rzeczowe. Podobnie jak w przypadku tradycyjnego opracowania zbiorów, również w zakresie metadanych opisowych wyróżnić można elementy opisu formalnego (na przykład autor, data utworzenia, wersja) oraz rzeczowego (na przykład słowa kluczowe, spis treści, abstrakt). Nowym zjawiskiem jest stosowanie części elementów, uznawanych tradycyjnie za formalne, do wyszukiwania rzeczowego; takim elementem jest na przykład tytuł dzieła lub serii.

Każdy polski, ale chyba niewyłącznie, bibliotekarz słysząc o metadanych pomyśli o dwóch standardach opisu dokumentów, w tym elektronicznych: formacie MARC i schemacie Dublin Core⁶⁰. Oba standardy pomyślane były jako swoje przeciwieństwa z punktu widzenia ich złożoności [Witten, Bainbridge 2003, s. 253]. MARC jest formatem bardzo bogatym, pozwalającym na pełen opis każdego rodzaju dokumentu, rozbudowanym i skrupulatnie zarządzanym, tworzonym przez bibliotekarzy i katalogerów najwyższej klasy od dziesiątek lat. Powszechnie stosuje się go w zintegrowanych systemach bibliotecznych. Jego używanie wymaga przygotowania teoretycznego i wieloletniej praktyki.

Dublin Core jest standardem, który w zamierzeniu miał umożliwić tworzenie minimalnego zestawu danych, dotyczących obiektów internetowych przez osoby często przypadkowe, które w żaden sposób nie były szkolone w tym zakresie; mogły one być na przykład autorami stron Web⁶¹, chcącymi zwiększyć ich dostępność i indeksowalność przez wyszukiwarki. Oba te schematy są interesujące nie tylko z powodu ich powszechnego obecnie stosowania, ale także ze względu na leżące u ich podstawy bardzo różne punkty widzenia na rolę opracowania dokumentów. Wkrótce po rozpowszechnieniu się Dublin Core okazało się zresztą, że z biegiem czasu oba standardy upodabniają się do siebie: MARC przystosowano do opisu dokumentów elektronicznych, w tym sieciowych, a Dublin Core został tak znacznie rozbudowany, że do korzystania z pełnej jego wersji potrzebne staje się pewne przygotowanie i doświadczenie.

Tworzenie rekordu MARC dla określonego dokumentu jest żmudnym zajęciem, obwarowanym szczegółowymi przepisami katalogowania, nie będącymi częścią formatu. W krajach anglosaskich przepisami tymi są AACR⁶², w Polsce stosowany jest ISBD.

Kodowanie danych w formacie MARC za pomocą ISO 2709 utrudnia ich wykorzystanie przez wyszukiwarki internetowe. W celu zmiany tej sytuacji powstają zasady zapisu danych MARC w syntaktyce XML. Przykładem jest MARCXML, schemat utworzony w Library of Congress dla umożliwienia pracy z danymi MARC w środowisku XML. Elementy danych MARC są dokładnie odtworzone w syntaktyce XML, co umożliwia dalsze transformacje, na przykład w kierunku konwersji do DC lub realizację procesów, takich jak walidacja⁶³. Schemat zachowuje semantykę formatu MARC; pola są traktowane jak elementy, których atrybutami są etykiety i wskaźniki, a podpola jak podelementy z etykieta podpola w roli atrybutu.

⁶⁰ Pełna nazwa: Dublin Core Metadata Element Set.

⁶¹ Dublin Core stawiany był w ten sposób niejako w opozycji do formatów typu MARC, które uważane są za trudne w implementacji i nie przeznaczone dla przypadkowego użytkownika. Podobny sposób myślenia reprezentują różnego rodzaju narzędzia stosowane w tzw. Webie 2.0, na przykład folksonomie, które są strukturami uwzględniającymi możliwości zupełnie nieprzygotowanego użytkownika, wykorzystującymi tzw. inteligencję zbiorową, czyli „mądrość tłumów”.

⁶² Od końca 2009 r. AACR miało być zastępowane przez RDA (Resource Description and Access – <http://www.rdaonline.org/>), przepisy katalogowania dostosowane do potrzeb biblioteki cyfrowej. Mają one być oparte na modelach FRBR i FRAR; przepisy tworzenia opisów mają zostać oddzielone od zasad ich wyświetlania (prezentacji), a ich budowa będzie hierarchiczna, od ogólnych zasad do szczegółowych przepisów [Kiorgard, Kartus 2006].

⁶³ Walidacja metadanych służy badaniu poprawności opisu.

Autorzy raportu dotyczącego przyszłości kontroli bibliograficznej zwracają uwagę na potrzebę stosowania nowych struktur danych [Library of Congress 2008, s. 24]. MARC oparty jest na technice zarządzania danymi sprzed 40 lat, co powoduje, że nie odpowiada współczesnym standardom tworzenia oprogramowania. Stosowany jest wyłącznie w bibliotekach, w innych aplikacjach bywa wykorzystywany do przekazywania danych. Nowe sposoby wykorzystywania danych bibliograficznych wymagają formatów pozwalających na gromadzenie metadanych, w tym adnotacji (opinii, komentarzy) i danych o wykorzystaniu tworzonych przez ekspertów, użytkowników i narzędzia zautomatyzowane.

Krytyce podlegają także przepisy katalogowania w rodzaju AACR, ISBD czy RDA. Według Caren Coyle i Diane Hillmann, oparte są na nieaktualnej ocenie sytuacji i zamkniętych już celach kontroli bibliograficznej [Coyle, Hillmann 2007]. Tymczasem sytuacja w tym zakresie zmieniła się i nadal ulega modyfikacjom: użytkownicy coraz mniej czasu poświęcają opisom bibliograficznym, a coraz więcej wyszukiwaniu pełnotekstowemu; coraz rzadziej wyszukują, a częściej współpracują w sieciach społecznych, dostarczających im poszukiwanych informacji. W związku z tym potrzebne jest zupełnie nowe spojrzenie na zasady opisu obiektów cyfrowych. Zamiast spisu szczegółowych zasad, uwzględniających wiele wyjątków, niezbędne jest stworzenie warunków zapewniających łańcuch i efektywność stosowania tworzonych narzędzi.

Jak piszą wspomniane autorki, wśród bibliotekarzy przyzwyczajonych do tradycyjnych standardów, możliwie pełny i szczegółowy opis katalogowy jest czymś w rodzaju wyidealizowanego celu. Uważają oni, że tego typu opis ma zasadnicze znaczenie dla funkcjonowania nauki i nie zauważają mnożących się problemów, nawet, gdy Google przejmują ich użytkowników. Jakiegokolwiek zmiany wychodzące poza linearny model katalogowania, oparty na katalogu kartkowym, związany ze stosowaniem zasad AACR2 czy podobnych, przyjmowane są nieprzychylnie. Bibliotekarze nie wierzą, że prostsze lub mniej ustyturowane zasady katalogowania, w znacznie większym stopniu wykorzystujące możliwości współczesnych komputerów, mogą zapewniać usługi na odpowiednim poziomie. W szczególności odrzucane są wszelkie zmiany godzące w zasady ujednolicenia opisów.

Jak wspominałem, schemat metadanych Dublin Core Metadata Element Set (DCMES) powstawał w pewnej opozycji do formatów MARC. W zamiśle twórców schematu Dublin Core i zgodnie ze swoją nazwą, zestaw elementów ma stanowić „rdzeń”, minimum niezbędne do tworzenia opisu dokumentu tekstowego. W razie potrzeby schemat ten może być rozbudowywany i uszczegóławiany przez użytkowników zgodnie z lokalnymi potrzebami. Nie ma elementów obowiązkowych, wszystkie są powtarzalne.

Od czasu powstania pierwszego zestawu elementów DCMES schemat ten uległ znacznej rozbudowie. Utworzenie modelu abstrakcyjnego Dublin Core⁶⁴ oraz tzw. Singapore Framework⁶⁵ stanowi poważną jego modyfikację, dzięki której obecnie Dublin Core stanowi podstawę tworzenia opisów dowolnego rodzaju dokumentów. Wymienione prace podstawowe wykorzystane zostały m.in. do budowy profili aplikacyjnych: SWAP (Scholarly Works Application Profile), opartego na modelu FRBR, służącego opisowi prac naukowych⁶⁶ [Allinson, Johnston, Powell 2007] oraz DCCAP (Dublin Core Collections Application Profile), stosowanego do opisu na poziomie kolekcji⁶⁷ [Lourdi, Papatheodorou, Doerr 2009].

⁶⁴ Dublin Core Abstract Model: <http://dublincore.org/documents/abstract-model/>.

⁶⁵ Singapore Framework przyjęty został podczas warsztatów DC w Singapurze w 2007 r. Składa się na niego zestaw dokumentów tworzących łącznie Profil Aplikacyjny Dublin Core (DCAP). Dokumenty te opisują między innymi wymagania funkcjonalne, modele dziedziny i syntaktykę oraz zawierają wskazówki dla użytkowników. Są one pomocne w tworzeniu profili aplikacyjnych i w ich formalnej ocenie.

⁶⁶ Zawiera on, oprócz podstawowych elementów metadanych, także elementy dodatkowe, takie jak źródło finansowania badań oraz numer grantu.

⁶⁷ Pozwala na opis kolekcji różnego rodzaju, a także katalogów i indeksów.

W Dublin Core, w porównaniu z MARC, słabo zorganizowana jest kontrola autorytar-na (istnieją na przykład wykazy terminów używanych dla elementów Typ i Format), przez co bardzo łatwo dwie osoby (a nawet ta sama osoba w większych odstępach czasu) mogą stworzyć dwa, różne opisy tego samego źródła. Schemat jest jednak wciąż rozwijany (co, jak mówiłem, już wpływa na wzrost stopnia jego skomplikowania), między innymi w kierunku tworzenia zestawów zdefiniowanych wartości dla wybranych elementów, jako sposobu na zapewnienie jednolitości opisu. Określono na przykład przydatne systemy klasyfikacyjne i inne języki informacyjno-wyszukiwawcze, możliwe do zastosowania w elemencie Opis rzeczowy. Oprócz tworzenia zestawów wartości poszczególnych elementów, standaryzacja polega również na tworzeniu reguł kodowania wartości elementów, na przykład dat czy języków.

Znaczenie elementów Dublin Core może być uszczegółowiane (przez tworzenie tzw. kwalifikatorów), co również podlega w pewnym stopniu standaryzacji, gdyż DCMI⁶⁸ podaje wykazy wyróżnionych przez siebie kwalifikatorów. Dla przykładu element Data może być kwalifikowany przez utworzenie bardziej szczegółowych elementów: Data utworzenia, Data ważności, Data dostępności, Data wydania, Data modyfikacji i jeszcze innych, według lokalnych potrzeb. Element Opis jest uszczegółowiany przez utworzenie podelementów Abstrakt lub Spis treści, natomiast element Relacja zawiera takie związki, jak Jest wersją, Jest częścią, Zastępuje, Wymaga, Odsyła wraz z relacjami odwrotnymi (typu Jest wersją – Ma wersję).

Powodem częstego wykorzystywania Dublin Core jest możliwość rozwiązywania za jego pomocą kilku istotnych problemów, związanych także z funkcjonowaniem GBC. Po pierwsze, potrzebny jest prosty, powszechnie znany format opisu różnego rodzaju zasobów; po drugie, możliwe jest stosowanie wszelkich używanych standardów syntaktyki, od HTML/XHTML przez XML po RDF, dzięki czemu może on być użyteczny w wielu serwisach GBC, w tym w Semantycznym Webie; po trzecie, może być on wspólną podstawą, „najniższym wspólnym mianownikiem” dla semantyki różnych, czasem bardziej obszernych schematów metadanych, co wpływa na ich zdolność do współdziałania. Twórcy tego schematu metadanych starają się także, aby był on używany w różnych kulturach i językach, promując jego tłumaczenia i tworząc mechanizmy transliteracji.

Obecnie sytuacja w zakresie schematów metadanych stała się nadzwyczaj złożona. Wystarczy wymienić przykłady stosowanych standardów: AACR2/RDA/ISBD, MARC 21, MARCXML, METS, MODS, Dublin Core, czy wiele innych, takich jak ONIX⁶⁹ (stosowany w środowisku wydawców), MIX⁷⁰ i VRA⁷¹ (obiekty graficzne, muzealnictwo), IMS⁷² (materiały edukacyjne) oraz CSMD (opis badań naukowych i danych z eksperymentów). Ta różnorodność utrudnia zarządzanie, prowadzi do nieporozumień w trakcie implementacji i spowalnia rozwój schematów. Z drugiej strony metadane, służące wyszukiwaniu informacji w Internecie, powstają podczas pełnotekstowego indeksowania, zarówno treści dokumentów, jak i rekordów metadanych, stąd maleje znaczenie ich ustrukturyzowania. Zauważalną tendencją jest konsolidacja schematów metadanych, przez zastosowanie XML jako platformy ułatwiającej współdziałanie, zapewniającej ujednoczoną syntaktykę, zrozumiałą dla narzędzi wyszukiwawczych. Szansę na unifikację semantyki wiąże się z RDA, w znacznie większym stopniu niż poprzednie standardy uwzględniającym potrzeby opisu obiektów w formatach cyfrowych [Gartner 2008, s. 14]. Schematy przedstawione poniżej są przykładem prac służących zwiększeniu indeksowalności metadanych.

⁶⁸ DCMI (Dublin Core Metadata Initiative) jest instytucją powołaną dla kierowania pracami nad rozwojem DCMES (<http://dublincore.org/>).

⁶⁹ ONline Information eXchange (<http://www.editeur.org/onix.html>).

⁷⁰ Metadata for Images in XML Schema (<http://www.loc.gov/standards/mix/>).

⁷¹ Virtual Resources Association (<http://www.vrweb.org/projects/vrcore4/index.html>).

⁷² IMS Learning Resource Metadata Information Model (<http://www.imsglobal.org/metadata/index.html>).

METS (Metadata Encoding and Transmission Standard) powstał z myślą o kodowaniu metadanych opisowych, administracyjnych oraz strukturalnych dla obiektów elektronicznych i łączeniu ich we wspólne pakiety, co ma ułatwić zarządzanie obiektami w repozytoriach cyfrowych, a także ich wymianę pomiędzy repozytoriami [Tedd, Large 2005, s. 96]. Jego obsługą zajmuje się Network Development and MARC Standards Office⁷³, funkcjonujący w Library of Congress wraz z Digital Library Federation. Idea pakietowania danych wraz z metadanymi była już wcześniej przedstawiana; jednym z najbardziej znanych modeli jest Warwick Framework powstały podczas drugich warsztatów Dublin Core w 1996 r. Była to architektura conceptualna dla logicznej agregacji różnych rodzajów metadanych w pakiety zwane kontenerami.

METS jest to Schemat XML dostarczający słownictwa i syntaktyki, służących tworzeniu informacji o obiektach cyfrowych w postaci „pakietów” metadanych opisowych, administracyjnych i strukturalnych, spisu plików oraz opcjonalnie samych danych złożonych obiektów cyfrowych [McDonough 2006, s. 149].

Zarządzający formatem METS wybrali niektóre standardy metadanych jako te, które najlepiej współpracować będą z tym formatem. Wśród metadanych opisowych znalazły się Dublin Core, EAD, MODS⁷⁴ (Metadata Object Description Schema) oraz MARCXML i MIX. METS stosowany jest w kilku produktach, zarówno typu open source (Fedora, Greenstone), jak komercyjnych (VITAL firmy VTLS).

MODS jest schematem opisu źródeł, wyrażonym jako Schemat XML. Został on utworzony we współpracy Library of Congress z Indiana Univ., OCLC i Univ. of Chicago. Jego etykiety nie są numeryczne, podobnie jak w Dublin Core, dzięki czemu zapewniając współpracę z formatem MARC (choć zawiera tylko podzbiór jego pól), może wspomagać tworzenie własnych opisów. Jest on szczególnie polecany do opisu zasobów dziedzictwa kulturowego i humanistyki. Dzięki rozróżnieniu pomiędzy treściami intelektualnymi źródła oraz jego formatem cyfrowym i lokalizacją udało się ograniczyć problemy związane z opisem obiektów cyfrowych, utworzonych z analogowych oryginałów. MODS upraszcza złożoną strukturę MARC, jednak jednocześnie wyraźnie ogranicza jego elastyczność [Goldsmith, Knudson 2006].

Wszystkie elementy MODS są opcjonalne (oprócz Tytułu) i powtarzalne (oprócz Informacji o rekordzie). Wersja MODS Lite zawiera tylko te pola z formatu MARC 21, które można bezpośrednio konwertować do Dublin Core.

Schemat **CSMD** (Core Scientific MetaData) służy realizacji nieco innych zadań, niż przedstawione wcześniej: ma pozwalać na opis przedsięwzięć naukowych oraz surowych danych powstających w ich wyniku. Rozwijany przez Science & Technology Facility Council⁷⁵ (STFC), składa się z kilku części:

- Zakres badań (słowa kluczowe, hasła przedmiotowe).
- Opis badań (nazwa, instytucja, zespół badawczy, rodzaj badań).
- Czynności badawcze (metoda, opis, niezbędne zasoby).
- Zasoby danych (opis, kolekcja, związki z czynnościami, identyfikator).
- Warunki dostępu.

CSMD pozwala na szczegółowy opis prowadzonych badań i danych uzyskanych w ich wyniku.

Poniżej prezentuję **zasoby metadanych opisowych** (danych bibliograficznych), dostarczane przez biblioteki i dostawców baz danych. Zostały one przedstawione łącznie w jednym miejscu, gdyż w globalnej bibliotece cyfrowej zasoby te tworzą coraz bardziej wyodrębnioną grupę serwisów GBC, stanowiąc w dużej części zasoby ukrytego Webu. Zagadnienia związane z wyszukiwaniem w tych zasobach będzie mowa w p. 4.3.

⁷³ <http://www.loc.gov/standards/mets/>.

⁷⁴ <http://www.loc.gov/standards/mods/>.

⁷⁵ <http://www.scitech.ac.uk/>.

Jak twierdzi Michael Nentwich, obecnie nieomal wszystkie wykazy zasobów bibliotek naukowych, a często także same zasoby dostępne są online [Nentwich 2003, s. 78]. Wyróżnić można następujące rodzaje tych wykazów:

- Katalogi biblioteczne (OPAC) są bibliograficznymi bazami danych, zawierającymi z zasady jedynie metadane opisowe dokumentów, głównie tradycyjnych, ale również zdarzają się opisy dokumentów elektronicznych (o dostępie lokalnym i zdalnym). Jako format danych stosuje się MARC w różnych odmianach, wraz z ISO 2709. Opisy rzadko pozwalają na bezpośrednie przejście do pełnego tekstu dokumentu. Można więc określić je jako elektroniczne wersje dawnych katalogów kartkowych, tylko oferujące większe możliwości wyszukiwawcze i inne usługi dodatkowe – zob. p. 4.3.
- Biblioteki cyfrowe wraz z cyfrowymi dokumentami pełnotekstowymi udostępniają także ich metadane opisowe. Zazwyczaj zebrane są one w formie bazy danych⁷⁶, rzadziej dostęp jest poprzez listę dostępnych obiektów⁷⁷. Metadane czasem występują w formacie MARC, jednak przeważnie w jednym z nowych schematów metadanych. Przeważa schemat Dublin Core, coraz częściej w powiązaniu z syntaktyką, opartą na XML. Dodatkowo tworzone są indeksy, powstałe w wyniku indeksowania pełnotekstowego tych dokumentów.
- Repozytoria, gromadzące bardzo zróżnicowane materiały, głównie jednak artykuły naukowe. Metadane mogą być tu gromadzone centralnie, wraz z tekstami dokumentów lub repozytorium może zawierać wyłącznie metadane i odpowiednie indeksy, a pełne teksty są rozproszone i udostępniane na serwerach wydawców; w taki sposób funkcjonuje na przykład archiwum RePEc⁷⁸.
- Bazy wiedzy, szczególnie zgromadzonej w Webie, są specjalnymi bazami danych, służącymi do udostępniania nie danych czy informacji, ale wiedzy. Zapewniają one reprezentację wiedzy w określony, sformalizowany sposób (na przykład w formie klasyfikacji czy ontologii), wraz z połączeniami pomiędzy jej elementami, zapisanymi w bazie danych. Bazy wiedzy stanowią będą podstawę Semantycznego Webu.
- Portale internetowe nie stanowią zasobów obiektów cyfrowych, zgromadzonych w jednym miejscu (serwerze), ale zbierają odesłania do zasobów zewnętrznych, rozproszonych w sieciach rozległych. Dla użytkowników oferują one tzw. metadane trzeciej strony, czyli wykonane przez niezależną instytucję, poza autorem i wydawcą. Portal jest więc serwisem GBC, składającym się wyłącznie z metadanych, zawierających m.in. odnośniki do dokumentów pierwotnych. Przykładem takich zasobów są projekty realizowane w ramach RDN⁷⁹ (Resource Discovery Network).
- Biblioteki hybrydowe udostępniają zarówno dokumenty tradycyjne, jak i cyfrowe. Często zdarza się, że dla obu rodzajów zbiorów powstają odrębne zasoby metadanych (katalogi), w których metadane tworzone są według różnych zasad, w różnych formatach [Grabowska 2007, s. 27]. Dla właściwej obsługi informacyjnej użytkowników niezbędne jest zintegrowanie wszystkich narzędzi wyszukiwawczych w sposób umożliwiający odszukanie wszystkich potrzebnych materiałów w jednym miejscu. Rozwój prac w tym zakresie wskazuje, że miejscem tym będzie wyszukiwarka internetowa.

Większość bibliotek, tradycyjnych, cyfrowych i hybrydowych, oprócz własnych zasobów metadanych udostępnia także komercyjne bazy danych, tworzone poza biblioteką. Bazy te są niezależne od bibliotek i mogą być udostępniane bezpośrednio poprzez sieć

⁷⁶ Przykładem mogą być polskie biblioteki cyfrowe stosujące oprogramowanie dLibra, <http://dlibra.pnsc.pl/>.

⁷⁷ W taki sposób funkcjonuje wspomniany wcześniej Projekt Gutenberg.

⁷⁸ RePEc (Research Papers in Economics) tworzone jest przez ochotników z ponad 60 krajów. Udostępnia prawie 550 tys. różnego rodzaju obiektów, w tym 433 tys. raportów, artykułów, książek, oprogramowania, wykazów publikacji, kontaktów instytucjonalnych dostępnych online. <http://repec.org/>.

⁷⁹ W ramach projektu funkcjonują portale BIOME (biomedycyna), EEVL (technika, informatyka), HUMBUL (nauki humanistyczne), PSIGate (fizyka), SOSIG (nauki społeczne); <http://www.portal.ac.uk/spp/>.

komputerową, na zasadzie prenumeraty indywidualnej lub instytucjonalnej. Zasoby te stanowią część ukrytego Webu i zapewne, ze względów komercyjnych, niesztybko zostaną udostępnione powszechnie, wspólnie z metadanymi bibliotecznymi. Są to:

- Abstraktowe bazy danych, powstałe na bazie wcześniej istniejących indeksów drukowanych i serwisów abstraktowych, w których gromadzono metadane dotyczące literatury naukowej (głównie czasopism). Schemat metadanych zwykle tworzony jest na potrzeby określonej bazy danych; często struktura opisu pochodzi jeszcze z drukowanych poprzedników. Opisy zawierają zwykle rozbudowany opis rzeczowy ze słowami kluczowymi oraz abstraktem, pochodzącymi z czasopisma lub tworzonymi w serwisie. Każda baza danych jest samodzielnym produktem, który ma najlepiej spełniać potrzeby informacyjne określonej grupy użytkowników. Z tego powodu brak jest jednolitego standardu struktury rekordu baz danych.
- Przeglądowe bazy danych, podobne do abstraktowych baz danych. Zawierają jednak obszerniejsze przeglądy literatury znanego autorstwa, zawierające opis i krytyczną ocenę artykułów. Oba rodzaje zasobów metadanych tracą mocno na znaczeniu na rzecz baz pełnotekstowych.
- Pełnotekstowe bazy danych zawartości czasopism, które mogą być częścią aparatu informacyjnego biblioteki hybrydowej lub występować samodzielnie. Funkcjonuje wielu dostawców tego typu baz danych, z których żaden nie oferuje pełnych zasobów. Są to zazwyczaj wielcy wydawcy komercyjni, tacy jak ScienceDirect Elseviera⁸⁰ czy SpringerLink⁸¹ albo konsorcja uczelni lub bibliotek, typu JSTOR⁸². Często metadane indeksowane są przez wyszukiwarki. Metadane stanowią tu pełny opis artykułu, zawierający pewne dodatkowe elementy, na przykład afiliację autorów. Tworzone są także wykazy bibliografii załącznikowych poszczególnych artykułów z aktywnymi odnośnikami do ich opisów lub pełnych tekstów. Dostępne są też indeksy cytowań.

Zasadą funkcjonowania bibliograficznych baz danych w Internecie, jako części serwisów globalnej biblioteki cyfrowej, jest bezpośrednie kierowanie użytkownika od metadanych do pełnych tekstów opisywanych dokumentów. Ich funkcja informacyjna jest więc w tym przypadku do tego stopnia bezpośrednio powiązana z funkcją udostępniania, że użytkownik może nie zdawać sobie sprawy z istnienia metadanych. Samo udostępnianie opisywanych obiektów może natomiast odbywać się na różnych zasadach, w trybie Open Access lub odpłatnie.

Wielkie zasoby metadanych są tworzone i gromadzone przez **wyszukiwarki internetowe**, chociaż często nie kojarzy się ich działalności z taką funkcją⁸³. Te skomplikowane, programistyczne narzędzia wyszukiwawcze gromadzą sporą ilość danych na temat indeksowanych stron (a więc ich metadanych). Sposób działania współczesnych wyszukiwarek opisany został w innej części książki, tutaj zajmę się tylko tymi funkcjami, z realizacją których wiąże się tworzenie, przechowywanie i stosowanie metadanych.

Podstawowymi danymi, którymi musi dysponować wyszukiwarka (a dokładnie jej szperacz), są adresy URL indeksowanych stron. Adresy te przechowywane są na specjalnych serwerach [Brin, Page 1998]. Pierwsze wyszukiwarki ograniczały się do indeksowania tytułów stron Web; obecnie indeksowane są zawartości całych stron Web, a także plików w innych formatach, na przykład PDF czy dokumentów Microsoft Office [Battelle 2006, s. 21].

⁸⁰ Baza zawiera niemal 9 mln artykułów oraz książek: <http://www.sciencedirect.com/>.

⁸¹ Baza zawiera ponad 3,6 mln dokumentów różnego typu: <http://www.springerlink.de/home/main.mpx>.

⁸² Dostęp do ponad 770 tytułów czasopism: <http://www.jstor.org/>.

⁸³ Piotr Gawrysiak na przykład metadane rozumie w sposób znacznie węższy, jedynie jako opis treści dokumentu sporządzony przez człowieka (tj. nie w wyniku działania programu komputerowego). Dostrzega on także pewne inne funkcje metadanych (oprócz opisowych), takie jak dane dotyczące sposobu uzyskania dostępu [Gawrysiak 2008a, s. 653].

Strony Web, przeznaczone do indeksowania, przechowywane są w specjalnym repozytorium, wraz z ograniczoną liczbą opisujących je metadanych, takich jak identyfikator strony, długość strony, długość URL oraz sam URL. Dane znalezione w tych dokumentach odsyłane są do wielkich baz danych, zwanych indeksami. Są one efektem pracy algorytmów analizy tekstu, o których (a raczej o ich twórcach) Piotr Gawrysiak twierdzi, że nie są dużo głupsze od przeciętnej osoby katalogującej materiały biblioteczne [Gawrysiak 2008c, s. 65].

W indeksie znajduje się lista słów, znalezionych na stronie Web, wraz z istotnymi danymi o niej (czyli metadanymi), takimi jak: na jakiej stronie wystąpiło słowo, jego pozycja w dokumencie, rozmiar czcionki, informacja o odnośnikach, tekst wokół i w obrębie odnośnika. Tworzone i przetwarzane są różnego rodzaju dane statystyczne. Specjalną rolę odgrywają metadane o odnośnikach, dzięki którym można określić, skąd i dokąd prowadzi każdy odnośnik oraz jaki tekst mu towarzyszy. Dane te są zorganizowane w taki sposób, aby znając adres URL można było odnaleźć związane z nim słowa.

Następnym krokiem jest odwrócenie bazy danych, czyli stworzenie indeksu odwróconego stron: listy słów z adresami URL, z których pochodzą. Indeks zawiera także inne informacje o słowach, na przykład miejsce w dokumencie, z którego słowo pochodzi. Szczególnie cenione są słowa z tytułu strony i z odnośnika, które podczas wyszukiwania uwzględniane są w pierwszej kolejności [Huuhka 2006]. Indeks można sobie wyobrazić jako ogromną bazę wybranych, istotnych danych o zindeksowanych stronach Web i udostępnianych na nich obiektach. Podstawą funkcjonowania współczesnych firm, tworzących światowe wyszukiwarki, takie jak Google i Yahoo, jest obecnie głęboka analiza tej bazy danych, której efektem jest odkrywanie zależności statystycznych i algorytmizowalnych zjawisk, wykorzystywanych do zapewnienia jak najwyższej trafności odpowiedzi na zapytania użytkownika.

W procesie analizy indeksy zapełniane są znacznikami, które także stanowią pewnego rodzaju metadane. Mogą one na przykład informować o tym, że strona została zapisana w określonym języku lub też należy do określonej kategorii stron, na przykład spamerskich, pornograficznych albo rzadko aktualizowanych. Strony mogą także być oznaczone na przykład jako biograficzne (tzn. oferujące dane biograficzne, nie tylko i niekoniecznie zawierające słowo „biografia”), udostępniające recenzje (na przykład filmowe), notowania giełdowe lub prognozy pogody. Dzięki właśnie tym metadanym wyszukiwarka zwiększa szanse na przedstawienie trafnej odpowiedzi użytkownikowi. Wyszukiwarka tworzy swojego rodzaju klasyfikację treści zasobów Web na podstawie zapytań użytkowników.

Metadane zbierane są i zapisywane także na kolejnych etapach wyszukiwania. Tak zwany serwer zapytań służy m.in. do obserwacji treści i sposobu zadawania pytań przez użytkowników. Na podstawie metadanych o zapytaniach użytkowników powstają statystyczne wzorce zapytań. Za ich pomocą można uwzględniać lokalne różnice w formułowaniu zapytań i próbować zaradzić problemom, związanym ze stosowaniem przez użytkowników wyszukiwarek niczym nie kontrolowanego słownictwa. W tym celu wyszukiwarki obserwują i zapisują tak zwane intencje użytkownika – co jest szukane, które wyniki z przedstawionej listy są wybierane, jakie strony później są odwiedzane.

Odzwierciedleniem liczby i rodzaju gromadzonych metadanych są także możliwości, jakie daje tzw. wyszukiwanie zaawansowane. Większość wyszukiwarek pozwala zawęzić wyszukiwanie według kryterium wybranej frazy, domeny, rodzaju pliku, położenia geograficznego, języka i liczby wyświetlonych adresów. Możliwe jest dodawanie lub wykluczanie słów, określanie zasięgu czasowego a także wyszukiwanie stron, które są podobne do już znalezionych.

Dzięki tego typu narzędziom i procesom statystycznym wyszukiwarka, na podstawie gromadzonych przez siebie metadanych oraz ich modelom i strukturom powiązanym relacjami, które można określić mianem ontologii (zob. p. 2.7), staje się systemem sa-

moucącym i dostosowującym do potrzeb użytkowników – im więcej osób wyszukuje w Internecie, tym wyszukiwanie staje się bardziej efektywne. Pomimo tego, że większość algorytmów stosowanych w wyszukiwarkach internetowych używa głównie metody statystycznej analizy tekstu, jakość ich działania jest wystarczająco wysoka [Gawrysiak 2008b, s. 658]. Metody głębszej analizy semantycznej tekstu, wykorzystywane także przez algorytmy wyszukiwarek, przedstawione zostały w p. 4.2.

3. PRZESTRZENIE

Przestrzenie, zgodnie z zasadami 5S, są to zestawy obiektów, wraz z operacjami na tych obiektach, podlegające określonym ograniczeniom. W odniesieniu do GBC, będącej częścią Internetu, można mówić o bardzo różnych przestrzeniach, począwszy od rzeczywistych, takich jak wizualne reprezentacje danych i ich projekcje, po bardziej abstrakcyjne, jak przestrzenie informacyjne. Mierzenie odległości pomiędzy obiektami w przestrzeni pozwala na określenie ich podobieństwa podczas wyszukiwania, a więc także wyznaczenie efektywności systemu, służącego opisowi tych przestrzeni. Przestrzenie dotyczą lokalizacji fizycznych, sposobu funkcjonowania interfejsów użytkownika, a także lokalizacji obiektów w przestrzeni Sieci.

Współczesny Internet jest rozwinięciem niewielkiej sieci Arpanet, stworzonej na zamówienie Pentagonu¹ w końcu lat sześćdziesiątych XX wieku [Castells 2008, s. 58]. Sieć ta miała być, ze względu na swoją specyficzną topologię, bardziej odporna na uszkodzenia niż wcześniej stosowane rozwiązania. Po dodaniu, w połowie lat siedemdziesiątych ubiegłego wieku, protokołu TCP/IP powstał Internet, który jest „siecią sieci komputerowych”. Zasadniczą architekturą Internetu jest architektura klient-serwer². Taka konfiguracja zwiększa efektywność pracy sieci, gdyż dane ściągane z serwerów przetwarzane są lokalnie przez programy klienckie, więc liczba niezbędnych przebiegów w sieci jest minimalizowana. Każdy serwer przyłączony do Internetu staje się jego częścią. Internet scala w jedną sieć wszystkie przyłączone serwery, obsługujące z kolei sieci lokalne, rozległe i używane przez dostawców usług internetowych. Jest to więc sieć sieci – brak w nim komputera czy systemu centralnego, właściciela i ciała zarządzającego całością.

Od lat dziewięćdziesiątych XX wieku najpopularniejszą usługą Internetu stał się Web. Przestrzenność Webu tworzona jest przez relacje istniejące pomiędzy obiektami go wypełniającymi. Relacje te wyrażane są strukturą hiperlinków. Topologię³ Webu można określić za pomocą opisu struktury funkcjonujących w nim odnośników. Tworzone są one swobodnie przez użytkowników Webu. Odnośniki konstytuują przestrzenie bliskie i odległe, rozciągające się od jednego węzła sieci do następnych. „Odległość” pomiędzy dwiema stronami Web zależy od liczby odnośników potrzebnych do „przejścia” między nimi [Kolb 2009]. Ponieważ powstają one bez żadnego planu i nie wymagają odgórnej akceptacji, również Web rozwija się w sposób niekontrolowany. Samoorganizująca się struktura odnośników w Webie może być uważana za agregację, na poziomie makro, interakcji funkcjonujących na poziomie mikro – za tworzoną we współpracy, zmieniającą

¹ W szczególności chodzi o agencję ARPA (Advanced Research Projects Agency), później przemianowaną na DARPA (Defense ARPA), stworzoną do finansowania innowacyjnych i biznesowo ryzykownych przedsięwzięć technologicznych. Jednym z jej pracowników był Joseph Licklider, który w 1968 roku przedstawił koncepcję sieci komputerów połączonych wydajnymi łączami, co powinno pozwolić na realizację usług wyszukiwania informacji [Licklider 1968].

² Dostawcy usług internetowych udostępniają jeden lub więcej serwerów sieciowych przyłączonych do Internetu. Użytkownicy lokalni przyłączani są do serwerów. Lokalny komputer staje się wtedy komputerem-klientem, który współpracuje z serwerem i używa jego cyfrowych linii transmisyjnych do uzyskania połączenia z resztą Sieci.

³ Termin topologia jest tu użyty w sensie stosowanym w analizach sieci i teorii grafów, jako geometryczny i przestrzenny układ połączeń w sieci.

się, globalną sieć dokumentów, kształtowaną przez wielką liczbę lokalnych twórców odnośników. Z tego punktu widzenia Web jest podobny do złożonych sieci społecznych, nie posiadających architektury w sensie produktu inżynierskiego, lecz samoorganizujących się na zasadzie lokalnych interakcji wielkiej liczby osób i ich grup. Na mikro poziomie twórcy lokalnych odnośników w Webie nie są w stanie zaplanować sposobu, w jaki na poziomie makro ich pojedyncze odnośniki tworzą złożone i dynamiczne struktury, gdyż nie istnieje globalny rejestr pozwalający na stworzenie pełnej mapy Webu. Samoorganizujące się topologie odnośników w Webie są tworzone przez kolektywne, nie inżynierskie i nie intencjonalne agregacje odnośników.

Prostota odnośników w Webie jest ich zaletą, ma jednak również pewne wady. Wskazywał na nie Theodor Nelson, twórca systemu Xanadu⁴, który bardzo krytykował współczesny Web, m.in. za jednokierunkowość⁵ połączeń, ich niestałość i wykorzystywanie do celów komercyjnych, co powoduje, że użytkownicy Webu traktowani są raczej jako bierni konsumenci, niż twórcy treści [Wright 2007, s. 228]. Sam twórca Webu, Tim Berners-Lee, planował nadanie odnośnikom znacznie więcej „inteligencji”; w pierwotnej koncepcji Webu z 1980 r., nazwanej przez niego Enquire [Berners-Lee 2000, s. 10], odnośniki miały wskazywać na relacje między łączonymi obiektami⁶. Później porzucił tę myśl na rzecz uproszczenia struktury hipertekstu, dzięki czemu zapewnił jego szybkie upowszechnienie. Do koncepcji inteligentnych odnośników powrócił jednak wraz z koncepcją Semantycznego Webu (zob. p. 4.9), gdzie zaproponował już niestandardowy zestaw relacji, jak w Enquire, lecz standardowy sposób opisu jakiegokolwiek użytecznej relacji, przydatnej do wskazywania cech obiektów. Zasady tworzenia tych relacji, określanych jako trójki: dwóch terminów łączonych przez nazwę relacji, zawarte są w RDF (zob. p. 2.3). Same relacje i ich znaczenia przedstawiane są w ontologiach.

Jak twierdzi Yochai Benkler, odnośniki nie są tylko technicznymi narzędziami, mechanicznie łączącymi dwie strony Web [Benkler 2008, s. 314]. Web posiada znacznie bardziej skomplikowaną strukturę, niż mogą ukazać najbardziej złożone schematy odnośników. Na te struktury nakładają się bowiem społeczne sieci osób dyskutujących, tworzących dokumenty i porozumiewających się lub nie zgadzających się ze sobą. Każdy węzeł sieci skupia więc grupy uczestników dyskursu. Dzięki temu Internet jest potężnym narzędziem otwartej wymiany opinii, co także wykorzystywane jest w zastosowaniach naukowych.

W opisanym przestrzeni funkcjonuje infrastruktura GBC, będąc jednym z największych źródeł odnośników i tworzonych przez nie struktur: obiektów cyfrowych i użytkowników. Na poziomie poszczególnych serwisów GBC można wyróżnić także przestrzenie udostępniane użytkownikom na ekranie komputera, na którym odwzorowywane są relacje pomiędzy obiektami Webu. Są one częścią interfejsu użytkownika, który opisany został w p. 3.6.

3.1. Internet jako przestrzeń publiczna

Rola Internetu jako przestrzeni publicznej dostępnej dla każdego obywatela (w odróżnieniu od Sieci jako przestrzeni dla wybranych grup, na przykład osób korzystających z niej zawodowo), może być określona przez dwie przeciwstawne cechy: Internet jest

⁴ <http://xanadu.com/>.

⁵ W przypadku wykorzystania proponowanych przez Nelsona odnośników dwukierunkowych autor strony Web może wiedzieć nie tylko do kogo sam stworzył odesłania, ale także skąd prowadzą odnośniki do jego strony.

⁶ Plany dotyczyły takich relacji, jak „utworzyny przez”, „zawiera”, „wykorzystuje”, „opisuje”, „podobny do” [Weinberger 2007, s. 190]. Relacje te miały opisywać znaczenie odnośników, byłyby więc rodzajem metadanych.

zarówno ogólnie dostępny, jak i osobisty. Cyberprzestrzeń⁷, inaczej niż media tradycyjne (w tym publikacje) oraz tradycyjne miejsca publiczne w świecie fizycznym (w tym biblioteki), umożliwia zastosowanie nowych sposobów współdziałania ludzi na poziomie ekonomicznym, politycznym i społecznym. Nie chodzi tu o przestrzeń fizyczną, materialne części maszyn i łączny telekomunikacyjnych, ale o przestrzeń logiczną, wyobrażeniową [Golka 2008, s. 99; Górska 2009, s. 49]. Uniwersalny dostęp do łącz internetowych daje potencjalnie szanse interakcji każdemu, w każdym miejscu. Jednak uniwersalność ta może być powodem ograniczenia praw jednostki czy organizacji, zmniejszając ogólną użyteczność Internetu [Camp, Chien 2000, s. 13].

Spacerując w parku, przesiadając się na dworcu kolejowym czy spędzając spokojnie popołudnie w bibliotece, za każdym razem korzysta się w specyficzny sposób z przestrzeni fizycznej w domenie publicznej⁸, zajmuje się w niej jakieś miejsce i postępuje zgodnie z obowiązującymi zasadami. Internet jako miejsce w domenie publicznej również potrzebuje podobnego zbioru zasad funkcjonowania, powszechnie przestrzeganych, choć niekiedy formalnych, rządzących jego funkcjami informacyjnymi o zasięgu globalnym.

Każda przestrzeń posiada wewnętrzne, fizyczne definicje zarówno tego, co jest dozwolone, dostępne, jak i wykluczone, zamknięte. Wykorzystywanie modeli przestrzennych powoduje potrzebę ponownego rozpatrzenia relacji pomiędzy przestrzeniami. Wszystkie przestrzenie fizyczne i część przestrzeni cyfrowych pozwalają na tworzenie wykluczeń. W Internecie przykładem mogą być nazwy domen, które muszą być niepowtarzalne. Wybrane elementy zastrzeżonych przestrzeni elektronicznych mogą mieć cechy dóbr publicznych. Z drugiej strony przestrzenie te stają się coraz bardziej dozwolone. Mieszają się w nich to, co prywatne i publiczne, treści przepływają w obu kierunkach. Przestrzenie cyfrowe mogą być doświadczane jednocześnie, na przykład w przypadku umieszczenia zasobów jakiejś firmy w intranecie i w przestrzeni publicznej. Zarządzanie przestrzeniami musi uwzględniać fakt, że funkcjonują w kontinuum, zawierającym elementy wyłączności, dozwolenia i jednoczesności w różnych wymiarach i sytuacjach. Różni to przestrzenie elektroniczne od fizycznych, gdyż te ostatnie zawsze znajdują się w przestrzeni wyłącznej. Stąd metafora przestrzeni Internetu, co prawda obiecująca, nie może zakładać ani całkowitego odrzucenia wyłączeń, ani też ich potraktowania jako jedynej dostępnej możliwości.

Warto przyrzeć się kilku cechom Internetu jako przestrzeni:

- Publiczna i prywatna. Inaczej niż zazwyczaj przestrzeń fizyczna, Internet łączący ludzi, urządzenia i zasoby informacji, jest jednocześnie publiczny i prywatny. Jest to być może jedna z bardziej charakterystycznych jego cech – wszyscy pracują i żyją w tej samej przestrzeni, współużytkują te same zasoby, a przy tym każdy może wyodrębnić sobie część tej przestrzeni i traktować ją jak własną. W Sieci wspólne zasoby są wykorzystywane i dostosowywane do własnych potrzeb. Z drugiej strony umieszcza się w Sieci własne, prywatne informacje (na przykład w Webie społecznościowym lub na osobistych stronach uczonych w Web), przedstawiane do użytku publicznego.
- Globalna – lokalna. Przestrzeń Internetu z definicji zawiera połączenia w skali globalnej, z drugiej jednak strony zasoby lokalne wzbogacają ją i zwiększają użyteczność. Teoretycznie również łatwo jest wejść na stronę Muzeum Narodowego w Krakowie, jak i na stronę Muzeum w Luwrze. Treści przeznaczone do użytku wybitnie lokalnego, mieszają

⁷ Marian Golka przytacza następującą definicję cyberprzestrzeni: przestrzeń otwartego komunikowania się za pośrednictwem połączonych komputerów i pamięci informatycznych pracujących na całym świecie [Golka 2008, s. 99].

⁸ W tym miejscu termin „domena publiczna” rozumiany jest bardzo szeroko, jako miejsce dostępne każdemu, do którego brak ograniczeń wynikających z praw własności do elementów przestrzeni fizycznej i cyfrowej. Szczegółowe rozumienie tego terminu w obszarze praw własności intelektualnej przedstawione zostało w p. 6.3.

się z zasobami o znaczeniu globalnym, następuje więc wspomniana już globalizacja. Użytkownicy przestają odróżniać otoczenie bliskie i dalekie na rzecz tych obszarów, które najlepiej korelują z realizowanymi celami. Jednak poza udostępnianiem treści, Internet realizuje także zasady polityki światowej, w tym naukowej, co utrudnia zachowanie równowagi między strukturami narodowymi i globalnymi.

- Transjęczyczna i międzykulturowa. Pomimo, że język angielski jest nadal, a może nawet coraz bardziej uniwersalnym językiem naszych cyfrowych czasów, Internet jest środowiskiem wielokulturowym i wieloetnicznym. Ta cecha wiąże się z pierwszą: w przestrzeni publicznej zazwyczaj używa się języka angielskiego (przynajmniej jako jednego z kilku), w przestrzeni prywatnej własnego języka etnicznego.
- Związki z obszarami niepublicznymi. Przestrzeń Internetu, nawet w domenie publicznej, może być powiązana, zarówno przez przypadek, jak i w sposób zaprojektowany, z przestrzeniami o charakterze prywatnym, takimi jak obszary pracy, obszary handlu. Przykładowo taka przestrzeń prywatna, jak komercyjna strona Web przedsiębiorstwa, może być bardziej efektywna biznesowo, gdy stworzona zostanie jako funkcjonalnie publiczna (na przykład dla wykorzystania faktu, że klienci wchodzący na stronę zostawiają o sobie mnóstwo informacji).
- Kontrolowana i wolna. Dotyczy to głównie sposobu zarządzania przestrzenią. W tym zakresie jeszcze trudniej znaleźć konsensus, gdyż poprzednio przedstawione cechy przestrzeni utrudniają go. Opinia publiczna chce nieograniczonego dostępu do informacji, ale jednocześnie bezpiecznego, chronionego środowiska elektronicznego. Próby „ograniczenia anarchii”, panującej w Internecie stwarzają jednak zagrożenie zmiany proporcji pomiędzy tym, co jest prywatne i zastrzeżone w Internecie, a tym, co publiczne. Właściciel dobra prywatnego zawsze dąży do ograniczenia jego używania przez innych, bez względu na to, czy jest to dobro fizyczne, czy wirtualne.

Jako przestrzeń publiczna Internet znaczy wiele więcej niż tylko moce obliczeniowe i możliwości komunikacyjne. Używany jest on przez społeczności, dla których ważne są wybrane cechy przedstawione powyżej. Jedną z tych społeczności są użytkownicy globalnej biblioteki cyfrowej, która pozwala na tworzenie nowych możliwości i wyznaczanie nowych barier w pojmowaniu Internetu jako miejsca. GBC jako miejsce przedstawiona zostanie w następnym rozdziale.

3.2. Globalna biblioteka cyfrowa jako miejsce

Globalna biblioteka cyfrowa traktowana jako miejsce i przestrzeń może być oceniana z wielu punktów widzenia. Z jednej strony cyfryzacja komunikacji naukowej może pogłębiać izolację pracowników nauki i w rezultacie całej społeczności uczonych, z drugiej sieci komputerowe dają szansę zmiany tej sytuacji dzięki wielkim ułatwieniom w jednoczesnej i natychmiastowej komunikacji, możliwości swobodnego wyrażania opinii i dotarcia ze swoimi poglądami do szerokiego, ale równocześnie bardzo wyselekcjonowanego kręgu odbiorców. Dla rozwiązania tych sprzeczności warto sieci globalne traktować i opisywać nie tylko w kategoriach technicznych, jako narzędzia informatyczne, ale również jako przestrzeń społeczną, a globalną bibliotekę cyfrową jako miejsce w tej przestrzeni, w której zachodzą nowe formy interakcji i nowe relacje pomiędzy jej użytkownikami. W społeczeństwie sieci, w którym technologie pozwalają na błyskawiczny przepływ wielkich zasobów danych i informacji, ludzkie doświadczenie przestrzeni (a także czasu) niemal się zatarło [Barney 2008, s. 40]. Oznacza to, że przymus miejsca nie ogranicza już aktywności związanych z komunikowaniem informacji. Miejsca są ważne o tyle, o ile ułatwiają lub utrudniają, zapośredniczone przez technologię sieci, pozbywanie się ograniczeń, stwarzanych dla komunikacji naukowej przez lokalizację.

GBC jako miejsce posiada cechy heterotopii, opisywanych przez Michela Foucaulta [Foucault 2005, s. 117-125]. Heterotopia jest miejscem rzeczywistym, w odróżnieniu od utopii. Jest to miejsce wewnątrz, ale jednocześnie na zewnątrz społeczeństwa, znajdujące się jakby obok zwykłych miejsc zajmowanych przez społeczności. Takim miejscem są biblioteki, w których artefakty i ich reprezentacje, nawarstwiane w czasie, umieszczane są równolegle w wielu przestrzeniach, w których uzyskują odrębne znaczenia. GBC może być wyrazem nieodległych idei stworzenia *bibliotheca universalis*, miejsca, w którym można umieścić wszystkie epoki, formy, gusta, obejmującego wszystkie okresy, a jednocześnie pozostającego poza czasem i odpornego na zniszczenia. Miejsce to powinno zapewnić nieskończoną akumulację artefaktów wszystkich czasów.

Miejsca w przestrzeni pojawiają się w sposób naturalny lub też są starannie projektowane. Tworzenie miejsc jest złożonym procesem, polegającym na porządkowaniu lub uzdatnianiu odpowiednich przestrzeni i ich obiektów w celu tworzenia środowiska wspomagającego realizację pożądaných działań z uwzględnieniem społecznych i kulturowych ról aktorów i ich grup (społeczności). W oparciu o takie pojęcie miejsca Jeffrey Pomerantz i Gary Marchionini zaproponowali uwzględnienie trzech podstawowych elementów składowych biblioteki jako miejsca:

- Kontinuum fizyczno-konceptualne.
- Ludzie funkcjonujący w miejscu (aktorzy).
- Cechy funkcjonalne miejsca powodujące, że ludzie z niego korzystają [Pomerantz, Marchionini 2007, s. 506].

GBC jako miejsce to kontinuum fizyczno-konceptualne, rozumiane jako cały zakres konkretyzacji przestrzeni: od bieżącej rzeczywistości fizycznej, w której znajdują się nasze ciała, przez przestrzenie fizyczne udające inne przestrzenie fizyczne, do symulacji rzeczywistej przestrzeni fizycznej, wykonywanej przez media takie, jak telewizja i film, po reprezentacje wymyślonych przestrzeni fizycznych oferowanych przez nowe media, w tym rzeczywistość wirtualną, aż po miejsca powstałe (wyobrazone) jedynie w naszych myślach. Aktorami w skali GBC są użytkownicy (autorzy i wyszukujący) i ich społeczności; pozostali: bibliotekarze, administracja, przedstawiciele instytucji nadrzędnych, takich jak władze uczelniane czy lokalne, zarząd instytucji, to aktorzy na poziomie serwisów GBC. Podstawowymi funkcjami serwisu GBC jako miejsca są: selekcja idei materializowanych w udostępnianych obiektach, archiwizacja tych idei oraz tworzenie i wykorzystywanie struktur technicznych oraz organizacyjnych do ich udostępniania. Różnice i podobieństwa pomiędzy funkcjami i społecznościami związanymi z bibliotekami fizycznymi i cyfrowymi (w szczególności globalną biblioteką cyfrową) zachodzą w obrębie kontinuum fizyczno-konceptualnego.

Koncepcja GBC jako miejsca posiada zarówno sens fizyczny, jak i konceptualny, przez co można ją umieścić w obu ekstremalnych punktach kontinuum: dosłownym i konceptualnym. Ludzie mający związek z serwisem GBC jako miejscem nie zmieniają się ze względu na to, w jakim stopniu świadczy on usługi oparte na zasobach cyfrowych, jednak niektórzy mogą być związani tylko z jednym rodzajem usług, świadczonych w określonym miejscu. Ponieważ wiele bibliotek tradycyjnych posiada elementy elektroniczne (choćby OPAC), a serwisy GBC są często powiązane z bibliotekami tradycyjnymi, jest prawdopodobne, że oba w dużym stopniu wykorzystywane są przez tych samych użytkowników. Idee są niezależne od przestrzeni, lecz dla swojej materializacji potrzebują materii lub energii. Materializacje obiektów fizycznych wymagają głównie materii, natomiast obiektów cyfrowych – energii. Ludzie (aktorzy) współdziałają z serwisem GBC jako miejscem w trakcie zdobywania nowych doświadczeń i poszerzania zakresu swojej wiedzy. Według Jeffreya Pomerantza i Garego Marchioniniego, biblioteki tradycyjne i serwisy GBC różnią się od siebie głównie rodzajem (formatem, nośnikiem) udostępnianych materiałów, natomiast idee materializowane w tych zasobach w znacznym stop-

niu się pokrywają. Fizyczna lub cyfrowa forma materializacji silnie wpływa na sposób realizacji funkcji i usług. W przypadku bibliotek tradycyjnych całość komunikacji naukowej odbywała się w jednorodnym środowisku; zarówno aktorzy, jak i materializacje idei należą do świata fizycznego. W serwisie GBC aktorzy pozostali fizyczni (o ile nie mówimy o agentach programowych), natomiast materializacje stały się wirtualne, co ma duże konsekwencje. Miejsce, w którym funkcjonuje GBC, ma cechy obu tych środowisk. Z tego faktu wynikają wszystkie inne różnice między tymi miejscami. Wymienione podobieństwa i różnice pozwalają charakteryzować biblioteki tradycyjne i GBC jako miejsce, a w związku z tym także sposoby współdziałania z nimi ludzi.

W modelu zaprezentowanym przez cytowanych autorów przedstawiona została biblioteka, zarówno tradycyjna, jak i cyfrowa, jako miejsce [Pomerantz, Marchionini 2007, s. 508]. Centrum modelu stanowi przestrzeń zajmowana przez bibliotekę, bez względu na to, czy jest ona fizyczna, czy cyfrowa. Przestrzeń musi istnieć najpierw, zanim możemy mówić o pozostałych elementach tego modelu: dopiero potem mogą być w niej gromadzone i przechowywane materializacje dzieł oraz funkcjonować osoby przychodzące do niej i działające dla różnych celów związanych z wiedzą (tworzenia, rozpowszechniania, wykorzystywania wiedzy). Materializacje, tradycyjne lub cyfrowe, są dostarczane do przestrzeni w wyniku realizacji procesu tworzenia zbioru. W tym wypadku przestrzeń odgrywa rolę magazynu dla materializacji. Struktury ułatwiające dostęp do zmateriałizowanych dzieł stanowią wartość dodaną w modelu. Muszą one trafić do przestrzeni przed użytkownikami, gdyż bez nich nie będzie żadnego powodu do korzystania z tej przestrzeni, chyba wyłącznie jako miejsca spotkań, co niektóre biblioteki również wykorzystują, ale co nie wystarcza dla odróżnienia biblioteki od jakiegokolwiek innego miejsca. Korzystanie przez użytkowników z dzieł umieszczonych w przestrzeni powoduje jej użycie poprzez dotarcie do idei zawartych w dziełach. Przedstawiony model uświadamia, że przestrzeń biblioteki może być fizyczna lub wirtualna, co z kolei wpływa na ważne cechy pozostałych elementów modelu, takich jak format przechowywanych materializacji, medium, poprzez które użytkownicy otrzymują do nich dostęp oraz wymagane umiejętności użytkowników. Nowa wiedza, wynikająca z przebywania w przestrzeni i korzystania z niej powoduje powstawanie nowych idei i dzieł, które również stają się częścią przestrzeni informacyjnej.

Z rozumieniem GBC jako miejsca koresponduje postrzeganie biblioteki cyfrowej w teorii konwersacji [Lankes i in. 2007]. Według niej mniej ważny jest stopień „fizyczności” biblioteki, a więcej uwagi poświęca się funkcjom, które ona realizuje. Najważniejsza jest funkcja ułatwiania konwersacji, która definiowana jest jako akt komunikacji i porozumienia pomiędzy grupą agentów (osób, organizacji, społeczności). Ułatwianie konwersacji polega na udostępnianiu różnorodnej i przetworzonej informacji, utrzymywaniu zapisów wcześniej prowadzonych konwersacji, dokumentowaniu osiągniętych porozumień i wyników dla ułatwienia dalszej wymiany zdań. Biblioteka jest więc miejscem, w którym prowadzona jest konwersacja i które powinno ją ułatwiać. Jak widać, takie pojmowanie roli biblioteki nie musi wiązać jej z jakimś miejscem w sensie fizycznym; wręcz przeciwnie, dzięki nowym technologiom i wyzwoleniu z przywiązań do miejsca fizycznego, możliwości wspomaganie konwersacji niezwykle się powiększyły. Dodatkowo zyskała na jakości, gdyż w środowisku cyfrowym jest znacznie łatwiej o dwustronną, a nawet wielostronną wymianę opinii (na przykład typu bibliotekarz-użytkownik-autor) w odróżnieniu od biblioteki tradycyjnej, gdzie konwersacja odbywała się zazwyczaj w jedną stronę, w kierunku bibliotekarz-użytkownik lub użytkownik-bibliotekarz. Miejscem prowadzenia dialogu służącego przekazywaniu wiedzy specjalistycznej jest obecnie GBC.

Określenie „biblioteka cyfrowa” jest metaforą, w sensie, w jakim zjawisko to opisywali George Lakoff i Mark Johnson. Jest ona bowiem częścią systemu konceptualnego, w warunkach którego myślimy i działamy, a który jest ze swej natury metaforyczny [La-

koff, Johnson 1988, s. 4]. W ten sposób powstają modele mentalne, których badanie pozwala zrozumieć i przewidywać zmiany w postawach i zachowaniach jej użytkowników [Batorowska 2009, s. 277]. O ile bowiem termin biblioteka (tradycyjna), w jednym ze swoich najczęściej używanych znaczeń⁹ oznacza miejsce obejmujące pewną fizyczną przestrzeń, o tyle dla biblioteki cyfrowej trudno ją tak jednoznacznie wyznaczyć, a dla globalnej biblioteki cyfrowej taką przestrzeń powinien być cały świat. Wyznacznikiem biblioteki cyfrowej nie jest przy tym, jak się często sądzi, brak fizyczności, gdyż urządzenia pamięci masowych i łączący telekomunikacyjnych¹⁰ są zupełnie fizyczne, podobnie jak miejsca, w których użytkownicy korzystają z tej biblioteki i sami użytkownicy. Różnica polega na organizacji przestrzeni – dla biblioteki tradycyjnej jej przestrzeń jest ograniczona i ściśle wyznaczona, choć w ograniczony sposób bywa rozciągliwa, natomiast dla GBC przestrzeń sięga wszystkich miejsc, do których dociera Internet, jest prawie nieograniczona i stale się rozszerza, a więc jest bardzo rozciągliwa. Użytkownicy globalnej biblioteki cyfrowej mogą stosować w odniesieniu do niej pojęcia miejsca raczej wyłącznie w sensie metaforycznym, gdyż szczególną zaletą GBC jest oderwanie się (z punktu widzenia użytkownika), od ograniczeń związanych z jakimkolwiek miejscem. Podobnie używa się pojęcia „przestrzeń informacyjna” mając w tym przypadku na myśli raczej zasoby informacyjne określonego rodzaju, użyteczne w realizacji potrzeb informacyjnych, niż zasoby związane z jakimkolwiek miejscem fizycznym. Jest to przestrzeń, będąca komfortowym miejscem do pracy, głównie umysłowej i twórczej, nawiązywania kontaktów i współpracy, oferująca wiele udogodnień w stosunku do fizycznej przestrzeni biblioteki tradycyjnej.

Wspomniani George Lakoff i Mark Johnson podzielili metafory na dwa rodzaje: metafory strukturalne, w których jedno pojęcie jest przedstawiane w języku za pomocą innego pojęcia (na przykład dyskusja w terminologii wojennej – potyczki słowne) oraz metafory orientacyjne, w których pojęcie przedstawione jest przestrzennie. Biblioteka cyfrowa jako miejsce w pewnym zakresie jest metaforą orientacyjną, gdyż odnosi się do biblioteki (tradycyjnej) jako budynku, czyli miejsca fizycznego. Z drugiej strony metafora ta opisuje także pewne zjawiska społeczne, zachodzące w bibliotekach jako miejscach spotkań i współdziałania ludzi. Również w tym sensie globalna biblioteka cyfrowa ma cechy metafory miejsca, choćby poprzez funkcjonowanie informatycznej infrastruktury nauki i laboratoriów.

Cechy serwisu GBC jako miejsca można przedstawić następująco:

- Miejsce dla zasobów cyfrowych. Jak już wspomniałem, jedną z podstawowych różnic pomiędzy biblioteką tradycyjną i serwisem GBC jest zapotrzebowanie na przestrzeń fizyczną. Nie ulega wątpliwości, że obiekty, stanowiące materializację idei w formie elektronicznej, wymagają znacznie mniej miejsca, przy czym nie ma sensu gromadzić tych materiałów w licznych lokalizacjach¹¹. Dla zapewnienia dostępu globalnego do

⁹ Należy zaznaczyć, że istnieją także inne znaczenia terminu biblioteka, na przykład biblioteka jako zbiór dokumentów (kolekcja) jest o wiele bliższa idei biblioteki cyfrowej, gdyż jest ona znacznie bardziej „rozciągliwa” w przestrzeni – książka wypożyczona i wywieziona poza bibliotekę, rozumiana jako księgozbiór nadal pozostaje częścią takiej biblioteki, a więc przestrzeń tej biblioteki powinna obejmować całą przestrzeń zajmowaną w danej chwili przez jej zbiory.

¹⁰ Michael Batty zwraca uwagę na dwa podstawowe pojęcia przydatne w opisie przestrzeni wszelkiej komunikacji wspomaganej komputerowo: są to węzły i sieć, czyli urządzenia służące do przetwarzania, magazynowania i prezentacji danych i informacji (komputery, dyski) oraz infrastruktura służąca do ich przesyłania (łącza, bramki). Zarówno węzły, jak i sieć funkcjonują jako przestrzeń i miejsce. Dla węzłów przestrzeń to abstrakcyjny konstrukt funkcjonujący w obrębie danej maszyny cyfrowej dzięki zainstalowanemu oprogramowaniu, a miejsce to punkt przyłączenia, gdzie dokonuje się przejście od przestrzeni lokalnej do systemu komunikowania (miejsce przechodzi w przestrzeń). Dla sieci miejsce to obiektywnie istniejąca infrastruktura, związana z miejscami rzeczywistymi, a przestrzeń funkcjonuje w matrycach komputerowych i rozwija się z nich poprzez ich użycie do komunikowania się z inną, odległą fizycznie przestrzenią komputerową [Batty 1997, s. 339].

¹¹ Pomijam tu problem tzw. mirrorów, które zresztą, wraz ze wzrostem przepustowości sieci, powoli zanikają. W takim przypadku można mówić o istnieniu obiektów cyfrowych w kilku miejscach fizycznych.

danego zasobu wystarczy jego opublikowanie w jednym, dowolnym miejscu Sieci. Dowolność tego miejsca oznacza, że dla użytkownika obojętne jest, czy wykorzystywane zasoby znajdują się za ścianą, czy na przeciwległej półkuli Ziemi. Z tego punktu widzenia czasem mówi się o „wirtualnym miejscu” zajmowanym przez te materiały.

- Organizacja informacji (wiedzy) – zob. p. 2.7. W GBC sposób organizacji wiedzy jest bardzo zróżnicowany pod względem narzędzi stosowanych w poszczególnych serwisach. Wraz ze zmianą sposobu organizacji zasobów serwisu GBC również interfejs powinien ulegać zmianie, gdyż przestrzeń cyfrowa i schemat organizacji wiedzy stanowią jedność; organizacja informacji w pewnym stopniu definiuje przestrzeń cyfrową. W przypadku zmiany w obrębie SOW, zmiana zachodzi także w sposobie postępowania użytkownika w tej przestrzeni. Każdy obiekt cyfrowy może zostać przyporządkowany do wielu miejsc w schemacie. Elastyczność przestrzeni cyfrowej, wynikająca z możliwości stosowania wielu SOW, pozwala na lepszą reprezentację idei materializowanych w obiektach cyfrowych. Stąd zdecydowanie większa liczba tzw. punktów dostępu do materiałów w formie cyfrowej.
- Tworzenie zasobu i archiwizacja. Ta cecha dotyczy sposobu podejmowania decyzji o rodzajach obiektów, umieszczanych w przestrzeni oraz ich przechowywaniu w długim okresie lub usuwaniu. Pomimo istnienia serwisów globalnej biblioteki cyfrowej, stosujących politykę gromadzenia zbiorów podobną do zasad wprowadzanych w bibliotekach tradycyjnych, częstym zarzutem w stosunku do zasobów Internetu jest brak jakichkolwiek zasad w tym zakresie. Sytuacja globalnej biblioteki cyfrowej przypomina miejsce, w którym wszystkie polityki gromadzenia zbiorów (z natury zawierające sprzeczne rozwiązania) wszystkich bibliotek zaczęłyby obowiązywać jednocześnie. W takim przypadku pozostaje gromadzić wszystko, co może zaspokoić jakąkolwiek potrzebę kogokolwiek. Zasadniczą rolę odgrywa oddzielenie zasobów naukowych od nienaukowych; to zadanie ma być realizowane przez nowe narzędzia typu Google Scholar. Nawet jednak przed ich udostępnieniem w miarę łatwo dotrzeć do materiałów reprezentujących odpowiedni poziom (naukowy, fachowy) przez wybór serwisu i właściwe sformułowanie zapytań.
- Przestrzeń osobista. GBC porównywana bywa do bibliotek osobistych, prywatnych, należących do pojedynczych osób i przez nie używanych. Przestrzeń zajmowana przez taką bibliotekę nazywana jest przestrzenią osobistą; nie jest ona dzielona z innymi osobami. Jej ideę można wywodzić aż z Memexu Busha. Według Christine Borgman pozwala ona na uelastycznienie i dostosowanie usług biblioteki do indywidualnych potrzeb użytkownika [Borgman 2003b]. Szczególnie globalna biblioteka cyfrowa dostosowana jest do obsługi bardzo zróżnicowanej grupy użytkowników i ich potrzeb. Obecnie wiele aplikacji, funkcjonujących w GBC, umożliwia tworzenie własnego zasobu na podstawie indywidualnych zainteresowań i potrzeb, stanowiącego część powszechnego zasobu globalnego. Zasób ten powstaje z obiektów sieciowych oraz własnych, dodawanych przez użytkownika. Przykładowo część systemów bibliotecznych oraz księgami internetowych pozwala na kopiowanie i gromadzenia własnych zasobów metadanych. Pliki użytkownika gromadzone lokalnie mogą być wyszukiwane przy pomocy globalnych wyszukiwarek. Systemy społecznościowe, funkcjonujące w ramach Webu 2.0 poszły dalej: tam nie tylko gromadzi się własne metadane, słowa kluczowe i same obiekty cyfrowe, ale także współdzielą się je swobodnie z innymi uczestnikami sieci. W taki sposób przestrzeń, nie przestając być osobistą, staje się także wspólną przestrzenią społeczną.

Metafora miejsca w odniesieniu do biblioteki cyfrowej była krytykowana, szczególnie w latach dziewięćdziesiątych XX wieku. Mark Ackerman twierdził, że metafora biblioteki cyfrowej jako miejsca jest nawet szkodliwa, gdyż pomija rolę społeczną biblioteki. uwypuklając tylko jej zadania gromadzenia i wyszukiwania informacji. Tymczasem bi-

biблиотеka jest miejscem spotkań i socjalizacji użytkowników [Ackerman 1994]. Można jednak zwrócić uwagę, że użytkownicy Internetu, również „spotykają się” (co prawda metaforycznie) i współpracują w sieci, nawet oglądając swoje twarze (brak tylko kontaktu fizycznego dotyku). Mogą się oni widzieć, słyszeć i wymieniać dokumentami, jakby siedzieli przy jednym biurku.

Metafora miejsca jest ważna ze względu na konceptualizację przestrzeni służącej wspólnemu przebywaniu i wykonywaniu czynności. Funkcjonowanie bibliotek naukowych wciąż opiera się na założeniu o zasadniczym znaczeniu ich lokalizacji fizycznej, a w rzeczywistości rola ich umiejscowienia w obrębie kampusu stale maleje [Ross, Senney 2008, s. 146]. Globalna biblioteka cyfrowa oferuje nowe rodzaje funkcjonalności, obsługuje znacznie szersze grupy użytkowników, i tworzy nowe znaczenie miejsca, które utożsamiane jest z interfejsem systemu¹². Pomimo, że miejsce w takim sensie nie może spełniać wszystkich cech miejsca fizycznego, GBC dostarcza jednak nowe, istotne możliwości, dotyczące głównie zwiększonej dostępności dzieł i idei w nich zawartych. Jak piszą Jeffrey Pomerantz i Gary Marchionini, biblioteki są zasadniczo przestrzenne, przy czym pojęcie „przestrzeni” należy definiować szeroko, gdyż najważniejszym elementem przestrzeni serwisu GBC nie jest to, czy jest on fizyczny czy wirtualny, ale jak bardzo jest intelektualny [Pomerantz, Marchionini 2007, s. 527].

3.3. Kryteria i miary oceny

Nauka osiągnęła obecnie postać wielkiego, globalnego i kosztownego systemu; jednocześnie wzrosły wymagania instytucji, finansujących badania naukowe, co do jego efektywności. Fundusze przekazywane są do tych ośrodków badawczych, które mogą pochwalić się lepszym, bardziej efektywnym funkcjonowaniem. Niezbędne stało się stworzenie i zastosowanie kryteriów oraz miar¹³ jako narzędzi dla oceny instytucji i uczonych, ale także służących do intelektualnego zróżnicowania i integracji społecznej programów badawczych w skali globalnej. Miary stosowane w zakresie GBC dotyczą zazwyczaj dwóch elementów systemu komunikacji naukowej, pozwalając na ocenę jakości publikacji naukowych, oraz jakości funkcjonowania GBC i jej serwisów (por. p. 4.8). O ile dla publikowania elektronicznego w Internecie istnieje możliwość zastosowania wielu miar jakości, także nowych, typowych dla środowiska cyfrowego i sieciowego, o tyle trudno mówić o efektywności GBC jako całości; badana jest raczej efektywność jej poszczególnych serwisów.

Miary jakości publikacji przyczyniają się nie tylko do zwiększenia poziomu dostępu, użyteczności i cytowań, ale także wykorzystywane są do umiejscawiania i indeksowania informacji [Armbruster 2008a, s. 10]. W zakresie badań efektywności komunikacji naukowej w Sieci wielu narzędzi dostarcza webometria. Różnorodność miar umożliwiła uczonym obserwację postępów własnych i innych z różnych punktów widzenia¹⁴. Poprzez klasyfikację i zliczanie hiperlinków określane są sposoby zachowań użytkowników. We-

¹² Cecha ta nie jest jedynie spotykana w bibliotekach, czy szerzej systemach elektronicznych: jest powszechnie znaną rzeczą, że użytkownicy bibliotek tradycyjnych oceniają ich pracę po sposobie funkcjonowania katalogu (OPAC) oraz oddziały udostępniania, który stanowi rodzaj „interfejsu” biblioteki. Dobrą opinię o bibliotece doskonale realizującej funkcje gromadzenia czy opracowania zbiorów zepsuć może nie tylko niekompetentny dyżurny w czytelni, ale nawet opryskliwa szatniarka.

¹³ Tefko Saracevic zwraca uwagę na potrzebę odróżniania kryteriów oceny (np. relewancja) od miar (np. trafność, kompletność) [Saracevic 2004].

¹⁴ Na stronie <http://www.webometrics.info/> znajduje się Webometric Ranking of World Universities, zawierający dane o globalnej i regionalnej sytuacji uczelni wyższych oraz wybranych największych instytucjach nauki, takich jak National Institute of Health czy Max Planck Society.

bometria pozwala na zestawianie wielu rankingów, ale także uświadamia, że dzięki tworzeniu zasobów informacyjnych, instytucje nauki uzyskują korzyści z powodu wzrostu ich „widzialności”. Dostępne są różnego rodzaju kryteria, które można podzielić na dwie ogólne grupy: cytowań (*citation*) i wykorzystania (*usage*).

Podstawowym kryterium służącym ocenie publikacji jest ich cytowalność (*citedness*) – liczba lub częstotliwość występowania cytowań, których celem jest określona osoba (autor). Wskaźnik cytowań (czyli znormalizowana lub średnia cytowalność) jest obliczany jako iloraz wszystkich cytowań do liczby lat, w których obserwowana jest cytowalność lub do liczby lat, w trakcie których cytowania były możliwe (na przykład poczynając od daty publikacji cytowanego dokumentu). Badania wskazują na dodatnią korelację pomiędzy liczbą cytowań, a ocenami recenzentów oraz liczbą przyznanych prestiżowych nagród, na przykład Nagrody Nobla [Borgman, Turner 2002, s. 10].

Istnieje wiele miar cytowań, z których najczęściej stosowane są Impact Factor (IF) i inne miary tworzone na podstawie baz danych ISI/Thomson oraz nowe wskaźniki, tzw. h-index¹⁵ i g-index¹⁶. Służą one zarówno do oceny publikacji, jak i uczonych w myśl zasady: im więcej cytowań danej publikacji lub danego autora, tym lepiej. Impact Factor traktowany jest zwykle jako jeden z kilku wskaźników ilościowych służących rankingowaniu, ocenie, kategoryzowaniu i porównywaniu czasopism. Jest to miara częstotliwości, z jaką „średni artykuł” w czasopiśmie był cytowany w określonym roku (lub innym okresie). Jest to więc stosunek artykułów cytowanych do artykułów opublikowanych (czyli możliwych do zacytowania) [Garfield 1994]. IF jest miarą liczby cytowań, otrzymanych w danym roku przez „przeciętny” artykuł publikowany w dwóch poprzednich latach w danym czasopiśmie.

Odzwierciedlaczem IF dla stron Web jest tak zwany Web Impact Factor (WIF). Ma on odzwierciedlać miarę zainteresowania daną stroną Web lub obecności poszczególnych narodów w Web [Nowak 2006, s. 149]. Jest on stosunkiem liczby stron z odnośnikami do danego dokumentu w sieci lub kraju do liczby stron Web tego dokumentu lub kraju, a więc jest wskaźnikiem znacznie mniej obiektywnym od IF. Zależy bowiem od liczby stron danego dokumentu Web.

Podobne kryteria oceny wyszukiwanych stron internetowych stosują także współczesne wyszukiwarki globalne z Google na czele. Stosowany przez nich PageRank (zob. p. 4.3) bierze bowiem pod uwagę ilość i jakość specyficznych cytowań – odnośników do badanej strony Web. Idę tę wykorzystują także twórcy innych kryteriów oceny jakości (zob. dalej).

Zastosowanie Webu 2.0 w nauce powoduje powstawanie propozycji nowych wskaźników, uwzględniających aktywność uczonych na forach społecznościowych. Taką propozycją jest Scholar Factor (SF), który można przedstawić, jako:

$SF = (h\text{-index}) + (\text{współczynnik grantów} / \text{recenzji} / 20) + (\text{współczynnik adnotacji} / \text{oprogramowania} / \text{zbiorów danych} / 5) + (\text{współczynnik Web} / 50)$.

SF zawiera więc h-index (obliczany na podstawie danych z Google Scholar), współ-

¹⁵ Jest on definiowany następująco: uczoney posiada indeks h, jeżeli h z jego N_p artykułów posiada przynajmniej h cytowań, a inne artykuły ($N_p - h$) mają nie więcej niż h cytowań każdy, przy czym N_p oznacza liczbę artykułów opublikowanych w ciągu n lat [Hirsch 2005, s. 16569]. Tak więc uczoney ma na przykład h-index równy 10, jeżeli opublikował 10 artykułów, z których każdy otrzymał co najmniej 10 cytowań. Według obliczeń twórcy wskaźnika, po 20 latach pracy dobry uczoney będzie posiadał h-index równy 20, wyróżniający się uczoney około 40, a do wartości 60 dochodzą tylko wybitne jednostki. Dzięki temu wskaźnikowi można odróżnić uczonych, którzy dokonali jednego odkrycia i stąd mają niewiele artykułów wysoko cytowanych lub też wiele artykułów nisko cytowanych od uczonych, którzy regularnie uzyskują znaczące wyniki.

¹⁶ Jest on określony następująco: dla danego zbioru artykułów uszeregowanych według malejącej liczby cytowań, g-index jest to (unikalna) największa liczba, taka że g artykułów o największej liczbie cytowań otrzymuje (łącznie) przynajmniej g^2 cytowań. Wskaźnik ten, choć jeszcze nie wypróbowany w praktyce wydaje się ważnym uzupełnieniem h-indexu.

czynnik grantów/recenzji (oparty na danych z instytucji finansujących i czasopism), współczynnik adnotacji/oprogramowania/danych (oparty na ilościowych danych z otwartych archiwów) oraz współczynnik Web wyrażony ilościowymi danymi dotyczącymi wpisów na blogach, Wiki i innych serwisach [Bourne, Fink 2008]. Widać więc, że oszacowanie udziału uczonego w Nauce 2.0 wymaga sporego wysiłku.

Oprócz wymienionych wskaźników wyliczyć można także inne, oparte na cytowaniach i stosowane w badaniach bibliometrycznych i webometrycznych, takie jak:

- wskaźnik szybkości oddziaływania (*immediacy index*);
- wskaźnik starzenia się czasopisma (*cited half-life*) [Nahotko 2007b, s. 118];
- relacyjny wskaźnik cytowań (Relative Citation Rate – RCR) [Borgman, Furner 2002, s. 12]. Jest on przykładem wskaźnika, za pomocą którego porównać można liczby zaobserwowanych cytowań z oszacowanymi wartościami cytowań „oczekiwanych”;
- wskaźnik odbicia (*echo factor*);
- wskaźnik popularności (*popularity factor*);
- waga wpływu (*influence weight*) [Nowak 2006, s. 112-115];
- wskaźnik znaczenia czasopisma (*Eigenfactor*¹⁷), stosujący algorytmy podobne do Page-Rank Google’a do mierzenia wpływu czasopisma na podstawie częstotliwości cytowań w innych, wpływowych czasopiśmiech.

Miary wykorzystania obliczane są na podstawie liczby takich zdarzeń, jak wejścia do systemu (logowania), otwarcia plików (metadanych, abstraktu, pełnego tekstu) do odczytu na ekranie, drukowanie. Analiza tzw. logów transakcji pozwala wyciągać wnioski na temat wielu aspektów pracy użytkowników, takich jak: częstotliwość wyszukiwania informacji; różnicowanie potrzeb z punktu widzenia różnych kryteriów (wiek materiałów, liczba wykorzystywanych czasopism, wskaźnik powrotów do tych samych źródeł); „płaskość” wyszukiwania (jak pobieżne jest przeglądanie dokumentów); stopień wykorzystania ze względu na długość materiału; wpływ swobodnego dostępu na wykorzystanie; typowe potrzeby użytkowników wyszukiwarek. Analizy takie wspierane mogą być badaniem zawartości osobistych zasobów, tzw. zakładek (*bookmark*), znanych z serwisu del.icio.us¹⁸, a w nauce tworzonych za pomocą narzędzi typu Zotero¹⁹, CiteULike²⁰, Connotea²¹ lub Mendeley²².

Za rodzaj miary wykorzystania uznać można także liczbę i treść komentarzy użytkowników, od prostego punktowania (z zastosowaniem takich narzędzi, jak Digg) po obszerniejsze recenzje. Mechanizm komentarzy, z powodzeniem stosowany choćby w środowisku programistów, w nauce działa znacznie słabiej. Spośród różnych przyczyn wymienić można jedną, związaną z innymi: zwykle grupy zainteresowanych osób są zbyt małe. Obowiązuje tu tzw. zasada 90-9-1²³: 90% zainteresowanych jedynie obserwuje, 9% bierze skromny udział, a dopiero 1% tak naprawdę jest autorami regularnych komentarzy. Oznacza to, że jeżeli sto osób przeczyta artykuł, to przy odrobinie szczęścia można liczyć na jeden komentarz [Neylon, Wu 2009].

W odróżnieniu od miar opartych na cytowaniach, miary wykorzystania są automatycznie zliczane przez ogólnie dostępne oprogramowanie, stosowane przez serwisy GBC. Wyniki uzyskuje się natychmiast (IF dostępny jest najwcześniej po roku od opublikowa-

¹⁷ <http://www.eigenfactor.org/>.

¹⁸ Serwis del.icio.us pozwala na tworzenie zakładek w wyszukiwarce ułatwiających szybki powrót do strony opisanej w zakładce, a więc jest to mechanizm znany z Ulubionych, jednak różnica polega na tym, że zakładki przechowywane są w Internecie wraz z informacją (metadany, np. jednowyrazowymi etykietkami) o sposobach opisu danej strony przez innych internautów i popularności strony.

¹⁹ <http://www.zotero.org/>.

²⁰ <http://www.citeulike.org/>.

²¹ <http://www.connotea.org/>.

²² <http://www.mendeley.com/>.

²³ <http://www.90-9-1.com/>.

nia artykułu), z możliwością tworzenia wielu zestawień statystycznych. Czasem wykazy wartości tych miar traktowane są jak listy bestsellerów. Miary te traktowane są często jako uzupełnienie miar opartych na cytowaniach. Niektórzy autorzy uważają, że na ich podstawie można wnioskować o późniejszych cytowaniach [Brody, Harnad, Carr 2006]. Problemy stwarza potrzeba ujednoczenia wyników tak, aby można było je porównywać dla czasopism o różnych objętościach, rodzajach i w odmiennych dyscyplinach. Powstaje także pytanie, czy skopiowanie artykułu przez studenta jest równie ważne, jak skopiowanie przez profesora.

Twórcy inicjatywy o nazwie Index Copernicus²⁴ wyszli z założenia, że żadna z miar pojedynczo nie odzwierciedla jakości publikacji. Stworzyli oni wieloparametrowy system oceny czasopism naukowych, biorąc pod uwagę pięć grup parametrów o różnych wagach: jakość naukowa (58% punktów do uzyskania), jakość wydawnicza (20%), zasięg międzynarodowy (13,5%), częstotliwość, regularność, stabilność rynkowa (5%) oraz jakość techniczna (3,5%) [Dankiewicz 2008]. Dopiero uwzględnienie wielu parametrów ze wszystkich grup umożliwiła obiektywną ocenę czasopisma oraz stwarza możliwość ich porównywania.

Miary jakości publikacji są szczególnie przydatne przy założeniu, że mogą objąć jak największą liczbę publikacji; najlepsza jest sytuacja, gdy wszystkie artykuły naukowe udostępniane są w GBC. Wówczas możliwości badań naukometrycznych znacznie przekraczają ocenę jakości publikacji, umożliwiając na przykład badanie kierunków rozwoju nauki, powiązań pomiędzy ośrodkami naukowymi oraz znaczenie poszczególnych ośrodków badawczych i uczonych.

Dla tak obszernego i rozległego systemu, jakim jest globalna biblioteka cyfrowa, niezbędne jest stworzenie metodologii, pozwalającej na standardową **ocenę i pomiar funkcjonowania GBC** jako całości, a także jej poszczególnych serwisów, informacji, którą zawierają i usług przez nie świadczonych. Ocena może być jednorazowa lub prowadzona długoterwale, pozwalając na stałe uściślanie i aktualizację celów, warunków i rodzaju usług. Ta druga realizowana jest często na początku wprowadzania określonej usługi biblioteki cyfrowej, czy jej podsystemu, służąc weryfikacji przyjętych założeń. Możliwe jest zbadanie reakcji użytkowników na udostępnienie określonych usług jako części systemu. Ocena długoterwała pozwala poznać potrzeby użytkowników oraz sformułować poprawne miary na początku tworzenia nowych usług.

Ocena jednorazowa prowadzona jest na końcu procesu tworzenia usług. Dzięki niej można się dowiedzieć, czy cele i warunki założone podczas projektowania zostały osiągnięte. Cele serwisów GBC mogą tworzyć wielopoziomowe struktury hierarchiczne, czasem nawet ze sobą nawzajem sprzeczne. Ocena poprzez cele nigdy nie jest ostateczna i niezmienna. Ten sam serwis może być oceniany na różnych poziomach, w związku z wyborem różnych celów, za pomocą różnych metod i dla różnych odbiorców.

Niektórzy autorzy rozróżniają także ocenę systemów wyszukiwania danych i informacji [Baeza-Yates, Ribeiro-Neto 1999, s. 73]. Dla oceny dowolnego, tradycyjnego systemu wyszukiwania danych, najprostszymi kryteriami jest czas i przestrzeń. Im krótszy jest czas odpowiedzi, im mniej pamięci dyskowej jest wykorzystywane, tym lepszy jest oceniany system. W przypadku tych systemów należy zwracać uwagę na efektywność funkcjonowania struktur indeksowych (tworzonych w celu przyspieszenia wyszukiwania), współpracy z systemem operacyjnym, opóźnienia spowodowane przez kanały komunikacyjne (łącza) oraz przez kolejne warstwy oprogramowania, zazwyczaj składające się na system. Taki sposób oceny nazywany jest oceną funkcjonalną (*performance evaluation*). Ta sama ocena dla wyszukiwania w Webie jest bardziej skomplikowana, zakłada bowiem uwzględnienie dwóch czynników: 1) czasu, niezbędnego do odszukania relewantnej²⁵ jed-

²⁴ <http://www.indexcopernicus.com/>.

²⁵ Relewantna to ocena wartości informacji. Jest to relacja, której argumenty umieszczone są rozmaicie: a)

nostki danych, oraz 2) liczby węzłów sieci, niezbędnych do przejścia w celu odszukania relewantnej jednostki danych [Palmquist, Kim 2000, s. 558].

Dla systemów wyszukiwania informacji, poza oceną czasu i miejsca, stosowane są także inne kryteria i miary. Ze względu na to, że zapytanie użytkownika jest ze swej natury nieściśle, dokumenty stanowiące wynik wyszukiwania w różnym stopniu odpowiadają jego potrzebom i muszą być grupowane według ich relewancji do zapytania. Ranking na podstawie relewancji wprowadza do wyszukiwania element, który nie jest obecny w systemach wyszukiwania danych, a który odgrywa ważną rolę w wyszukiwaniu informacji. Systemy wyszukiwania informacji wymagają więc określenia trafności (*precision*), z jaką dobierany jest zbiór wyników. Taki sposób oceny nazywany jest oceną funkcjonalną wyszukiwania (*retrieval performance evaluation*).

Dla oceny funkcjonowania komunikacji naukowej w środowisku cyfrowym, w tym roli serwisów GBC, przydatne jest ustalenie zestawu wskaźników²⁶, które są podstawą standardowej i poddającej się zarządzaniu metody pomiaru poziomu działalności serwisu, pozwalają także na dokonywanie porównań pomiędzy jego elementami. Powszechnie znane, klasyczne miary efektywności systemów wyszukiwania informacji – trafności i kompletności – opierają się na ocenie relewancji technicznej, w małym stopniu dotyczą więc poziomu zadowolenia użytkownika (pertynencji). O ile jednak miary te mogą być stosowane do poszczególnych serwisów GBC, to ocena efektywności wyszukiwarek internetowych i GBC jako całości przy ich pomocy napotyka na wiele trudności, a nawet negowany jest sens takiej oceny [Bojar 2009, s. 19]. W systemach tradycyjnych (przedinternetowych) zakładana była potrzeba uzyskiwania wysokiej kompletności. W większości współczesnych systemów interaktywnych użytkownicy oczekują wyszukiwania tylko kilku najbardziej relewantnych dokumentów, więc zależy im raczej na uzyskaniu wysokiej trafności. Dodatkowo badanie kompletności wyszukiwania, realizowanego przez wyszukiwarki internetowe, jest praktycznie niemożliwe, gdyż do tego celu niezbędna jest znajomość liczby wszystkich dokumentów relewantnych, tej zaś nie można uzyskać. Istniejące metody pozwalają jedynie na zgrubne szacowanie tej liczby [Choroś 2002]. Dużo łatwiejsza do wyliczenia jest trafność wyszukiwania. Ta jednak istotna jest dla badania efektywności wyszukiwania w małych zbiorach, chociaż również tam nie dotyczy pertynencji.

W zastosowaniach internetowych twórcy wyszukiwarek typu ogólnego nastawiają się na podwyższanie trafności (chodzi zwłaszcza o trafność wyników wyświetlanych na pierwszym ekranie), dążąc do wydobycia obiektów możliwie wysoko relewantnych z ogromnej masy (często dziesiątków milionów stron Web) pozycji odszukanych. Stąd też wyszukiwarki automatycznie i standardowo łączą wprowadzane terminy wyszukiwawcze funktorem AND, którego użycie poprawia trafność kosztem kompletności wyszukiwania. Zdarza się także rozróżnianie trafności wyników wyszukiwania, od trafności opisów, wyświetlanych przez wyszukiwarki jako wyniki; trafność tych ostatnich jest zwykle wyższa [Lewandowski 2008, s. 929].

W każdym warunkach słabą stroną obu miar jest ich uzależnienie od relewancji; działają one tylko w przypadku, gdy relewancja, jako prawdopodobieństwo zbieżności pomiędzy relewancją systemową a relewancją użytkownika może być jednoznacznie określona. Nie zawsze się to udaje [Saracevic 2007, s. 1918]. Relewancja pragmatyczna, użytkowni-

w sferze relacji zachodzących między zbiorami wyszukiwawczymi SIW i tekstami JIW w planie wyrażania (relewancja techniczna) i/lub w planie treści (relewancja semantyczna); b) w sferze relacji między zbiorami informacyjnymi SIW i elementami jego otoczenia, np. potrzebą informacyjną użytkownika, jego wiedzą oraz celami działania (relewancja pragmatyczna) [Bojar oprac. 2002, s. 229].

²⁶ Wskaźnik funkcjonowania serwisu jest zazwyczaj definiowany jako określona wartość lub charakterystyka służąca do pomiarów wyników lub rezultatów. Są to miary jakości usług, efektywności działania i satysfakcji użytkowników. Wskaźniki mogą być ilościowe (np. czas przetwarzania) lub jakościowe (np. satysfakcja użytkownika).

ka, zwana też pertynencją, jest pojęciem bardzo subiektywnym i intuicyjnym, zależnym wyłącznie od indywidualnej oceny ludzkiej.

Ze względu na trudności z określeniem trafności i kompletności wyszukiwania w Web stosowane są inne miary. Dirk Lewandowski wymienia takie miary efektywności wyszukiwawczej, jak wskaźnik czisty wyszukiwawczej, zdolność do wyszukiwania stron umieszczonych najwyższej w rankingu oraz średniej odległości [Lewandowski 2008, s. 934]. Autor podkreśla także, że globalne wyszukiwarki, takie jak Google i Yahoo, rzadko w odpowiedzi na zapytanie nie przedstawiają żadnych wyników. Kolejnym jest wskaźnik Content-Bearing Clic (CBC). Jest on definiowany jako stosunek kliknięć na odnośnik hipertekstowy, zastosowanych do wyszukiwania prawdopodobnie relewantnej informacji, do kliknięć na odnośnik hipertekstowy w innych celach.

W przypadku serwisów GBC można stosować inne kryteria, często związane z oceną interfejsu systemu. Mogą być one uzyskiwane głównie w efekcie badań empirycznych. Gromadzenie danych empirycznych pochodzących od użytkowników informacji jest jednak czasochłonne, a wnioski bywają niejednoznaczne. Jest to spowodowane dużą zmiennością charakterystyk użytkowników i ich motywacji, a także zróżnicowaniem działań służących uzyskaniu dostępu do informacji. Podejmowanie więc były próby zastosowania innej metodologii. Jakob Nielsen zaproponował na przykład bardziej nieformalne podejście, nazywane przez niego oceną heurystyczną, polegające na badaniu interfejsu użytkownika na podstawie bardziej ogólnych właściwości [Nielsen 1993, s. 115].

Dla przeprowadzenia oceny jakiegokolwiek systemu czy projektu należy uwzględnić rachunek zysków i strat, oczywiście nie tylko finansowych. Wówczas ocena oznacza badanie funkcjonowania systemu lub jego części w odniesieniu do pewnych przyjętych celów. Może być ona wykonywana jako:

- Badanie efektywności funkcjonalnej: jak dobrze system (lub wybrana jego część) wykonuje zadania, dla realizacji których został zaprojektowany.
- Badanie sprawności kosztowej: koszty mogą być finansowe lub dotyczyć czasu bądź wysiłku.
- Połączenie obu typów oceny, jako efektywności kosztowej [Saracevic 2000].

Początkowo badając efektywność większy nacisk kładziono na mierzenie kosztów związanych z działalnością biblioteczną. Od początku XXI wieku większego znaczenia nabiera tendencja badania i wskazywania na korzyści funkcjonalne dostarczane przez serwis GBC [Choudhury i in. 2002].

Strategie oceny efektywności funkcjonalnej serwisów GBC mogą być oparte na trzech podstawowych kryteriach: funkcjonalności, użyteczności i dostępności [Bertot i in. 2006]. Kryteria te pozwalają na ocenę poziomu zaspokajania potrzeb użytkowników, w szczególności:

- Badanie funkcjonalności określa stopień, w jakim serwis GBC, w całości lub ocenianej części, jest zdolny do realizacji wyznaczonych działań (na przykład wyszukiwania podstawowego, wielojęzyczności).
- Wysoka użyteczność pozwala na intuicyjne i efektywne wykorzystywanie dostępnych funkcji.
- Badanie dostępności pozwala na określenie poziomu możliwego współdziałania (interakcji) użytkownika z serwisem oraz przydatności udostępnianych treści.

Jak widać, kryteria te związane są z użytkownikiem i pozwalają na ocenę poziomu zaspokojenia jego potrzeb [Fuhr i in. 2007, s. 26]. Metodologia ta uwzględnia tzw. punkty kluczowe usług serwisów GBC konstruowane przez dostawcę usług, gdzie: informacje i usługi mają być zrozumiałe do wszystkich użytkowników; funkcje niezbędne dla realizacji usług informacyjnych muszą być dostępne i działać niezawodnie; uwzględniać należy potrzeby zróżnicowanych grup użytkowników, co jest istotne ze względu na globalny zasięg usług każdego serwisu GBC.

Pomimo wyraźnych postępów w zakresie poprawy użyteczności systemów, ze względu na poważne różnicowanie zastosowań i duży rozrzut potrzeb użytkowników, trudno jest stworzyć w tym zakresie ogólny pakiet zasad, funkcjonujących w każdych warunkach, nawet, gdy swoje zainteresowania zawężymy tylko do środowiska nauki. Na podstawie podobnych kryteriów oceniać można zarówno użyteczność systemów IIN, jak i serwisów GBC: powinny one być łatwe do zrozumienia i nauczania się, winny tolerować błędy użytkownika, być elastyczne, adaptujące się, odpowiednie i efektywne w stosunku do wykonywanych zadań, niedrogie, przenośne, kompatybilne, inteligentne, wspomagające interakcje pomiędzy społecznościami i grupami, wiarygodne (bezpieczne, chroniące prywatność, godne zaufania), a nawet przyjemne w użyciu.

Użyteczność serwisów GBC związana jest z poziomem ergonomii interaktywnych urządzeń oraz aplikacji, głównie funkcjonujących w Webie. Norma ISO 9241²⁷ definiuje użyteczność jako miarę wydajności, efektywności i satysfakcji użytkownika, cechujących dany produkt, używany dla osiągnięcia określonych celów przez określonych użytkowników. Jest ona wówczas rozumiana jako:

- Zapewnienie intuicyjności nawigacji.
- Ułatwienia w zakresie poszukiwania informacji.
- Zapewnienie komunikacji zrozumiałej dla użytkownika.

Tego typu kryteria mogą stanowić zbiór ogólnych wskazówek w zakresie projektowania serwisów GBC, jednak nie mogą służyć jako instrukcja w konstruowaniu jakiegokolwiek konkretnego systemu informacyjnego. Ich zastosowanie do konkretnego systemu wymaga uwzględnienia potrzeb celowej grupy użytkowników i porównań z podobnymi, funkcjonującymi implementacjami.

Korzyści, jakie daje wzrost użyteczności systemu wydają się oczywiste; wraz z umiłowaniem zastosowań systemów komputerowych stała się ona celem także projektantów i programistów. Zdarza się, że około połowa kodu współczesnego oprogramowania dotyczy interfejsu użytkownika [Borgman 2003a, s. 121]. W takiej sytuacji nawet proste metody oceny użyteczności systemu mogą dawać poważne efekty, wprost przeliczalne na zyski finansowe. Stąd powszechne stosowanie technik optymalizacji stron Web z punktu widzenia algorytmów wyszukiwarki, czyli tzw. SEO (*search engine optimization*), które mają zapewnić długotrwałą obecność strony w wynikach wyszukiwania na jak najwyższym miejscu [Jaskowska 2009, s. 231].

Od niedawna zwraca się uwagę także na dostępność systemu, rozumianą nie tylko jako możliwość technicznego dostępu do sieci, ale także jako na przykład ułatwienia oferowane osobom niepełnosprawnym, zapewniające im jednakowy lub ekwiwalentny dostęp do informacji i usług.

Kryteria oceny GBC można również dzielić na grupy, na przykład takie, jak:

- Kryteria dotyczące transakcji, służące do pomiaru liczby otwartych sesji, aktywnych użytkowników i ich działań, w tym częstotliwości i kolejności, odsetka trafień, odsetka błędów, działań użytkownika dla poprawy błędów, liczby użytkowników jednoczesnych, długości sesji, liczby transakcji na sesję, lokalizacji użytkowników [Jose 2007, s. 232].
- Kryteria dotyczące czasu, służące pomiarowi czasu dostępu i wykonywania czynności, na przykład czasu oczekiwania na odpowiedź.
- Kryteria dotyczące kosztów, mierzące koszty działania i wydatki w różnych grupach kosztów.
- Kryteria dotyczące użytkowników, służące do mierzenia działalności i aktywności użytkowników.
- Kryteria dotyczące przydatności serwisu, takie jak dostępność, jakość interfejsu, poprawność prezentacji, kompletność wyszukiwania.
- Kryteria dotyczące poprawności funkcjonowania: poprawność odpowiedzi, trafność.

²⁷ PN-EN ISO 9241-1:2001 Wymagania ergonomiczne dotyczące pracy biurowej z zastosowaniem terminali wyposażonych w monitory ekranowe (VDT). Ogólne wprowadzenie.

Istnieją także prace, służące ocenie serwisów GBC na podstawie badania ich stron internetowych, przy czym gromadzone informacje mogą pochodzić z badanych stron lub spoza nich (badanie kontekstowe). Remigiusz Sapa za pomocą metody benchmarkingu oceniał na przykład strony Web bibliotek akademickich [Sapa 2005]. Ocena ta była prowadzona na podstawie takich kryteriów, jak: przejrzystość strony Web, jej komunikatywność, spójność i konsekwencja, przewencyjność, udostępnianie narzędzia nawigacji, sprawność, sekwencyjność, elastyczność i aktualność. Również inni autorzy podają w swoich pracach wykazy kryteriów, nazywane „listami kontrolnymi” (*check list*), według których oceniać można poszczególne serwisy GBC [Dobratz, Schoger, Strathmann 2007]. Kryteria te bywają podstawą certyfikacji tych serwisów [Dale, Ambacher 2007]. Inny sposób całościowej oceny funkcjonalnej wyszukiwania w Web przedstawił Yazdan Mansourian, który zaproponował metodologię WSE (*web search efficacy*), pozwalającą na ocenę stopnia przygotowania użytkownika, przydatności narzędzi wyszukiwawczych, strategii wyszukiwawczej, jakości wyszukanych treści oraz kontekstu, w którym odbywa się wyszukiwanie (hardware, software, jakość łącza) [Mansourian 2008, s. 351]. Badanie kontekstowe może polegać na rekomendacjach serwisu przez zaufane źródło, porównywaniu z serwisem o ustalonej reputacji oraz weryfikacji uzyskanych informacji w kilku źródłach [Metzger 2007, s. 2083-2084].

Niektóre projekty, zajmujące się oceną funkcjonowania GBC, traktują ją zbyt instytucjonalnie, przez co przyjęte miary nie są odpowiednie dla globalnej biblioteki cyfrowej jako całości, mogą natomiast być stosowane dla poszczególnych serwisów. Takie wskaźniki stosowane są na przykład w projekcie EQUINOX²⁸. Jak twierdzi Rowena Cullen, w wyniku prac prowadzonych przez różne ośrodki w ramach wielu projektów, takich jak EQUINOX, LibQUAL²⁹, eVALUED³⁰ czy COUNTER³¹ wypracowano pewien konsensus dotyczący miar opisujących funkcjonowanie serwisów GBC oraz standaryzację gromadzonych danych (zarówno przez bibliotekarzy, jak i dostawców treści), co umożliwiła porównywanie wyników [Cullen 2003, s. 203]. Standardy tworzone przez ARL, EQUINOX oraz ISO WP TC46/SC8/WG4³² stanowią podstawę dla stosowania ujednoczonych miar. Miary dla wielu obszarów, takich jak efektywność, jakość a także miary wpływu i oddziaływania serwisu GBC nie zostały dotąd ustandaryzowane. Brak jest nadal modeli, które by odzwierciedlały złożoność procesów informacyjnych w globalnej bibliotece cyfrowej.

3.4. Przestrzenie nazw

W działalności naukowej bardzo istotne znaczenie ma możliwość śledzenia powiązań pomiędzy poszczególnymi publikacjami i ich autorami w celu określenia źródeł nowych pomysłów i idei [Borgman 2007, s. 92]. Już w środowisku publikacji drukowanych wyko-

²⁸ EQUINOX jest projektem finansowanym przez Unię Europejską. Wskazuje on na potrzebę rozwoju i stosowania przez wszystkie biblioteki metod pomiaru funkcjonalności w nowym środowisku sieciowym (<http://equinox.dcu.ie/>).

²⁹ LibQUAL+ jest zestawem narzędzi dla bibliotekarzy, służących badaniu opinii użytkowników na temat jakości usług. Narzędzia te oferowane są przez ARL. Dzięki niemu bibliotekarze mogą doskonalić usługi biblioteczne, podnosić kulturę organizacyjną oraz urynkowanie biblioteki (<http://www.libqual.org/>).

³⁰ EVALUED obsługuje bazę danych wyników badań w dziedzinie bibliotekoznawstwa posiadowaną w Birmingham City University. Celem projektu jest stworzenie modelu oceny biblioteki cyfrowej (<http://www.ebase.bcu.ac.uk/index.htm>).

³¹ COUNTER powstał w 2003 r. w Wlk. Brytanii. Utworzył zestaw standardów i protokołów dotyczących zapisu i wymiany danych dotyczących wykorzystania zasobów online. Opublikował zasady gromadzenia statystyk wykorzystania baz danych online (<http://www.projectcounter.org/>).

³² Głównie chodzi o normę ISO 11620:1998 Information and documentation -- Library performance indicators.

nywano wiele pracochłonnych czynności, pozwalających na śledzenie powiązań pomiędzy pracami cytującymi i cytowanymi. W środowisku cyfrowym powiązania pomiędzy obiektami cyfrowymi mogą być reprezentowane przez tworzone pomiędzy nimi odnośniki. Reprezentacja intelektualnych powiązań pomiędzy publikacjami i zawartymi w nich informacjami, z wykorzystaniem odnośników, pozwala na rozwój dyskursu naukowego.

W środowisku GBC odnośnik, wychodzący z obiektu cyfrowego, powinien wskazywać unikalny egzemplarz dzieła, a nie jedną z wielu kopii, jak to jest praktykowane w bibliotekach tradycyjnych. Dzięki temu, że odnośniki bezpośrednio wiążą połączone obiekty, ich opisy bibliograficzne nie są niezbędne, a przynajmniej odgrywają inną rolę. Mogą one służyć określeniu autentyczności powiązanych obiektów (pozwalać na stwierdzenie czy obiekty są tym, za co je uważamy). Śledzenie powiązań między obiektami cyfrowymi możliwe jest więc dzięki zdolności do identyfikowania poszczególnych obiektów i ich zestawów funkcjonujących w GBC.

Właściwa identyfikacja odbywa się za pomocą identyfikatorów (zob. nast. podrozdz.), umieszczonych w pewnej, uzgodnionej przestrzeni czy systemie nazw. Przestrzeń nazw jest to zestaw nazw, wyznaczany przez zbiór zasad lub algorytmów służących do zarządzania nimi. Nazwy mogą być stosowane jako etykiety lub odesłania do zestawu obiektów. Zagadnienia związane z przestrzeniami nazw mają zasadnicze znaczenie dla infrastruktury Web (w tym w szczególności dla XML) [Duval i in. 2002]. Mechanizm przestrzeni nazw umożliwia ujednoczenie schematów metadanych na poziomie syntaktycznym, pozostawiając odpowiednią elastyczność na poziomie semantycznym, gdzie zapewnione jest niezbędne różnicowanie.

Najlepiej, aby przestrzenie nazw miały wsparcie instytucjonalne, co zwiększa ich wiarygodność i ułatwia przejmowanie przestrzeni przez kolejne instytucje. W celu dotarcia do określonego obiektu wykorzystywanych może być wiele przestrzeni nazw. Elementy metadanych opisowych książki mogą być na przykład pobierane z przestrzeni nazw odpowiednich dla oznaczenia odpowiedzialności, wydawcy, ISBN, haseł przedmiotowych. Częstym sposobem organizacji przestrzeni nazw jest stosowanie kartotek wzorcowych i autorytarnych na różnych poziomach organizacyjnych oraz słowników języków informacyjno-wyszukiwawczych. Elementy metadanych składające się na rekord pochodzą więc zazwyczaj z wielu przestrzeni nazw.

Dla obiektów cyfrowych, funkcjonujących w obrębie GBC, istnieje obecnie wiele konkurencyjnych przestrzeni nazw. Część z nich to modyfikacja przestrzeni nazw przygotowanych dla publikacji tradycyjnych, inne tworzone są wyłącznie z myślą o obiektach cyfrowych. Podstawowe znaczenie ma na przykład główny protokół Web, HTTP, który jest przestrzenią nazw gwarantującą unikalność każdego URI.

Pomimo pomyślnego prowadzenia prac nad standardami, łączącymi przestrzenie nazw, wciąż na pełne rozwiązanie oczekują problemy ich jednolitości i długotrwałości. Dzięki odpowiedniemu wsparciu instytucjonalnemu istnieje dobrze zorganizowana kontrola nad elementami komunikacji formalnej, gorzej wygląda sytuacja w zakresie zapewnienia identyfikacji i długotrwałego przechowywania w obszarze komunikacji nieformalnej. Poziomą zapewnianą stabilność dla tego typu materiałów może zależeć od tego, czy są przeznaczone do krótkotrwałego udostępniania, czy długotrwałego przechowywania.

Przestrzenie nazw szczególną rolę odgrywają w XML, gdzie stosuje się je jako mechanizm tworzenia globalnie unikalnych nazw dla elementów i atrybutów w języku kodowania. Dzięki unikalnym nazwom zapobiega się konfliktom znaczeniowym w zakresie nazw, stosowanych w dokumentach XML, możliwe jest także łączenie różnych elementów kodu z zachowaniem jednoznaczności nazw. Wyróżnienie przestrzeni nazw powoduje rozdział stosowanych terminów (elementy i atrybuty) na zestawy (słowniki) odpowiednie dla poszczególnych dziedzin i zastosowań. Zestawy nazw elementów i atrybutów tworzą autory dokumentów XML. Są to tak zwane nazwy lokalne, ponieważ zakres ich zastosowania

i kontekst znaczeniowy nie wykracza poza dany dokument [Zieliński 2000, s. 65]. Aby uniknąć konfliktu nazw, polegającego na istnieniu identycznych nazw mających odmienne znaczenia, stworzono możliwość określenia kontekstu używanych nazw elementów i atrybutów. Dokonuje się tego poprzez łączenie nazwy z odpowiednim adresem URL, co określane jest mianem nazwy uniwersalnej.

Szczegółowy opis sposobów deklarowania i zarządzania przestrzeniami nazw w XML wykracza poza zakres tej publikacji, jednak ogólnie mówiąc idea ta ma zasadnicze znaczenie dla tworzenia infrastruktury niezbędnej dla stosowania modularnych systemów metadanych w GBC. Deklaracja przestrzeni nazw pozwala projektantowi schematu metadanych na zdefiniowanie kontekstu każdego stosowanego terminu, co gwarantuje uzyskanie unikalności terminów w granicach zdefiniowanej przestrzeni nazw. Deklaracja przestrzeni nazw w obrębie zestawu metadanych pozwala na identyfikację elementów metadanych jako należących do określonego zestawu elementów.

Wykorzystując tego typu infrastrukturę, projektanci systemu metadanych mogą wybierać elementy z przydatnych im, istniejących zestawów elementów metadanych, dzięki czemu wykorzystują prace i doświadczenia różnych społeczności, związanych z tworzeniem opisów obiektów cyfrowych. To pozwala unikać wielokrotnego tworzenia schematów metadanych od początku dla każdego nowego zastosowania (por. także profile aplikacyjne w p. 2.7). Zgodnie z przyjętymi zasadami utworzenie nowego elementu metadanych, nie istniejącego we wcześniej utworzonych przestrzeniach nazw, powoduje konieczność stworzenia dla niego nowej przestrzeni nazw i zarządzania nią³³ [Heery, Patel 2000].

Schematy metadanych mogą być także integrowane w obrębie RDF. W tym modelu danych wprowadzono możliwość tworzenia lokalnych słowników, odpowiednich dla realizacji potrzeb lokalnych społeczności, w tym wymiany terminologii. Dla efektywnego łączenia wyrażen RDF z określonym słownikiem lub schematem RDF wykorzystywany jest mechanizm przestrzeni nazw XML. W ten sposób rekord RDF scala wiele opisów, które mogły być tworzone w różnych okresach dla różnych celów, w jeden, nowy opis.

3.5. Identyfikatory obiektów cyfrowych

Pojawienie się naukowych publikacji elektronicznych spowodowało wzrost oczekiwań co do ich dostępności. Uczni chcą mieć możliwość korzystania z tekstów na ekranach swoich komputerów bez potrzeby odwiedzania w tym celu biblioteki. Realizacja tych oczekiwań wymaga jednak dalszej rozbudowy infrastruktury oraz tworzenia nowych narzędzi. W dodatku oczekiwania wciąż rosną: standardem w zastosowaniach naukowych staje się możliwość bezpośredniego przejścia od wykazu literatury do tekstów cytowanych dokumentów; system wyszukiwania informacji jest w stanie dostarczyć nie tylko metadanych interesujących obiektów, ale także łatwy dostęp do pełnego tekstu [Apps, MacIntyre 2006]. Istotne może być także zapewnienie jednoznacznych powiązań pomiędzy fizycznymi przedmiotami i cyfrowymi informacjami o nich; typowym przykładem może być stosowanie ISBN dla książki drukowanej w sieciowym systemie bibliotecznym lub księgarni. Te i inne usługi GBC mogą być realizowane dzięki właściwej identyfikacji obiektów cyfrowych o różnym poziomie agregacji, co w rozległym i heterogenicznym środowisku stwarza zasadnicze problemy³⁴. Wynikają one m.in. ze stosowania nowych

³³ Tworzenie własnego schematu metadanych przestawić można jako realizację etapów: 1) wybór podstawowej przestrzeni nazw metadanych, 2) wybór elementów z innych przestrzeni nazw metadanych, 3) określenie własnych elementów metadanych i definicja ich przestrzeni nazw, 4. rozwiązywanie problemów aplikacyjnych elementów metadanych [Chan, Zeng 2006].

³⁴ Nie mniejsze problemy stanowi identyfikacja i autoryzacja użytkowników serwisów GBC, o czym mowa jest w p. 5.5.

narzędzi; na przykład prowadząc metawyszukiwanie należy identyfikować zarówno zasób, jak i serwis wyszukiwawczy. To jeszcze nie wszystko; Juha Hakala wyróżnia siedem poziomów identyfikacji: organizacje, ludzie (kartoteki autorytarne), zasoby i usługi, dzieła i realizacje, materializacje, części składowe oraz atrybuty wyszukiwawcze [Hakala 2006, s. 363]. W dalszej części rozdziału zajmę się identyfikacją dzieł i ich cyfrowych materializacji.

Pierwsze rozwiązania polegały na implementacji statycznych odnośników do publikacji tworzonych przez ich dostawców. Odnośniki od odsyłaczy bibliograficznych do cytowanych publikacji były kodowane na stałe, początkowo w obrębie zamkniętego zasobu obiektów, na przykład treści publikowanych przez danego wydawcę lub zagregowanej bazy danych czasopism elektronicznych. Następnie kolejno wprowadzono je w obrębie treści posiadanych przez dostawców na podstawie indywidualnych porozumień pomiędzy nimi, co czasem powodowało potrzebę zawierania dodatkowej umowy o przenieśnięcie z biblioteką. Odnośniki te tworzone były zazwyczaj przy użyciu własnej syntaktyki. W efekcie po to, aby tworzyć odnośniki od wyszukanych metadanych do cytowanego tekstu, system informacyjny musiał implementować odrębny zestaw odnośników dla każdego wydawcy.

Kolejnym etapem było powstanie kilku syntaktyk odnośników, wykorzystujących standardowe metody identyfikacji lub opisu za pomocą metadanych wyszukanych obiektów, dzięki czemu powstać może odpowiedni Uniform Resource Locator (URL). Do wykonania odnośnika do czasopisma można więc wykorzystać jego ISSN, natomiast przez dodanie danych o woluminie, zeszytach i paginacji otrzymujemy identyfikator artykułu, co wykorzystane zostało w SICI³⁵. Odnośniki w obrębie strony właściciela treści publikacji elektronicznych mogą zawierać metadane obiektu umieszczone w jego adresie, przez co URL jest efektywnie sparametryzowany i opisany. Proces ten jest zazwyczaj wykonywany przy pomocy programu CGI, który pobiera metadane z ciągu będącego zapytaniem URL, co pozwala na dostęp do wielu obiektów przez tę samą bramkę. Odnośniki te nadal używają własnych formatów, różnych dla poszczególnych dostawców treści.

Jak pisze Christine Borgman, uwidocznienie nazw domen w formie URL, które mogą służyć jako identyfikatory, było konsekwencją powstania Webu [Borgman 2007, s. 16]. W czasach, gdy Web dopiero się rozwijał i miał znacznie mniejsze rozmiary niż obecnie, często proponowano użytkownikom zgadywanie nazw domen, jako sposób na szybkie dotarcie do nich. Nadal wielką wartość komercyjną mają nazwy łatwe do odgadnięcia (a przy okazji i do zapamiętania), co wiąże się także z procederem zastrzegania tysięcy nazw, które mogłyby mieć dla kogoś wartość, w celu ich odsprzedawania.

Identyfikator obiektu cyfrowego, aby właściwie spełnić swą rolę w zakresie publikacji naukowych, musi być unikalny, co umożliwi tworzenie jednoznacznych powiązań. Musi być on także niezmienny, aby odnośnik do obiektu mógł być używany przez długi czas, w którym identyfikowany obiekt może być przenoszony na różne komputery, do innego formatu (na przykład do nowej wersji oprogramowania), a nawet zmieniać właściciela. URL nie jest wystarczająco stabilny, aby mógł być używany w tym celu w formie, w jakiej funkcjonuje obecnie. W krótkim okresie poprawił on współdziałanie serwisów GBC, lecz nie może być rozwiązaniem docelowym.

W celu zwiększenia pewności odszukania identyfikowanego obiektu powstały zasady tworzenia dodatkowego poziomu narzędzi pomiędzy wyszukiwarką i tym obiektem. Niezmienny identyfikator przestał pełnić rolę odsyłacza do określonego pliku (wystąpienia) obiektu cyfrowego, zamiast tego stanowi formę opisu obiektu cyfrowego. Takie pośred-

³⁵ SICI – Serial Item and Contribution Identifier, standard Z39.56, wykorzystujący ISSN do identyfikacji czasopism i artykułów. Może też być stosowany do tworzenia odnośników z bibliografii załącznikowej do artykułów w formie elektronicznej.

nie identyfikowanie wymaga istnienia dodatkowego poziomu resolvera³⁶, który odsyła do właściwego pliku. Opis obiektu (metadane), przygotowany dla resolvera, pozwala wyszukiwarce na odnalezienie konkretnego obiektu. Usługi resolvera polegają na przekierowywaniu użytkownika do odpowiedniego (na przykład najnowszego) egzemplarza obiektu. Przekierowywanie jest zwykle niewidoczne dla użytkownika [Tonkin 2008].

Technologia resolverów wykorzystana została w systemach pośredniczących, określających miejsce, w którym obiekt może być odnaleziony. Najbardziej znanym przykładem tych prac jest Digital Object Identifier (DOI³⁷), oparty na resolverze Handle³⁸. W oparciu o DOI utworzony został CrossRef³⁹, serwis pozwalający na tworzenie odnośników pomiędzy wydawcami. W ten sposób DOI zapewnia syntaktykę dla niezmiennego identyfikatora, skierowującego do zidentyfikowanego obiektu [Chandrakar 2006, s. 446].

Zaprezentowane rozwiązania pozwalają na tworzenie odnośników od metadanych do opisywanego obiektu, jednak nie biorą pod uwagę uwarunkowań użytkownika, szczególnie jego uprawnień do korzystania ze wskazanego obiektu, który może być dostępny z wielu miejsc, w których stosuje się różne zasady dostępu, w tym opłaty. Biblioteki naukowe zazwyczaj zawierają umowy na prenumeratę czasopisma w formie elektronicznej lub też na dostęp do serwisu agregatora. Użytkownik biblioteki, posiadający bezpłatny dostęp do zakupionych (zaprenumerowanych) przez bibliotekę obiektów, będzie zaskoczony niepoprawnym odnośnikiem i żądaniem zapłaty w przypadku, gdy spróbuje uzyskać dostęp do tych obiektów w innym miejscu (na przykład w księgarni internetowej). Problem ten, wynikający z potrzeby dostarczania odnośników do egzemplarza obiektu, do którego użytkownik ma uprawnienia lub do jego wersji Open Access, znany jest jako problem „odpowiedniego egzemplarza” [Beit-Arie i in. 2001].

Próbą jego rozwiązania jest OpenURL, ujednociający syntaktykę odnośników. Jest to protokół, który pozwala na łączenie różnego rodzaju zasobów: metadanych w bazach danych i OPAC z opisywanymi dokumentami pełnotekstowymi [Zhu 2004, s. 254]. Metadane zasobu, takie jak jego ISBN, autor, tytuł są kodowane w postaci URL, co pozwala na tworzenie powiązań pomiędzy zasobami informacji a usługami serwisów GBC. Biblioteki naukowe i inne instytucje udostępniają resolvery, kupowane lub tworzone we własnym zakresie, interpretujące OpenURL. Wykorzystywany jest on przez dostawców baz danych do przesyłania metadanych do resolverów linków, które skierowują zapytania OpenURL do odpowiednich serwisów, udostępniających pełne teksty. Identyfikator ten został w swojej otwartej wersji szeroko zaimplementowany przez społeczności związane z informacją naukową. Serwisy udostępniające pełne teksty, takie jak czasopisma elektroniczne, wykorzystujące odnośniki z bibliografii załącznikowych czy też serwisy abstraktowych baz danych z odnośnikami do pełnych tekstów, nie muszą wysyłać zapytania do strony określonego czasopisma, ale mogą skierować je do resolvera OpenURL w instytucji użytkownika. Ogólnie mówiąc OpenURL definiuje kilka elementów opisu (pól), które przekazują informację bibliograficzną do serwera Web. Częścią resolvera OpenURL jest zazwyczaj

³⁶ Resolver to program, który zajmuje się wysłaniem zapytania do serwerów nazw (*name server*) i dostarczeniem odpowiedzi (informacji) o pytaną domenę internetową. Udziela informacji na pytania zadawane przez komputery – klientów. Resolver odczytuje OpenURL i wyszukuje informacje w nim zawarte w bibliotecznej bazie danych dla określenia ich dostępności. Następnie wyświetla informację użytkownikowi o tym, czy pozycja znajduje się w bibliotece i o sposobie dostępu (np. dostęp online, wypożyczalnia międzybiblioteczna). Istnieje wiele resolverów dostarczanych przez firmy komercyjne: Resorce Linker (Ulrichs), LinkResource (EBSCO), Discovery Resolver (Eandavor), WebBridge (Innovative), LinkSolver (Ovid), Serials Solution Article Linker (ProQuest), SFX (Ex Libris) i in.

³⁷ <http://www.doi.org/>.

³⁸ <http://www.handle.net/>.

³⁹ <http://www.crossref.org/>.

„baza wiedzy”, zawierająca dane o obiektach elektronicznych, udostępnianych w danym serwisie⁴⁰, wraz z informacjami o sposobie kodowania odnośników do wielkiej liczby zasobów elektronicznych, zarówno publikowanych komercyjnie, jak i w Open Access.

Usługi OpenURL mogą być udostępniane na stronach Web serwisów GBC. Na przykład SFX dostarcza narzędzie zwane Citation Linker, pozwalające użytkownikowi końcowemu na ręczne wprowadzanie metadanych za pomocą odpowiedniego formularza. Na tej podstawie tworzony OpenURL wysyłany jest do resolvera SFX, dzięki czemu użytkownikowi przedstawiane są serwisy przydatne z punktu widzenia wprowadzonych metadanych.

Sukces OpenURL w dużej mierze wynika z jego prostoty [Dahl, Banerjee, Spalti 2006, s. 45]. Może on być jednak stosowany do przekazywania metadanych w ograniczonym zakresie, chociaż możliwe są rozszerzenia, polegające na przykład na zastosowaniu DOI do uzyskania dostępu do dowolnych informacji o zasobie i do samego zasobu. OpenURL nie zastępuje DOI, ani na odwrót. Są to raczej uzupełniające się technologie, które stosowane mogą być wraz z lokalnymi usługami przekierowującymi, pozwalającymi na dostęp użytkownika do źródła.

W oparciu o technologię OpenURL, w środowisku bibliotekarzy i wydawców powstał schemat Info URI⁴¹. Wykorzystuje on URI wyłącznie w roli identyfikatora, a więc narzędzia nie stosowanego do innych celów, na przykład do wyszukiwania. URI pozwala na prostą identyfikację zasobów w obrębie globalnej architektury Web. Wykorzystywany jest jako identyfikator zasobów informacyjnych, posiadających identyfikatory w publicznych przestrzeniach nazw, takich jak LC Control Number czy DOI, ale nie mających swojej reprezentacji w URI. Dzięki temu identyfikowane zasoby informacyjne mogą być opisywane z zastosowaniem technologii Web, takich jak XLink, RDF czy mapy pojęć⁴² (*Topic maps*).

Wspomnieć można także o próbach łączenia identyfikatorów obiektów cyfrowych z identyfikatorami osób, które są pomocne w rozwiązywaniu takich problemów jak odróżnianie osób o tym samym nazwisku, zmiana nazwiska przez autora, różne formy tej samej nazwy osobowej. W tym zakresie prowadzone są różne eksperymenty; wymienić można prace nad połączeniem OpenID⁴³ dla autorów z DOI w celu łączenia autorów z ich tekstami.

Nie istnieje technologia pozwalająca stworzyć niezmienny identyfikator obiektu cyfrowego w takim sensie, w jakim termin ten używany jest w stosunku do bibliotek tradycyjnych. Jest to spowodowane głównie przez fakt, że długotrwałość większości schematów identyfikatorów zależy w znacznym stopniu od zdolności dostawcy obiektów do ich stałego udostępniania i aktualizacji. Oferowanie identyfikatora obiektu poprzez dany resolver oznacza potrzebę długotrwałej obsługi również resolvera. Wybór odpowiedniego rozwiązania w tym zakresie wpływa na poziom dostępności tej części infrastruktury GBC.

⁴⁰ Zazwyczaj w bazie znajdują się takie metadane, jak: tytuł czasopisma, nazwa wydawcy/agregatora czasopisma, ISSN, rok wydania, URL strony, z której czasopismo może być przeszukiwane.

⁴¹ <http://info-uri.info/registry/docs/misc/faq.html>.

⁴² Mapy Pojęć (czasem nazywane mapami tematycznymi) służą reprezentacji i wymianie wiedzy. W postaci grafu prezentują charakterystyki formalne tekstów wraz z ich relacjami bibliograficznymi, instytucjonalne i personalne powiązania twórców tekstów oraz pojęcia zawarte w tekstach [Sosińska-Kalata 2008, s. 113]. Tworzone zgodnie z normą ISO/IEC 13250:2003 Topic Maps. Information Technology.

⁴³ OpenID jest zbiorem protokołów i specyfikacji, umożliwiających tworzenie zdecentralizowanych systemów obsługi tożsamości cyfrowej, kontrolowanej przez użytkownika. Jest on oparty na URI. OpenID pozwala na logowanie się do wielu serwisów przy pomocy jednego hasła oraz tworzenie wirtualnych wizytówek. <http://openid.net> oraz <http://openid.pl>.

3.6. Interfejsy użytkownika

W globalnej bibliotece cyfrowej istnieje wielka mnogość interfejsów użytkownika oraz jednocześnie dość silna unifikacja w tym zakresie. Interfejs użytkownika do usługi serwisu GBC jest ogólnym, abstrakcyjnym opisem serwisu, określającym jego zadania, dostępne obiekty wraz z metodami przedstawiania relacji między nimi oraz sposoby prowadzenia interakcji z użytkownikiem. Interfejs definiuje rodzaje komunikatów, oczekiwanych i obsługiwanych przez serwis oraz sposoby ich dostarczania. Określone są także wszelkie warunki i wymagania dotyczące komunikatów. Jeżeli przyjmiemy, że interfejs jest formalnym przedstawieniem możliwości serwisu, jego zawartości i oczekiwanego sposobu otrzymywania i wysyłania komunikatów wraz z wymogami dotyczącymi struktury i relacji między komunikatami, to łatwo zauważyć, że sam interfejs zawiera wiele szczegółowo ustrukturyzowanych informacji.

Najogólniej można wyróżnić dwa poziomy interfejsów: interfejs Web i powszechnie stosowanych systemów operacyjnych (głównie Windows) oraz interfejsy poszczególnych serwisów, wykorzystujących rozwiązania przyjęte w interfejsach wyższego poziomu. Pierwszym interfejsem, z którego korzysta użytkownik informacji jest zazwyczaj interfejs przeglądarki Web i wyszukiwarki internetowej. W dalszej kolejności posługuje się on interfejsami udostępnianymi przez poszczególne serwisy, zazwyczaj opartymi na jakiegoś rodzaju Windowsach. Jednak praktycznie wszystkie interfejsy, współcześnie stosowane przez użytkowników, tworzą jedną, spójną całość, choć różniącą się szczegółami projektu interakcji użytkownika z systemem i sposobem prezentacji wyników pracy systemu komputerowego. Sprawia to, że użytkownik znający jakikolwiek interfejs graficzny (GUI) szybko nabywa umiejętności działania w dowolnym innym interfejsie⁴⁴. Dzięki temu GBC pozostaje stosunkowo łatwym w użyciu systemem, pomimo tego, że poszczególne serwisy mogą być systemami złożonymi technologicznie, oferującymi wiele różnorodnych funkcji i możliwości.

Interfejsy użytkownika, zarówno na poziomie globalnej biblioteki cyfrowej, jak i poszczególnych jej serwisów, odgrywają wielką rolę jako narzędzia umożliwiające interakcję użytkownika z zasobami informacyjnymi [Tedd, Large, s. 129]. Interfejs użytkownika jest częścią systemu komputerowego (sprzętu i oprogramowania), którą można obejrzeć, usłyszeć, dotknąć, mówić do niego lub w jakiś inny sposób porozumiewać bądź zwracać się. Składa się on z dwóch części: część wejścia zarządza sposobami komunikowania przez użytkownika potrzeb i żądań do systemu, a część wyjścia określa sposób prezentacji użytkownikowi wyników pracy systemu.

Interfejs, dostosowany do potrzeb i możliwości użytkownika, zasadniczo wpływa na poziom jego satysfakcji, zarówno dotyczącej wyników samego wyszukiwania informacji, jak i ogólnej satysfakcji wynikającej z korzystania z systemu. Jak stwierdził William Arms, biblioteka cyfrowa jest tylko tak dobra, jak jej interfejs [Arms 2000, s. 160]. Ponieważ jednak publiczność GBC jest mocno zróżnicowana, twórcy systemów dodatkowo sta-

⁴⁴ Graficzny interfejs użytkownika (GUI), podczas tworzenia którego zastosowano metaforę biurka (desktop metaphor) w odróżnieniu od interfejsu rozkazów, zazwyczaj zawiera następujące elementy: wskaźnik (symbol ukazujący się na ekranie, który jest przesuwany w celu wyboru obiektów i rozkazów); urządzenie wskaźnika (urządzenie służące do przesuwania wskaźnika i dokonywania wyborów, takie jak mysz); ikonki (małe obrazki reprezentujące rozkazy, pliki, foldery lub okna), wybór następuje przez kliknięcie (kliknięcia); pulpit (obszar na ekranie, na którym umieszczone są ikony, co symbolizuje rzeczywiste obiekty na rzeczywistym pulpicie biurka); okna (ekran monitora można podzielić na kilka obszarów, w każdym z nich można uruchomić inny program lub wyświetlić plik; okno można przesuwać na pulpicie, można też zmieniać jego rozmiar i kształt), pomiędzy oknami można przenosić ikonki; rozwijane menu (interfejsy graficzne zazwyczaj pozwalają na wybór rozkazu także z menu różnego rodzaju). Poza graficznymi elementami wizualnymi GUI ułatwiają wymianę danych pomiędzy aplikacjami ze względu na stosowane standardowe formaty danych dla tekstu i grafiki. Różne programy pracujące pod wspólnym GUI mogą więc pracować na tych samych danych.

rają się umożliwić kastomizację sposobu interakcji z systemem. Jest ona tym łatwiejsza, im bardziej twórcy systemów przestrzegają zasad obowiązujących podczas projektowania interfejsów.

Z punktu widzenia globalnej biblioteki cyfrowej należy zwrócić uwagę nie tylko na podstawowe zasady projektowania, ale także na specyficzne zagadnienia związane z różnorodnością kulturową i językową. W szczególności chodzi tu o wpływ języka i wzorców kulturowych użytkownika na percepcję interfejsu, co powinno znaleźć odbicie w decyzjach jego projektanta.

Interfejsy systemów komputerowych funkcjonują w różnych środowiskach sprzętowych i programowych. Najważniejszymi elementami środowiska są zadania, które będą w nim realizowane oraz profil obsługiwanego użytkownika. Interfejs GBC musi umożliwiać użytkownikom realizację określonych działań w celu zaspokojenia ich potrzeb informacyjnych – przez odszukanie relewantnych obiektów cyfrowych. Interfejsy poszczególnych serwisów GBC różnią się między sobą z wielu powodów. Po pierwsze, serwisy te są bardzo odmienne ze względu na rozmiar, zakres udostępnianych materiałów, stosowane języki i normy kulturowe. Po drugie, różnią się one stosowaną platformą technologiczną, na której są posadowione: większość dostępna jest poprzez Web za pomocą typowych komputerów PC. Wówczas serwisy GBC dostępne są dla użytkownika za pomocą protokołu TCP/IP, głównie HTTP jako nośnika zarówno warstwy informacyjnej, jak i prezentacyjnej, czyli interfejsu. Interfejsem użytkownika w takim przypadku jest wybrana strona Web, wyświetlona w przeglądarce. Coraz częściej należy brać także pod uwagę inny sprzęt, głównie mobilny. Tego rodzaju urządzenia, ze względu na zupełnie inne rozmiary ekranów, mogą wymagać całkowitego przeprojektowania interfejsów użytkownika. Duże możliwości w tym zakresie daje XHTML, który umożliwia przenoszalność stron na urządzenia typu palmtopów czy telefony komórkowe. Po trzecie, społeczności użytkowników także różnią się między sobą, co może pociągać potrzebę zmiany interfejsu. Należy również pamiętać o tym, że użytkownikami serwisów GBC są także bibliotekarze czy pracownicy informacji. Dla nich tworzy się zazwyczaj specjalne, odrębne interfejsy „robocze”. Różnice pomiędzy interfejsami serwisów GBC mogą wynikać też z tego, że nie istnieje jedynie słuszna droga do uzyskania interfejsu idealnego; projektowanie interfejsów, podobnie jak każde projektowanie, umiejscawia się pomiędzy sztuką, nauką i techniką.

Pomimo tych zróżnicowań można przedstawić kilka zasadniczych funkcji, które muszą być realizowane przez każdy interfejs serwisu GBC. Można wyróżnić przynajmniej dwa aspekty dotyczące użyteczności strony Web: wizualno-funkcjonalny oraz treściowy [Malak 2007, s. 257]. Pierwszy aspekt dotyczy takich zagadnień, jak:

- Przedstawianie treści: zasoby poszczególnych serwisów mogą być przedstawiane w różny sposób, na przykład graficznie lub przy pomocy wykazów tekstowych; istotne jest tu rozmieszczenie elementów na stronie oraz możliwość nawigacji.
- Wyszukiwanie i przeglądanie: wyszukiwanie jest opisane w innej części tej książki (zob. p. 4.3). Najczęściej stosuje się dwa sposoby wyszukiwania: proste (pojedyncze okienko do wprowadzania danych) i złożone, a także różnego rodzaju menu i odnośniki tekstowe i graficzne. Ostatnio interfejs służący wyszukiwaniu w różnych serwisach ulega ujednoliceniu, o czym świadczy na przykład upodobnienie interfejsu WorldCat do Amazon i Google.
- Wyświetlanie wyszukanej informacji: użytkownik otrzymuje wyszukaną informację lub reprezentującą ją metainformację, dzięki czemu może wybrać najbardziej relewantne obiekty.

Drugi aspekt (treściowy) dotyczy sposobu przekazywania treści użytkownikowi, w tym właściwy dobór języków kodowania. W tym zakresie można zauważyć jednokierunkowe funkcjonowanie przeglądarek internetowych; służą one jedynie dla umożliwie-

nia użytkownikowi dostępu do odległego źródła informacji. W celu utworzenia własnej informacji poprzez przeglądarkę, użytkownik musi zastosować mediację poprzez to źródło. Przeglądarka nie służy bowiem tworzeniu nowych treści [Wright 2007, s. 201].

Obecnie, w związku z olbrzymim nagromadzeniem danych i informacji, coraz większego znaczenia nabiera taka ich prezentacja, która maksymalnie ułatwia percepcję. Jednym ze sposobów jest tworzenie interfejsów zapewniających odpowiednią wizualizację informacji. Jak pisze Veslava Osińska, proces wizualizacji wiąże dwa najpotężniejsze systemy przetwarzania informacji – ludzki mózg oraz współczesny komputer. Wizualizacja informacji ma na celu zwiększenie ludzkich możliwości w zakresie przetwarzania i rozumienia informacji poprzez uruchomienie ludzkiego systemu przetwarzania wizualnego. Badania nad wizualizacją danych i informacji są efektem wysiłków uczonych, starających się doskonalić jakość interpretacji danych, gromadzonych przez urządzenia pomiarowe lub wygenerowanych w trakcie symulacji komputerowych. Dotyczą one bezpośrednio takich aspektów, jak: wizualizacja danych naukowych, dialog człowiek-komputer, analiza danych, eksploracja danych, wyobraźnia i grafika. Wizualizacja informacji jest obecnie uważana za odrębną część grafiki komputerowej, wykorzystuje najnowsze osiągnięcia nauk biologicznych oraz inżynierii. Metody wizualizacji informacji ewaluowały od interfejsów z minimalną ilością elementów graficznych, używających zagnieżdżonych drzew klasyfikacji, tabel oraz wykresów dwuwymiarowych, przez diagramy relacji między dokumentami przy użyciu takich abstrakcyjnych kształtów, jak koła, kwadraty, linie oraz łąca, do przeglądarek hiperbolicznych i geoprzestrzennych map z włączoną osią czasu [Osińska 2006]. Dzięki odpowiedniej wizualizacji informacji możliwe jest na przykład graficzne obrazowanie relacji, występujących pomiędzy uczonymi (relacje nieformalne) i publikacjami (cytowania) oraz pomiędzy dokumentami i osobami, na temat których dane przechowywane są w repozytoriach [Baptista, Ferreira 2007].

Wizualizacja informacji wiąże się z trzema, wzajemnie powiązаныmi zagadnieniami: wizualną reprezentacją, interakcją i rozumowaniem [Fast, Sedig 2005]. Reprezentacja wizualna polega na kodowaniu informacji do formy graficznej w sposób umożliwiający ludzką, efektywną percepcję wizualną. Znacznie podnosi ona zdolność do przyswajania wielkich ilości informacji, odkrywania złożonych zależności, zrozumienia skomplikowanych struktur oraz tworzenia nowej wiedzy. Jednak wizualizacja informacji to coś więcej niż reprezentacja i percepcja wizualna. Dotyczy ona także zagadnień związanych z interakcją ludzi i tych reprezentacji. W kontekście wizualizacji informacji interakcja jest działaniem podejmowanym przez użytkownika na reprezentacji w formie kolejnych odpowiedzi. Dodanie interakcji do reprezentacji wizualnej może wyraźnie zwiększyć możliwości użytkownika w zakresie badania danych, tworzenia zapytań, nawigacji, transformacji, manipulacji i pracy na elementach i cechach reprezentacji. Rolą interakcji jest wspomaganie i doskonalenie inteligentnych czynności, takich jak wnioskowanie, rozwiązywanie problemów, podejmowanie decyzji i określanie znaczenia. Właściwa interakcja może prowadzić do powstawania nowych poglądów, odkryć i rozumienia zjawisk. Wizualizacja informacji ma więc zastosowanie przede wszystkim tam, gdzie istnieje złożona informacja, którą ludzie muszą zarządzać, nawigować, wyszukiwać, rozumieć, wnioskować na jej podstawie i ogólnie współdziałać z nią w celu zdobywania i tworzenia wiedzy. Stąd łatwo jest przewidzieć dobre efekty zastosowania wizualizacji informacji w serwisach GBC.

Do niedawna zastosowania nowoczesnych interfejsów, w tym wizualizacji informacji, były ograniczane przez dwa czynniki: po pierwsze badania w zakresie wizualizacji informacji i bibliotek cyfrowych prowadzone były niezależnie od siebie; po drugie preferowanie Webu jako mechanizmu udostępniania i rozpowszechniania informacji powodowało istotne ograniczenia techniczne, w tym zmniejszenie możliwości przesyłu danych i interakcji. Obecnie oba te czynniki są mniej istotne; przeglądarki internetowe zostały znacznie udoskonalone, a przepustowość sieci szybko rośnie. Oznacza to możliwość stosowania coraz bardziej rozbudowanych interfejsów.

W tym rozdziale opiszę procesy, według terminologii 5S nazywane Scenariuszami; określają one sposoby wykorzystania serwisów GBC. Łącznie wyznaczają one ogólną funkcjonalność GBC w komunikacji naukowej. Pozwalają na właściwą reprezentację połączenia funkcji z danymi i niezbędną infrastrukturą, co ułatwia poprawne projektowanie serwisów GBC.

Od początku XXI w. funkcjonuje nowy sposób całościowego ujmowania zagadnień, związanych z funkcjonowaniem danych i informacji cyfrowych w infrastrukturze nauki. Podejście to zaczęto nazywać cyfrową ochroną lub kuratelą nad danymi (*digital curation*). Termin ten stosowany jest coraz częściej dla całości niezbędnych działań, związanych z zarządzaniem danymi tworzonymi przez e-naukę i innymi obiektami cyfrowymi w trakcie całego cyklu życia oraz w czasie ich bieżącego i przyszłego użytkowania [Beagrie 2006, s. 4].

Termin „cyfrowa ochrona” jest wciąż bardzo nowy, przez co jego zakres nie jest do końca określony. W p. 2.6 wymieniłem cztery funkcje formalnej komunikacji naukowej, realizowanej głównie za pomocą publikacji; jedną z nich jest archiwizacja. Właśnie z takimi procesami, jak cyfrowa archiwizacja i długotrwałe przechowywanie początkowo kojarzona była cyfrowa ochrona. Obecnie dotyczy pełnego cyklu życia obiektu cyfrowego, począwszy od jego konceptualizacji, poprzez kolejne transformacje, aż po likwidację, chociaż zagadnienia długotrwałej archiwizacji zajmują wciąż bardzo ważne miejsce. Sarah Higgins wyróżniła następujące działania w cyklu życia obiektu cyfrowego, wraz z czynnościami związanymi z cyfrową ochroną¹:

- Konceptualizacja: wymyślenie i zaplanowanie procesu tworzenia obiektu cyfrowego, w tym metoda tworzenia i opcje zapisu.
- Tworzenie: powstawanie obiektu cyfrowego wraz z metadanymi administracyjnymi, opisowymi, strukturalnymi i technicznymi.
- Dostęp i użytkowanie: zapewnienie stałego dostępu dla określonych użytkowników. Udostępnienie odbywa się przez opublikowanie, wraz ze stosowaniem kontroli dostępu i procedur uwierzytelnienia (*authentication*) (zob. p. 4.4).
- Ocena i selekcja: ocena obiektu cyfrowego i selekcja do długotrwałej ochrony i archiwizacji. Jej podstawą powinny być uzgodnione wytyczne, polityka lub wymagania prawne i techniczne (na przykład możliwość obsługi określonego formatu), a nawet organizacyjne (na przykład istnienie odpowiednich metadanych).
- Usunięcie: usuwanie obiektów nie wybranych do długotrwałej ochrony i archiwizacji. Bezpieczne wykonanie tych czynności może opierać się na wytycznych, polityce lub przepisach prawnych.
- Akcesja: transfer obiektów do archiwum, repozytorium, bazy danych lub innego miejsca przechowywania. Jej podstawą powinny być uzgodnione wytyczne, polityka lub wymagania prawne.
- Czynności archiwizacyjne: realizacja prac zapewniających długotrwałą archiwizację i zachowanie wartości autorytarnej obiektu cyfrowego. Pozwalają one na utrzymanie integralności obiektów, dzięki czemu są one autentyczne, wiarygodne i gotowe do za-

¹ Działania te dotyczą i są realizowane w poszczególnych serwisach GBC.

stosowania. Działania te to walidacja, tworzenie metadanych archiwalnych, tworzenie metainformacji reprezentującej obiekt oraz właściwy dobór struktur danych i formatów plików.

- Ponowna ocena: przekazanie obiektów cyfrowych, które nie przeszły procedur walidacyjnych do kolejnych ocen i ewentualnego wyboru.
- Magazynowanie: przechowywanie danych w bezpiecznych warunkach, odpowiadających przyjętym standardom.
- Dostęp w celu późniejszego stosowania: zapewnienie dostępu do obiektu zarówno bieżącym, jak i późniejszym użytkownikom (*reusers*). Udostępnienie może nastąpić w formie publicznie dostępnej informacji opublikowanej wraz ze stosowaniem kontroli dostępu i procedur uwierzytelnienia.
- Transformacja: tworzenie nowego obiektu cyfrowego z istniejącego wcześniej [Higgins 2007, s. 85].

Wyróżnione elementy modelu cyklu życia obiektu cyfrowego świadczą o potrzebie całościowego podejścia do funkcjonowania cyfrowych publikacji naukowych; nie wystarczy już tylko przygotowanie pliku w jakimś formacie i umieszczenie go na stronie Web. Elementy te, stanowiące scenariusze realizowane w GBC, bardziej szczegółowo opisałem w kolejnych podrozdziałach tego rozdziału.

4.1. Publikowanie

Jedną z dziedzin, w której zmiany, związane ze stosowaniem nowych technologii informacyjnych są najbardziej widoczne, jest system publikowania naukowego. Jest on częścią systemu komunikacji naukowej, w którym funkcjonuje GBC. Pojawienie się w nim masowych ilości publikacji elektronicznych w różnych formach, szczególnie czasopism elektronicznych, spowodowało zachwianie ustabilizowanych wcześniej relacji pomiędzy uczestnikami komunikacji naukowej: autorami (uczonymi), wydawcami, dostawcami (w tym agregatorami) i bibliotekami naukowymi. Pomimo wątpliwości zgłaszanych przez niektórych autorów [Cisek, Sapa 2007, s. 43], publikowanie stało się tańsze (szczególnie w przypadku zastosowania nowych modeli publikowania) [Fisher 2008], łatwiejsze i szybsze. Dotyczy to szczególnie sposobów publikowania zrywających z wymogami druku. Coraz śmiało wykorzystywane są możliwości medium elektronicznego, dzięki czemu publikacje elektroniczne zyskują cechy nigdy niedostępne w świecie publikacji tradycyjnych. Obecnie autorzy nie potrzebują już pośredników w docieraniu do swoich odbiorców – za publikowanie uznaje się umieszczenie danego obiektu (tekstu, multimediów, danych, programów) w sieci przez kogokolwiek, na przykład twórcę obiektu cyfrowego, chociaż współczesne naukowe publikowanie elektroniczne przybiera coraz bardziej zorganizowaną formę (zob. p. 2.6).

Publikowanie elektroniczne z punktu widzenia stosowanych technologii i ich możliwości powinno być traktowane jako nowy rodzaj działalności, a nie część lub przedłużenie publikowania tradycyjnego [Feather 2003, s. 168]. Podstawą publikowania elektronicznego są procesy digitalizacji i cyfryzacji, a więc różne sposoby tworzenia obiektów cyfrowych. Współczesny system publikowania naukowego wciąż charakteryzuje się współistnieniem form drukowanych i elektronicznych. Prawdopodobnie właśnie duży stopień hybrydyzacji powoduje, że z punktu widzenia zastosowania nowych technologii można wyróżnić dwa rodzaje publikacji elektronicznych. Dużą ich część stanowią materiały, dla których kopiuje się wprost rozwiązania stosowane w druku; dla nich wciąż dominującym paradygmatem jest metafora dwuwymiarowej „stronicy”. Wydawcy starają się, aby publikacja na ekranie wyglądała dokładnie tak samo, jak na papierze [Nentwich 2003, s. 322]. Cyfryzacja procesu publikowania w środowisku sieciowym pozwala jednak na

zastosowanie nowych reprezentacji informacji, pozwalających na nowe interakcje [Wright 2007, s. 229]. Stąd pojawiają się głosy, że czasopisma elektroniczne także są formą przejściową i należy spodziewać się nowych sposobów komunikowania naukowego, które w pełni (lub przynajmniej pełniej) wykorzystają elektroniczne medium online. Początkowo nowe formy zapewne występować będą obok elektronicznych kopii publikacji tradycyjnych, powodując ich mutowanie w kierunku form typowych wyłącznie dla środowiska elektronicznego, udostępniających funkcje niemożliwe do realizacji w środowisku druku. Przykładem mogą być nowe formy publikacji, korzystające z modelu komunikacji sieci społecznej Web 2.0 [Landow 2006, s. 77-80].

Obecnie znaczną większość dokumentów przygotowuje się w środowisku cyfrowym, przy użyciu narzędzi typu edytorów tekstów, programów do przygotowania grafiki, wizualizacji, zarządzania danymi. Na realizację wszystkich procesów związanych ze współczesnym publikowaniem wpływ ma elektronika. Inny sposób ich realizacji jest obecnie nieefektywny ekonomicznie. Nadal jednak znaczna część publikacji przygotowanych w formie cyfrowej, w celu dystrybucji, konwertowana jest na wersję analogową [Gawrysiak 2008a, s. 190], przez co tracone są korzyści wynikające z rozpowszechniania elektronicznego (takie, jak duży zasięg i szybkość).

Wydaje się, że przyczyną konwersji do druku dużej liczby publikacji naukowych jest nienadążanie zmian społecznych, charakteryzujących się znaczną inercją, za rozwojem technologicznym. Technologie najlepsze z czysto technicznego punktu widzenia często bywają zapomniane na rzecz innych, mniej efektywnych. Na poziom ich rozpowszechnienia składa się kompleks zagadnień społecznych i ekonomicznych [Hine 2006, s. 27]. Gdyby zagadnienia związane z wyborem formy publikacji były wyłącznie natury technologicznej, wszystkie publikacje od dawna byłyby cyfrowe, gdyż z technologicznego punktu widzenia brak ku temu przeszkód. Jednak technologia jest tylko jednym z wyznaczników najlepszych sposobów produkcji i dystrybucji treści naukowych. Ważne są też przyzwyczajenia odbiorców, możliwości percepcji, wyznaczone latami stosowania technologii analogowej i poziom oczekiwania (akceptacji) zmian. Wyraźnie działa tu prawo opóźnienia kulturowego.

Przejście od druku do publikowania elektronicznego powoduje zmiany trojkiego rodzaju: doskonalenie (ulepszanie), zwielokrotnienie (zwiększenie) i transformację (zmianę sposobu postępowania) tego, co udało się osiągnąć wcześniej przy pomocy druku [Treloar 1999]. Publikowanie elektroniczne:

1. Doskonali:

- nawigację w obrębie dokumentów i pomiędzy nimi;
- szybkość publikowania i rozpowszechniania informacji.

2. Zwielokrotnia:

- strukturę i nowe formy treści (multimedia);
- połączenia pomiędzy dokumentami, ich częściami oraz ich zasobami;
- objętość publikacji, pozwalając na jej (niemal) dowolny rozmiar;
- możliwości funkcji wyszukiwawczych;
- dostępność i umożliwia stosowanie nowych modeli dystrybucji.

3. Transformuje:

- tekst w hipertekst;
- stabilność (niezmiennność) tekstu w kierunku dokumentu dynamicznego;
- dokument tradycyjny w dokument interaktywny;
- publikacje tradycyjne w nowe rodzaje publikacji;
- sposoby kontroli jakości, w tym recenzowania w szczególności.

Tak jak każde zjawisko o charakterze technologicznym, digitalizacja i cyfryzacja nie funkcjonują w oderwaniu od całego otoczenia technologicznego. Procesy te dotyczą nie tylko transformacji komunikacji od analogowej do cyfrowej, ale także wykorzystania ca-

tego zakresu urządzeń i technologii, systemów i procedur. Do najważniejszych elementów technologii elektronicznego publikowania zaliczyć można komputer osobisty i sieci, w szczególności globalne. Komputer osobisty umożliwia funkcjonowanie technologii informacyjnej, zwielokrotniającej możliwości ciała ludzkiego, jako narzędzia do pisania i czytania [Owen 2007, s. 130]. Połączenie komputerów w sieci znacznie wzmacnia to zwielokrotnienie dzięki tworzeniu powiązań z wielką liczbą obiektów cyfrowych, wytworów osób i grup².

Opublikowanie informacji w Internecie powoduje, przynajmniej w teorii, jej powszechną, globalną dostępność. Oznacza to, że twórca źródła informacji umieszcza je w węzle sieci, z którego jest dostępne dla każdego, zawsze i zewsząd. Z punktu widzenia komunikacji naukowej dostępność globalna jest podstawową cechą cyfrowej informacji sieciowej. Publikacja cyfrowa jest dostępna globalnie z określonego miejsca w Sieci, a odnośniki, prowadzące od publikacji do innych źródeł, pozwalają na dostęp do tych źródeł. Właśnie takie cechy, jak globalna dostępność, możliwość kontrolowania przez twórców (autorów, wydawców) i nieograniczona „łączliwość”³ są głównymi cechami charakterystycznymi odróżniającymi zasoby informacji cyfrowej i sieciowej od zasobów drukowanych. Dzięki gęstej sieci powiązań i etykiet, tworzonych przez każdego użytkownika (zarówno autora treści naukowych, jak i ich odbiorcę) powstają niejawne relacje, tworzące infrastrukturę wiedzy [Weinberger 2007, s. 171].

Oczywiście nawet, gdy informacja jest cyfrowa i sieciowa, część wymienionych cech może ulegać ograniczeniu, w wyniku kontroli dostępu przez pośrednika (na przykład wydawcę, bibliotekę naukową), usytuowanego pomiędzy twórcą a użytkownikiem. Pośrednik może ograniczyć globalną dostępność publikacji, na przykład dla ochrony praw autorskich czy zwiększenia własnych zysków. W przypadku lokalnego przechowywania i udostępniania obiektu w serwisie GBC, traci się jego cechy dynamiczne. Oprócz technicznych możliwości rozszerzenia kontroli i dostępności, istnieją także ograniczenia społeczne, wynikające z negocjacji prowadzonych pomiędzy uczestnikami procesu publikowania.

Zastosowanie technologii informacyjnych wpływa na przyspieszenie realizacji wielu działań związanych z nauką, w tym publikowania. Ze względu na przedłużającą się okres przygotowania publikacji do druku, już nie tylko publikacje książkowe, ale nawet czasopisma tradycyjne, stają się zbyt wolnym medium. Nie jest czymś wyjątkowym, że artykuł ukazuje się w czasopiśmie dopiero po dwóch lub trzech latach od jego przygotowania; w wielu dziedzinach taki tekst jest już nieaktualny. Podobnie lub jeszcze gorzej sytuacja wygląda w zakresie publikacji książkowych, przy czym w wielu dyscyplinach, rozwijających się w szybkim tempie, publikacje książkowe utraciły walor naukowości, najpierw na rzecz artykułów naukowych, a obecnie źródeł internetowych [Evans, James 2008, s. 399].

Technologie komputerowe, szczególnie związane z publikowaniem elektronicznym, pozwalają na udostępnienie publikacji cyfrowej znacznie wcześniej, niż w wersji drukowanej. Można wyróżnić następujące przyczyny przyspieszenia:

- Przyspieszeniu ulega praca redakcyjna dzięki zastosowaniu EJS⁴ (Electronic Journal

² Koncepcja przedłużenia ciała zgodna jest z koncepcją uniwersalnej książki Paula Otleta, który uważał ją za mechanizm dynamiczny; na podobieństwo maszyn, będących przedłużeniem ludzkiego ciała, jest ona przedłużeniem ludzkiego mózgu, w szczególności pamięci. Wizja uniwersalnej książki Otleta jako systematycznego, pełnego i aktualnego rejestru faktów dotyczących poszczególnych dziedzin wiedzy może być porównywana z Xanadu Nelsona czy Memexem Busha.

³ Publikacja drukowana również zapewnia podobną „łączliwość” poprzez stosowanie odsyłaczy różnego rodzaju (w tym bibliograficznych), jednak możliwość swobodnego przechodzenia między działami i ich częściami w środowisku cyfrowym jest tak duża, że tworzy nową jakość.

⁴ Przykładami takich systemów są Eprints (<http://www.eprints.org/>), Open Journal Systems (OJS – <http://pkp.sfu.ca/?q=ojs>), Fedora (<http://www.fedora-commons.org/>) czy DSpace (<http://www.dspace.org/>).

System) oraz poczty elektronicznej jako środka komunikacji niemal symultanicznej. Znacznemu przyspieszeniu ulega komunikacja pomiędzy autorami, recenzentami, redaktorami i pracownikami wydawnictwa, dzięki przekazywaniu niezbędnych materiałów w formie bezpapierowej i w dużej części w sposób zautomatyzowany.

- Format i rytm publikowania wydawnictw drukowanych, zbudowanych z kolejnych zeszytów, o ograniczonej liczbie stron, a przez to objętości, w wyniku komputeryzacji podlega znacznej modernizacji. Nowe obiekty, odpowiedniki wcześniejszych artykułów, mogą być umieszczane w sieci, czyli publikowane natychmiast po przygotowaniu (ocenie jakości, zredagowaniu, sformatowaniu). Cecha ta bywa nazywana „publikowaniem nieprzerwanym” (*continuous publishing*). W efekcie usunięte mogą być opóźnienia, wynikające z tradycyjnej formy publikacji (w odróżnieniu od przyczyn dotyczących bezpośrednio tekstu).
- Czas niezbędny na druk i przesyłkę obiektów fizycznych redukowany jest do zera. Co prawda konieczne jest przygotowanie układu strony, jednak sam druk i dystrybucja obiektów fizycznych do użytkowników są eliminowane z cyklu wydawniczego przez ich zastąpienie umieszczeniem plików na serwerze. Pliki te trafiają więc bezpośrednio do użytkownika, a nie do drukarni, w której odbywa się czasochłonna transformacja na formę analogową. Odbiorca końcowy może drukować je lokalnie.

Długość cyklu wydawniczego artykułu jest bardzo zróżnicowana i wynosi od kilku dni do kilku lat, w zależności od przyjętego sposobu komunikacji i modelu czasopisma (drukowane, elektroniczne, repozytorium). Tak duże różnice wynikają z wielu przyczyn, takich jak kłopoty telekomunikacyjne, spóźnione recenzje, różnorodność standardów dla układu tekstu i jego kontroli. Analizy przepływów czasowych pozwalają na podział cyklu wydawniczego na dwie fazy: od dostarczenia tekstu do jego końcowej akceptacji po recenzjach i zredagowaniu oraz od ostatecznej akceptacji do ukazania się artykułu, czyli jego publikacji [Mabe 2009, s. 6]. Zastosowanie narzędzi komputerowych przyspiesza wyraźnie realizację drugiej fazy, podczas gdy czas trwania pierwszej nie ulega zasadniczym zmianom, gdyż środki komunikacji elektronicznej, stosowane w pierwszej fazie, są obecnie rozpowszechnione we wszystkich wydawnictwach, także tych, które przygotowują publikacje drukowane. Wynika z tego, że z wymienionych trzech przyczyn przyspieszenia tylko dwie ostatnie stwarzają różnicę pomiędzy publikowaniem tradycyjnym i elektronicznym. Szczególnie potrzeba oceny jakości publikowanych materiałów (recenzowanie) powoduje znaczne spowolnienie procesów, bez względu na stosowane technologie⁵. Największe oszczędności czasu uzyskać można dzięki wyłamaniu się z obowiązującego rytmu publikowania czasopisma, wynikającego z jego sztywno ustalonej częstotliwości.

Równie poważne zmiany następują w zakresie dystrybucji treści naukowych. Publikacje naukowe stają się coraz bardziej dostępne. Dzięki usługom powiadamiania, takim jak RSS użytkownicy mogą uzyskiwać informacje o dokonywanych zmianach i nowych materiałach na stronach Web. Publikacje elektroniczne dostępne są online, dzięki czemu, z technicznego punktu widzenia, w każdej chwili istnieje możliwość skopiowania interesującej publikacji na własny komputer. Są one lepiej ustrukturyzowane, przy czym nie chodzi tu o struktury hipertekstowe, ale dostarczanie publikacji w wygodnych częściach, bez potrzeby kompletowania całej książki lub zeszytu czasopisma przed opublikowaniem. Rezultatem jest dostęp kastomizowany: użytkownik uzyskuje dostęp tylko do tych elementów treści publikacji, które są dla niego interesujące.

Publikowanie elektroniczne przesuwają odpowiedzialność za selekcję materiałów w procesie publikacji, która wiąże się z kontrolą ich jakości, przedstawioną w dalszej części rozdziału. Procesy publikowania tradycyjnego zapewniały selekcję wielostopnio-

⁵ Wraz z publikowaniem elektronicznym proponowane są także nowe formy oceny jakości (np. po opublikowaniu), wpływające na tempo prac (zob. p. 4.8).

wą – najpierw autor decydował co i gdzie opublikować⁶, następnie wydawca podejmował decyzję o przyjęciu materiału do druku, a w końcu bibliotekarz ustalał, czy włączyć publikację do zbiorów i udostępnić ją użytkownikom. Obecnie wystarczającą jest decyzja autora, aby użytkownicy otrzymali bezpośredni dostęp do utworzonych przez niego treści (zob. model, rozdz. 6). Z drugiej strony decyzje o tym, co rozpowszechniać elektronicznie i w jakiej kolejności, coraz częściej podejmują także bibliotekarze, publikujący w bibliotekach cyfrowych, czy też tworzący elektroniczne reedycje. Selekcja materiałów w formie cyfrowej ma szczególne znaczenie dla funkcjonowania wielu serwisów globalnej biblioteki cyfrowej, gdzie podejmuje się decyzje o długotrwałej archiwizacji obiektów cyfrowych; niezbędne jest tu znalezienie właściwej proporcji między efektywnością kosztową i zapewnieniem dostępu do dziedzictwa kulturowego. Liczy się również szybkość podejmowanych decyzji.

Inną różnicą w stosunku do publikowania tradycyjnego jest zróżnicowanie funkcji publikacji cyfrowych. Egzemplarz książki drukowanej spełniał jednocześnie wiele zadań: umożliwiał zapoznanie się z treścią dzieła, jednocześnie pozwalając zachować te treści w stałej, niezmiennej formie w długim okresie. W trakcie publikowania cyfrowego w formie digitalizacji publikacji tradycyjnych, zachodzi często potrzeba podejmowania decyzji o rozdzieleniu tych funkcji, gdyż ze względu na stan oryginału należy zdecydować: czy zachować formę zbliżoną do oryginału (aby można było zobaczyć na ekranie „jak oryginał wyglądał”), czy też umożliwić odczytanie tekstu dokumentu (w oryginale niedostatecznie czytelny). W przypadku podjęcia decyzji o zachowaniu obu funkcji niezbędne może być utworzenie dwóch odrębnych, ale związanych ze sobą obiektów cyfrowych dla tej samej publikacji, pozwalających na realizację różnych zadań.

Proces transformacji publikowania naukowego, wynikający ze zmian technologicznych i organizacyjnych, jeszcze się nie zakończył. Powstają i powstawać będą nowe formy i sposoby tworzenia publikacji naukowych; część z nich związana jest z funkcjonowaniem publikacji udostępnianych w trybie Open Access.

Według wielu autorów jednym z podstawowych czynników wpływających na rozwój globalnej biblioteki cyfrowej może być umożliwienie swobodnego i globalnego dostępu do zasobów publikacji naukowych w trybie tzw. **Open Access (OA)** [Brody 2006; Harnad 1992, 2001; Yiotis 2005]. Świadczy o tym jedna z pierwszych deklaracji przyjętych w sprawie rozwoju Open Access w Budapeszcie⁷, na początku 2002 r. Mowa jest w niej o rozproszonych zasobach obiektów cyfrowych, powszechnie dostępnych i tworzonych z wykorzystaniem standardów OAI, które dzięki narzędziom indeksującym mogą być traktowane jako jeden wielki zasób danych. Użytkownicy nie muszą martwić się o to, jakie archiwa funkcjonują, ani gdzie są one zlokalizowane, żeby móc wyszukiwać w nich potrzebną informację i uzyskiwać do niej bezpośredni dostęp.

W deklaracji BOAI znaleźć można definicję Open Access, która zawiera kilka zasadniczych elementów, dobrze charakteryzujących podstawowe cechy OA:

- publikacje Open Access udostępniane są bezpłatnie (dla użytkownika końcowego), co nie znaczy, że ich przygotowanie nie powoduje powstawania żadnych kosztów;
- są one udostępniane online, co znaczy, że są to obiekty cyfrowe dostępne w Internecie. Dzięki temu uniknąć można dużej części kosztów, gdyż w publikowaniu cyfrowym istnieje wyłącznie koszt uzyskania pierwszego egzemplarza⁸; w środowisku online jest to jedyny egzemplarz, co odróżnia publikacje elektroniczne online od drukowanych;

⁶ Christine Borgman wymienia wśród kryteriów wyboru przez autora miejsca publikacji takie czynniki, jak: prestiż czasopisma, etap realizacji opisywanego projektu, zwyczaj panujący w danej dziedzinie wiedzy [Borgman 2003a, s. 102].

⁷ Budapest Open Access Initiative: <http://www.soros.org/openaccess/read.shtml>.

⁸ Koszt uzyskania pierwszego egzemplarza stanowi największą część kosztów, bez względu na sposób publikowania.

- są to publikacje naukowe, więc nie chodzi o literaturę piękną, gazety czy popularne poradniki;
- OA nie powoduje obniżenia jakości czasopisma: zarówno czasopisma OA, jak komercyjne, są różnej jakości; w obu grupach funkcjonuje jej kontrola [Oppenheim, Rowland 2009, s. 280];
- publikowanie w trybie Open Access zwiększa stopień wykorzystania tych materiałów, między innymi dzięki ułatwieniu dostępu [Nahotko 2007c; Nicholas, Huntington, Jamali 2007, s. 858]. Oprócz obowiązku podawania danych o autorstwie nie istnieją żadne inne ograniczenia w wykorzystaniu materiałów⁹;
- istnieją dwie główne strategie Open Access: autoarchiwizacja w różnego rodzaju repozytoriach (dziedzinowych, instytucjonalnych – zob. p. 2.6) czy na własnej stronie Web (zieleny OA, strategia BOAI-I) oraz publikowanie w czasopismach OA¹⁰ (złoty OA, strategia BOAI-II) [Bailey 2006, s. 15].

Jak z tego wynika, OA jest raczej ogólnym celem niż strategią marketingową, gdyż przedstawia różne sposoby dojścia do wyznaczonego stanu, bez względu na wybrany model biznesowy.

Istnieją definicje, które w istotny sposób zawężają zakres OA, na przykład nie uznają za publikacje OA materiałów umieszczanych na własnych stronach Web, pomijających proces recenzowania i obróbki redakcyjnej [Borgman 2007, s. 100]. Z drugiej strony właśnie rozpowszechnianie materiałów, w tradycyjnym systemie komunikacji naukowej często w ogóle nie uznawanych za publikacje, jak dane z badań, dokumenty instytucjonalne czy materiały instruktażowe, może być szansą dla rozwoju repozytoriów OA.

Według Josepha Esposito, w komunikacji naukowej istnieje miejsce dla wielu form publikowania, gdyż spełniają one różne funkcje [Esposito 2008]. Uczni z kolegami prowadzącymi podobne badania, stanowiący dobrze znający się krąg osób, komunikują się w sposób nieformalny; niepotrzebne jest im pośrednictwo wydawcy ani jego usługi, takie jak recenzowanie, gdyż w tej grupie każdy sam najlepiej potrafi ocenić jakość pracy naukowej. Również staranne opracowanie redakcyjne tekstu nie ma tu wielkiego znaczenia. Uczni z tej samej dyscypliny, ale zajmujący się innymi zagadnieniami, być może potrzebują już pewnego wspomnienia w ocenie materiałów. Dla nich istnienie formalnej publikacji może oznaczać, że warto poświęcić opisanemu zagadnieniu nieco czasu. Tym bardziej cenne są usługi wydawców i bibliotekarzy dla specjalistów z innych dyscyplin, w tym pokrewnych. Kolejna grupa odbiorców, luźno związanych z nauką, jak administratorzy różnych poziomów, poprzestają na wskaźnikach typu Impact Factor, sugerujących im, w co warto inwestować publiczne fundusze. Użytkownikami są też osoby, chcące mieć jakieś wyobrażenie o sensie pracy uczonych; dla nich powstają wydawnictwa popularnonaukowe. Dla każdej z tych i być może wielu innych, niewymienionych grup uczestników komunikacji naukowej należy zastosować inny rodzaj publikacji: OA lub komercyjnej – drukowanej lub elektronicznej.

Rozdźwięk pomiędzy wzrostem dostępności, możliwym dzięki nowym narzędziom elektronicznym, głównie Internetowi, a wzrostem monopolizacji rynku publikacji (powodującym gwałtowny wzrost cen), realizowanym przez wydawców, stał się istotną siłą napędową dla OA. Open Access pojawił się więc jako reakcja na kryzys i niestabilność tradycyjnego modelu publikowania naukowego (zob. p. 1.2), jednak w toku licznych dyskusji środek ciężkości przesunął się od prób rozwiązywania szczegółowych, indywidual-

⁹ Istnieją jednak sposoby bardziej szczegółowego określenia powinności użytkownika, na przykład na zasadzie licencji Creative Commons, o których będzie mowa w dalszej części książki.

¹⁰ Strategia polegająca na publikowaniu w czasopismach elektronicznych bywa jeszcze dzielona na złoty OA, gdy koszty publikacji ponosi autor (tak jak w przypadku czasopism PLoS i innych) oraz na platynowy OA, gdy od autora nie są pobierane żadne opłaty. W obu przypadkach z punktu widzenia użytkownika czasopismo jest bezpłatne.

alnych zagadnień ku badaniu zjawisk towarzyszących komunikacji naukowej w całości. Szczególnie interesujące wydają się próby odpowiedzi na pytanie, jak wyglądałaby komunikacja naukowa i w ogóle świat nauki, gdyby OA zdobył co najmniej przewagę, jeżeli nie całkowicie zawładnął publikowaniem naukowym [Shadbolt i in. 2006, s. 195-202]. Z tych prac można wyciągnąć dwa wnioski:

- Uczeni publikują głównie w tym celu, aby być czytanim i cytowanym, w znacznie mniejszym stopniu dla osiągnięcia bezpośrednich dochodów w postaci honorariów autorskich¹¹. Udostępnienie publikacji poza systemem publikowania tradycyjnego, blokującym dostępność do publikacji przez zbyt wysokie koszty prenumeraty, pozwala na zwiększenie wskaźników wpływu (Impact Factor), czyli liczby cytowań uzyskanych we wpływowych publikacjach naukowych.
- Słaby rozwój OA utrudnia dostęp uczonych do potencjalnie ważnych informacji w przypadku, gdy ich instytucje nie prowadzą prenumeraty odpowiednich publikacji. Możliwość dostępu jest więc zależna od zamożności instytucji, przez co dystans do optymalnej sytuacji pełnego dostępu waha się, od nieistotnego w bogatych instytucjach, do dramatycznego w ubogich ośrodkach badawczych. To z kolei oznacza, że wiele wyników badań nie jest rozpowszechnianych w optymalny sposób, więc nie docierają one do wielu zdolnych uczonych. Sytuacja, w której na opłacany dostęp do wyników badań naukowych mogą sobie pozwolić tylko najbogatsi, stanowi poważny bodziec do zmian w kierunku OA [Guédon 2006, s. 28].

Połączenie nowych technologii informatycznych, w tym przetwarzania danych masowych oraz zasobów informacji w postaci publikacji naukowych, daje nowe możliwości, trudne do wyobrażenia nie tylko w środowisku publikacji tradycyjnych, ale także współczesnych systemów online, czyli w czasach, gdy dalece nie wszystkie dane są udostępniane jako OA. Możliwości te dotyczą nie tylko cyfrowego publikowania, funkcjonowania komunikacji naukowej, ale także sposobu realizacji badań naukowych w ogóle. W związku z tymi korzyściami należy spodziewać się szybkiego rozwoju publikowania OA w różnych jego przejawach. OA powoduje powstawanie uniwersalnego dobra publicznego: im wyższej jakości informacja jest udostępniana, tym dobro publiczne jest większe [Hess, Ostrom 2006, s. 13].

Nie tylko to jednak jest efektem rozwoju ruchu OA. Według Clifforda Lyncha, głównym wkładem OA w funkcjonowanie komunikacji naukowej jest gruntowna rekonceptualizacja sposobów korzystania z publikacji naukowych jako efektu badań naukowych [Lynch 2006a, s. 193]. W rekonceptualizacji tej bezpłatne udostępnianie publikacji naukowych nie jest najważniejszym elementem; również w tradycyjnym systemie komunikacji naukowej biblioteki naukowe zapewniają użytkownikom bezpłatny dostęp do posiadanych materiałów. OA powoduje gruntowną zmianę ról pośredników – wydawców i bibliotek. Są oni zastępowani przez samych autorów (pracowników nauki), wspieranych przez IIN i wdrażających nowe technologiczne, organizacyjne i prawne modele publikowania. W rezultacie dostęp do materiałów umieszczanych w sieci staje się globalny, a dotychczasowi pośrednicy uparczywie poszukują dla siebie nowych ról (zob. p. 5.1).

4.2. Tworzenie metadanych

Jak już wspomniałem, metadane mogą być tworzone, najogólniej rzecz ujmując, przez człowieka lub automatycznie, bez jego udziału lub z udziałem ograniczonym. W pierwszym przypadku powstają one na poziomie serwisów GBC, w drugim chodzi często

¹¹ Publikacje służą raczej zapewnieniu prawa pierwszeństwa do odkrycia naukowego i zdobyciu prestiżu w środowisku, niż bezpośrednim dochodom. Te pojawiają się raczej w związku z osiągnięciami naukowymi (m. in. pierwszeństwo) oraz posiadaniem prestiżem.

o efekt pracy wyszukiwarek globalnych. To, w jaki sposób powstają zależy od tradycji dyscypliny, której dotyczą dane, rodzaju zasobu informacji, posiadanych lub dostępnych narzędzi oraz oczekiwanych efektów. Często są to prace zespołowe, wykorzystujące różne metody i narzędzia.

Według badań przeprowadzonych wśród naukowych bibliotek amerykańskich, metadane najczęściej tworzy się manualnie, co nie jest wynikiem zaskakującym, ze względu na to, że to właśnie biblioteki naukowe dysponują najlepiej wyszkolonym i doświadczonym personelem. Na 68 badanych bibliotek tylko jedna tworzyła metadane całkowicie automatycznie; 9 bibliotek łączyło obie możliwości, a 16 do automatycznie utworzonych metadanych wprowadzało ręczne poprawki. Tworzenie metadanych jest funkcją często wykonywaną we współpracy między instytucjami, o czym świadczy fakt, że 52% bibliotek przyjmowało metadane z instytucji współpracujących, a 29% również od dostawców dokumentów. 73% bibliotek korzysta z konwersji metadanych [Ma 2007, s. 12].

Metadane tworzone są także automatycznie przez wyszukiwarki internetowe, podczas indeksowania treści obiektów Web. Wówczas cały tekst dokumentu stanowi jego reprezentację. Takie metadane nazywane są czasem metadanymi nieustrukturyzowanymi, w odróżnieniu od tworzonych manualnie, z wykorzystaniem skomplikowanych i wielopoziomowych struktur danych. Nazwa ta może być myląca, gdyż wyszukiwarki również tworzą struktury w swoich zasobach indeksowych [Battelle 2006, s. 22]. Metadane, powstające w wyszukiwarkach, stają się podstawą współczesnego wyszukiwania informacji w GBC (zob. p. 2.8).

Dużą część metadanych administracyjnych i strukturalnych jest tworzona przez pracowników technicznych, najczęściej tych samych, którzy digitalizują czy w inny sposób tworzą prymarnie obiekty cyfrowe (dane). Tego rodzaju metadane mogą być również stosunkowo łatwo tworzone automatycznie. Inaczej wygląda sytuacja metadanych opisowych, podobnych do tradycyjnych opisów bibliograficznych.

W latach ubiegłych częste były opinie, że tworzenie metadanych opisowych jest zajęciem typowo intelektualnym, a więc właściwym wyłącznie człowiekowi [Milstead, Feldman 1999]. Trudno sobie było wyobrazić bardziej efektywny sposób tworzenia surogatów dokumentów, o bogatej strukturze wewnętrznej i odzwierciedlających różnorodne relacje bibliograficzne. Brano pod uwagę dwie grupy twórców metadanych:

- autorów zasobów, na przykład stron Web; zaletą tego rozwiązania jest rozproszenie prac: metadane powstają tam, gdzie dane;
- osoby zawodowo zajmujące się katalogowaniem i indeksowaniem, a więc bibliotekarzy i specjalistów z zakresu informacji. Nie brak im profesjonalizmu i umiejętności, więc można spodziewać się opisów dobrej jakości. Jednak bardzo szybko okazało się, że zasobów Internetu przybywa tak szybko, że ich pełne skatalogowanie przez specjalistów jest fizycznie niemożliwe.

Stosunkowo dogodna jest sytuacja, gdy twórca obiektu cyfrowego sam go udostępni wraz z metadanymi (na przykład w bibliotece cyfrowej lub repozytorium). Jest to szczególnie przydatne w przypadku zestawów danych, dla których twórca posiada wyłączną wiedzę o ich przydatności, wartości i zawartości, a automatyczne indeksowanie utrudnione jest przez brak informacji tekstowej.

Czasami podczas tworzenia metadanych współpracują ze sobą twórca zasobu i specjalista z zakresu tworzenia metadanych. Ten pierwszy może przygotowywać podstawowy, pierwszy szkic rekordu metadanych, tworząc elementy, co do zawartości których nie ma wątpliwości. Bibliotekarz może uzupełniać lub kontrolować rekordy metadanych dla zapewnienia jednolitości opisów i ich zgodności z przyjętymi zasadami.

Twórcy obiektów cyfrowych, podczas tworzenia metadanych wspomagani są przez różnego rodzaju narzędzia, ułatwiające te procesy. Są to zarówno narzędzia dostępne bezpłatnie, jak również tworzone przez firmy komercyjne. Można wyróżnić kilka rodzajów tych narzędzi:

- Formularze pozwalają użytkownikom na wprowadzanie wartości metadanych w obrębie zdefiniowanego zestawu elementów, w wyniku czego tworzony jest rekord metadanych.
- Narzędzia służące kodowaniu metadanych w odpowiednim języku kodowania (takim, jak HTML, XML lub SGML).
- Narzędzia konwertujące, pozwalające zmienić pierwotny schemat metadanych na inny.
- Narzędzia automatycznego tworzenia metadanych na podstawie analizy treści obiektu cyfrowego, o której mowa w dalszej części rozdziału.

Obecnie wciąż stosunkowo niewiele obiektów Web zawiera ręcznie kodowane metadane. Być może w przyszłości ich liczba wzrośnie dzięki temu, że autorzy rozumieją korzyści płynące z dołączania metadanych, rozpowszechnione zostaną standardy służące ich kodowaniu, powstaną narzędzia udostępniające interfejsy maksymalnie ułatwiające ich przygotowanie. Można na to liczyć szczególnie w komunikacji naukowej, gdzie częścią kwalifikacji zawodowych pracownika nauki jest umiejętność przygotowania tekstu zawierającego wyniki badań, a obecnie dołączyła do tego umiejętność zakodowania go w sposób czytelny dla komputera; na tym w dużej mierze polegać może także zadanie bibliotekarzy, współpracujących z pracownikami nauki w kolaboratoriach. Jest jednak wysoce prawdopodobne, że wraz z pojawianiem się coraz doskonalszych algorytmów, rosnąć będzie rola schematów służących automatycznemu tworzeniu metadanych w wyniku analizy nieustrukturyzowanego tekstu, natomiast metadane ustrukturyzowane, powstające dzięki rozumieniu zawartości obiektu (tekstu lub multimediów) będą tworzone w stosunkowo niewielkiej ilości w serwisach GBC, dostarczających informacji selekcyjonowanej przez pracowników informacji naukowej.

Interesującą, nową propozycją w zakresie „oddolnego” tworzenia metadanych przez twórców opisywanych obiektów cyfrowych, głównie stron Web w HTML/XHTML, są mikroformaty¹². Do kodu strony dodawane są standardowe elementy, określające zawartość dokumentu. Na przykład mikroformat hCard służy do oznaczania danych teledreślowych, Geo – pozycji geograficznej, co może być wykorzystane do wywołania mapy, hReview – opinii, recenzji, natomiast rel-tag – opisu rzeczowego. Mikroformaty można dodawać do istniejących stron, na przykład do informacji na blogu. Prowadzone są prace dla wykorzystania mikroformatów w Dublin Core [Méndez i in. 2008, s. 142].

Ręczne tworzenie metadanych dla wszystkich wartościowych zasobów GBC nie jest jednak możliwe. Dla funkcjonowania globalnej biblioteki cyfrowej duże znaczenie ma więc możliwość **automatycznego tworzenia metadanych** na podstawie treści dokumentów. Większość metadanych przed powstaniem Webu zestawiana była przez ludzi podczas kosztownych, bo pracochłonnych procesów katalogowania, realizowanych w bibliotekach. Później pojawiły się wyszukiwarki internetowe, które przeglądają i indeksują znaczną część stron internetowych, pozwalając na dostęp do nich po niewielkich kosztach¹³. O tym, że indeksy wyszukiwarek także mogą być uważane za zbiory metadanych już wspominałem wcześniej. Teraz omówię metody automatycznego wykrywania znaczenia tekstu, w celu tworzenia metadanych ustrukturyzowanych, na wzór inteligentnego katalogowania. Dają one także wyobrażenie o sposobach funkcjonowania algorytmów wyszukiwarek.

Dzięki zaawansowanej technologii przetwarzania języka naturalnego i wyszukiwania informacji, system informacyjny może pobierać metadane bezpośrednio z treści przeszukiwanych zasobów [Gawrysiak 2008b, s. 658]. Automatyczne wyodrębnianie (eksplora-

¹² <http://microformats.org/>.

¹³ Dostęp do informacji o stronach internetowych, dostarczanych przez wyszukiwarki globalne, jest z punktu widzenia użytkownika wyszukiwarki bezpłatny, gdyż twórcy wyszukiwarek uzyskują dochody w inny sposób (m. in. poprzez reklamy, specjalne pozycjonowanie), bez bezpośredniego pobierania opłat od użytkowników.

cja) informacji z tekstu, zwane po angielsku *text mining* lub *data mining*¹⁴, jest przedmiotem wielu aktualnie prowadzonych badań [Witten, Bainbridge 2003, s. 266], jako część zagadnień wydobywania wiedzy z baz danych (*knowledge discovery in databases*), polegającego na przetwarzaniu chaotycznych danych w uporządkowaną wiedzę [Chen, Chen 2007, s. 712]. Znaczenie automatycznego tworzenia metadanych w środowisku GBC, zawierającym ogromne ilości danych tekstowych, jest trudne do przecenienia. Obecnie narzędzia indeksujące generują streszczenia dokumentów, automatycznie tłumaczą teksty, dobierają słowa kluczowe i kategorie pochodzące z ontologii. Niestety, w tej dziedzinie trudno wskazać na jakąś podstawową i ogólnie przyjętą teorię, natomiast istniejące metody heurystyczne są skomplikowane, dostosowane do indywidualnych przypadków i trudne do wielokrotnego stosowania, a przez to do obiektywnej oceny.

Eksploatacja tekstu rozpoczyna się zwykle od stworzenia zestawu leksyki, czyli podzielenia tekstu na słowa kluczowe z wykorzystaniem stop-listy i oddzielenia tematów słów od przed- i przyrostków. Następnie stosowane są metody statystyczne, jak obliczanie częstotliwości występowania słów i grupowanie słów ze względu na tendencję do współwystępowania. Można także obliczać ich odległość semantyczną, co pozwala tworzyć grupy terminów bliskich znaczeniowo (*clusters*). W dalszej kolejności możliwe jest stosowanie niżej opisanych heurystyk.

Automatyczne tworzenie metadanych oparte jest obecnie na wcześniejszych osiągnięciach i zasadach automatycznego indeksowania [Andersen, Pérez-Carballo 2001, s. 232], polegającego na komputerowym wspomagananiu albo w pełni komputerowym indeksowaniu dokumentów. Pierwsze prace w tym zakresie polegały na badaniu możliwości tworzenia opisów treści: określaniu ich przedmiotów, wyodrębnianiu słów kluczowych i abstraktowaniu. Obecnie automatyczne tworzenie metadanych znajduje raczej zastosowanie w opisie formalnym, wyróżnianiu takich elementów opisu, jak autor, tytuł, data, format i inne [Greenberg, Spurgin, Crystal 2005, s. 3]. Prace te można podzielić na dwie grupy: badanie struktury tekstu dokumentu oraz stosowanie systemów organizacji wiedzy (SOW).

Badaniem struktury tekstu, w tym strukturą tematyczno-rematyczną wypowiedzenia, strukturą akapitu, zależnościami między miejscem wypowiedzi w tekście a ich wartością informacyjną, zajmuje się lingwistyka tekstu [Bojar 2009, s. 18]. Emma Tonkin i Henk Muller przedstawili pięć rodzajów struktur tekstów, które mogą być badane w celu pobierania metadanych:

- Dokumenty mogą posiadać strukturę wynikającą z ich formatu pliku, takiego jak XML, ale także XHTML, a nawet HTML, z którego można na przykład pobierać etykiety typu <title>, znajdujące się w nagłówku (sekcja HEAD). Języki kodowania, takie jak XML, narzucające wewnętrzną strukturę tekstom, ułatwiają „zrozumienie”, czy raczej wychwycenie wybranych, zwykle najważniejszych elementów treści dokumentu zarówno komputerom, jak i ludziom. Pozwalają one na zakodowanie wybranych elementów treści wprost, w taki sposób, aby mogły one być łatwo wydobyte przez analizę składniową struktury dokumentu.
- Dokument może posiadać strukturę służącą wizualizacji treści. Przykładem są dokumenty w formacie PostScript oraz PDF, w których określa się sposób rozmieszczenia tekstu na stronie, co może być wykorzystane do określenia jego struktury. Na przykład grupa liter umieszczona w obok siebie zapewne oznacza słowo, a grupa słów obok siebie na tym samym poziomie może oznaczać część zdania.
- Struktura tekstu dokumentu może wynikać z tradycji lub uzgodnionych zasad. Podstawowe metadane o dokumencie, szczególnie naukowym, takie jak jego tytuł, autor, wydawca, data wydania, słowa kluczowe czy abstrakt, mają swoje standardowe miejsce na

¹⁴ Dokładnie mówiąc, *text mining* polega na zastosowaniu metod *data mining* do zasobów tekstowych [Solka 2008, s. 96]. Można więc uważać eksploatację tekstu za szczególny przypadek eksploatacji danych.

stronie tytułowej i na początku tekstu, co ułatwia ich percepcję [Lawrence, Bollacker, Giles 1999]. Co więcej, forma ich prezentacji jest zazwyczaj ujednoczona. Podobne spostrzeżenia dotyczą cytowań bibliograficznych, umieszczanych zwykle na końcu tekstu. Wiele serwisów GBC stosuje heurystyki, służące do wykrywania tych i podobnych jednostek. Metody te można podzielić na:

– wykorzystujące preprogramowane heurystyki; po wstępnym oprogramowaniu heurystyki, do doskonalenia powstałych schematów wykorzystywana jest ludzka inteligencja, co pozwala na uwzględnienie wyjątków, występujących w języku naturalnym;

– gromadzące konwencje tekstu na podstawie ręcznie etykietowanych danych ćwiczebnych; systemy te zawierają zdefiniowaną strukturę, która może być adaptowana do bieżąco napotykaných wzorców tekstowych z uwzględnieniem parametrów oceny zgodnych z etykietowanymi dokumentami ćwiczebnymi;

– potrafiące samodzielnie podejmować działania adaptacyjne, wykorzystując dane nieetykietowane; raz wprowadzone techniki autoadaptacyjne mogą funkcjonować autonomicznie dla dużej liczby nieetykietowanych dokumentów.

• Wielu metadanych dostarczają analizy bibliometryczne cytowań lub współautorstwa. Analizy odnośników służą do określania poziomu wpływu (na wzór IF) i jakości dokumentu. Poziom połączeń pomiędzy dokumentami świadczy też o relacjach między nimi, w tym o podobieństwie zakresów.

• Dokumenty mogą być analizowane lingwistycznie. Na przykład odszukiwane mogą być wyrażenia typu „stwierdzono, że...” wskazujące na znaczenie dalszej części zdania. Analizy takie wymagają zarówno znajomości języka dokumentu, jak również dziedziny wiedzy. Najczęściej służą wyróżnianiu słów kluczowych i związków między cytowaniami. Do tej grupy zaliczyć można też probabilistyczne metody badania struktur gramatycznych [Tonkin, Muller 2008, s. 31].

Struktura dokumentu może być łatwo wyrażona i wykorzystana, jeśli mamy do czynienia z dobrze zdefiniowaną i silnie kontrolowaną grupą dokumentów. Dokumenty naukowe szczęśliwie należą do takiej grupy. W praktyce jednak kolekcje dokumentów zawierają zazwyczaj obiekty wyjątkowe (z punktu widzenia struktury); różnice te, mało istotne dla człowieka, dla heurystyk służących pobieraniu metadanych mogą być mylące.

Technologia cyfrowa spowodowała duży wzrost dostępności narzędzi organizacji i reprezentacji wiedzy (SOW – zob. p. 3.7), takich jak ontologie, tezaury, listy autorytatywne i inne. Wynikiem ich stosowania jest m.in. tworzenie rejestrów metadanych [Nahotko 2004, s. 116], służących udostępnianiu SOW oraz schematów metadanych. Stanowią one kolejne źródło zasobów do automatycznego tworzenia metadanych. Nazwy wyróżnione w dokumencie porównywane są z zawartością list autorytatywnych, takich jak na przykład Library of Congress Authority File. Stosowane są także metody łączące wykorzystanie technik przetwarzania języka naturalnego oraz słowników dziedzinyowych [Greenberg, Spurgin, Crystal 2005, s. 4].

Jak twierdzą James Anderson i José Pérez-Carballo, automatyczne indeksowanie dokumentów funkcjonuje już obecnie całkiem sprawnie i nie gorzej niż indeksowanie wykonywane przez człowieka, chociaż odmiennie [Anderson, Pérez-Carballo 2001, s. 236].

4.3. Wyszukiwanie informacji

Procesy wyszukiwania informacji mogą być opisywane z różnych punktów widzenia. Po pierwsze, można system wyszukiwania informacji traktować jako czarną skrzynkę, ocenianą według poziomu zaspokajania potrzeb użytkowników. W tym przypadku często przyjmuje się, że poszukiwanie informacji jest spowodowane koniecznością osiągnięcia celu przez wyszukującego [Wilson 2000, s. 49]. Można ten kierunek określić jako huma-

nistyczną koncepcję poszukiwania informacji, koncentrującą się na użytkowniku, zwaną po angielsku *information seeking* lub *information behavior* (stosowane jest również określenie *information seeking behavior*). Po drugie, można skupić się na procesach realizowanych wewnątrz systemu, prowadzących do przedstawienia wyników wyszukiwania w odpowiedzi na zapytanie. Jest to tradycyjna teoria wyszukiwania zorientowana na systemy (*information retrieval*) [Cisek 2007]. Oczywiście, oba punkty widzenia muszą być zbieżne, gdyż w innym przypadku użytkownicy i systemy pozostawałyby obok siebie, bez szans na pozytywny efekt w postaci pertinentnych wyników wyszukiwania.

Z punktu widzenia **zachowań użytkowników** wyszukiwanie informacji może być podzielone na cztery kategorie: pasywna uwaga, pasywne wyszukiwanie, aktywne wyszukiwanie i wyszukiwanie ciągłe. Uwaga pasywna uruchamiana jest podczas takich zachowań, jak słuchanie radia czy oglądanie telewizji. Możliwe jest przyswajanie informacji nawet, gdy jest to proces nieintencjonalny. Wyszukiwanie pasywne występuje, gdy informacja relewantna jest pozyskiwana w trakcie poszukiwania innej informacji lub nawet podczas wykonywania innych czynności. Wyszukiwanie aktywne, będące najczęściej przedmiotem badań, jest intencjonalnym poszukiwaniem informacji. Wyszukiwanie ciągłe ma miejsce, gdy wyszukiwanie aktywne przebiega w pewnych stałych ramach; realizowane jest periodycznie, w wyniku czego aktualizowane są wyniki wyszukiwania i wcześniej przyjęte zasady.

Potrzeby informacyjne użytkowników i sposoby ich zaspokajania mogą być opisywane w różny sposób, w zależności od przyjętego modelu. Jak pisze Maria Próchnicka, rozwój tego typu modeli wyszukiwania wiedzie od wąskich, odwzorowujących selekcję tekstów z bazy danych, które spełniają założone kryteria wyszukiwawcze (są to modele algorytmiczne), poprzez rozbudowane, uwzględniające aspekty konsytuacyjne (modele zadaniowe), po kontekstowe, wskazujące na rolę interakcyjności i wielopoziomowości procesu wyszukiwania informacji (modele kognitywne i konwersacyjne). Modele te różnią się głównie miejscem przyznawanym użytkownikowi w wyszukiwaniu informacji. W modelach algorytmicznych użytkownik uznawany jest za element względnie stabilny i dobrze zdefiniowany z punktu widzenia badania jego potrzeb informacyjnych. W modelach kognitywnych i konwersacyjnych użytkownik jest głównym elementem zmiennym i nieprzewidywalnym, powodującym niedookreśloność procesu wyszukiwania informacji [Próchnicka 2004, s. 25-26].

Wyszukiwanie informacji jest procesem powtarzalnym, z którym związane są zachowania aktywne i pasywne oraz komunikacja formalna i nieformalna. Cykl czynności polegających na tworzeniu, wykorzystywaniu i poszukiwaniu informacji może być postrzegany jako kolejne, wzajemnie powiązane etapy tego samego procesu. W trakcie zachowań informacyjnych ludzie „przemieszczają” się pomiędzy tymi etapami, jednocześnie tworząc, poszukując i wykorzystując informacje. Zazwyczaj wykonują wiele zadań związanych z zachowaniami informacyjnymi, z których każde może być realizowane w dowolnym czasie na innym etapie cyklu. W większości modeli procesów informacyjnych cykl inicjowany jest przez „potrzebę informacyjną” użytkownika [Belkin, Oddy, Brooks 1982; Dervin 1999]. Samo wyszukiwanie informacji natomiast może być traktowane jako szczególnie przypadek zachowań, związanych z rozwiązywaniem problemów (*problem-solving behaviour*) [Wilson 1999; Liu, Coleman 2004].

Potrzeby informacyjne są zawsze subiektywne, występują wyłącznie w umyśle poszukującego informacji, co utrudnia realizację badań opisanego procesu. „Potrzeba” jest terminem psychologicznym, podobnie jak związane z nim takie terminy, jak motywacje czy wartości. Pojęcie potrzeby jest niejednoznaczne; może być interpretowane jako stan niedoboru, proces motywacyjny lub obiektywnie istniejący warunek realizacji celu [Próchnicka 2004, s. 24]. Potrzeby nie mogą być bezpośrednio obserwowane przez badacza czy bibliotekarza, a tym bardziej przez programy komputerowe. Musimy więc poprzestać na

obserwacji wskaźników i zewnętrznych przejawów istniejących potrzeb. Paradoxem wyszukiwania informacji jest to, że użytkownik musi opisać swoje potrzeby (a więc dokumenty, które je zaspokajają poprzez dostarczanie odpowiedniej informacji), zanim jeszcze dotrze do odpowiednich źródeł. Opis taki jest utrudniony przez wiele czynników: niski poziom znajomości problemu przez użytkownika, nieznanostwo SIW lub języków, w których informacja jest dostępna.

Użytkownicy przejawiają zróżnicowane zachowania w procesie poszukiwania informacji [Fourie 2006, s. 23]. Niektóre są typowe dla poszczególnych osób czy panujących warunków; inne wydają się tworzyć stałe wzorce, które mogą stanowić podstawę doskonalenia projektowania serwisów GBC. Wykazują one zróżnicowanie w obrębie poszczególnych dyscyplin, przejawiające się na przykład odmiennymi wyborami kanałów komunikacji, co tłumaczy różnice pomiędzy dyscyplinami w stosowanych modelach publikowania elektronicznego.

Inną przyczyną różnicowania wzorców jest poziom zaangażowania w działalność naukową. W ciągu kolejnych lat kariery uczeni tworzą bogate sieci interpersonalne z kolegami o podobnych zainteresowaniach, nazywane niewidzialnymi uniwersytetami. Dzięki nim starsi pracownicy nauki uniezależniają się od katalogów, indeksów i spisów treści, oferowanych przez systemy informacyjne. Powstawanie repozytoriów preprintów zmienia nieco sytuację na korzyść młodszych pracowników nauki, jednak nie niwelują one braku rozbudowanych kontaktów interpersonalnych oraz kompleksowej, heurystycznej znajomości literatury przedmiotu. Taka sytuacja prowadzi do wniosku, że serwisy GBC powinny być tworzone z uwzględnieniem potrzeb młodych pracowników nauki [Borgman 2003a, s. 114].

Związki interpersonalne wpływają także na wzorce zachowań podczas wyszukiwania poprzez realizację tak zwanych wyszukiwań wspólnych (*collaborative information retrieval*). W takim modelu wyszukiwania informacji dochodzi nie tylko do interakcji pomiędzy systemem i użytkownikiem, ale także pomiędzy wieloma użytkownikami i systemem (systemami). Współpraca w zakresie wyszukiwania może być scharakteryzowana według dwóch kryteriów: poziomu synchronizacji i intencjonalności (rys. 7). Współpraca asynchroniczna oznacza, że użytkownicy nie pracują nad danym zagadnieniem jednocześnie; współpraca niejawną występuje, gdy system wykorzystuje informację z innego użycia systemu dla zastosowania w kolejnych wyszukiwaniach, ale bez gwarancji jednolitości celów wyszukiwania. Synchroniczne i jawne wyszukiwania prowadzone są przez małe grupy użytkowników, pracujących równocześnie dla osiągnięcia wspólnych celów. Mogą to być użytkownicy usytuowani w pewnej odległości od siebie, z których każdy dysponuje własnym komputerem lub też grupa użytkowników współużytkująca sprzęt komputerowy. Mamy tu więc jeszcze jedno kryterium podziału: kryterium przestrzenne, gdyż współpracujący użytkownicy mogą znajdować się w tym samym miejscu lub przebywać w nawet bardzo odległych lokalizacjach, połączeni siecią.

Tego typu wyszukiwanie informacji stwarza interesujące perspektywy dla zawodu bibliotekarza. Staje się on bowiem członkiem grupy wspólnie poszukującej potrzebnych informacji. Może uczestniczyć we wszystkich rodzajach wyszukiwania wspólnego, w różnych rolach, na przykład jako członek kolaboratorium, broker informacji, czy na jeszcze innych zasadach. Tego typu działalność bibliotekarze podejmowali i podejmują od dawna; opisuje to m.in. Lidia Derfert-Wolf jako *digital reference service* lub *virtual reference desk* [Derfert-Wolf 2006].

Kolejnym czynnikiem wpływającym na wzorce zachowań informacyjnych jest stopień dostępności źródeł informacji, zróżnicowany ze względu na odmienny poziom finansowania w różnych instytucjach i krajach. Bogatsze instytucje skupiają uczonych najwyższej klasy, jednocześnie zapewniając im znacznie lepszy dostęp do bieżącej literatury naukowej, spotkania podczas konferencji, możliwości publikowania. Do pewnego stopnia funk-

cjonowanie sieci komputerowych i GBC niweluje różnice w dostępie do niewidzialnych uniwersytetów i zwiększa dostępność literatury. Jednym z pośrednich efektów rozwoju GBC jest bowiem także wzrost zasobów Open Access i interaktywności wyszukiwania.

Synchroniczne	Wspólne wyszukiwanie informacji naukowej	Powiadamianie w czasie rzeczywistym. Stała, kontekstowa aktualizacja treści
	Grupowe przeglądanie asynchroniczne	Wspólne filtrowanie Wyszukiwanie społeczne Web 2.0 Inteligencja zbiorowa
	Jawne	Niejawne

Rys. 7. Rodzaje wyszukiwania wspólnego

Christine Borgman zaproponowała model opisujący niezbędną wiedzę i umiejętności, wykorzystywane podczas wyszukiwania informacji w serwisach GBC [Borgman 2003a, s. 133-140]. Zawiera on trzy elementy:

- Wiedza konceptualna, pozwalająca użytkownikowi na stworzenie modelu ułatwiającego zrozumienie systemu, z jakim ma do czynienia. Użytkownik stosuje wiedzę konceptualną w procesie wyszukiwania informacji w celu przetłumaczenia potrzeby informacyjnej na plan realizacji wyszukiwania oraz ewentualnej jego modyfikacji.
- Wiedza semantyczna i syntaktyczna służące bezpośredniej realizacji wyszukiwania. Pierwsza pozwala na zrozumienie operacji dostępnych w celu realizacji planu, umożliwiając na przykład wybór pomiędzy opcjami wyszukiwania. Druga służy zrozumieniu komend lub czynności wykonywanych w konkretnym systemie.
- Umiejętności techniczne, niezbędne do uzyskania wiedzy syntaktycznej, semantycznej oraz konceptualnej. Są nimi na przykład umiejętność korzystania z klawiatury komputerowej, znajomość sposobów prezentacji informacji na ekranie, czy znaczenia podstawowych klawiszy funkcyjnych. Zdobywanie tych umiejętności jest ułatwione dzięki funkcjonowaniu jednolitego GUI w Internecie.

Użytkownik wykorzystuje wiedzę i umiejętności, obejmujące wymienione trzy elementy, w sposób zależny od tego, czy korzysta z systemu wyszukiwania informacji, skształowanego w okresie przed powstaniem Webu, czy później. W pierwszym przypadku mamy do czynienia z takimi zasobami GBC, jak bazy danych, w tym komercyjne, OPAC biblioteczne i inne, zawierającymi głównie ustrukturyzowane dane, przechowywane w dobrze zorganizowanych systemach. Wyszukiwanie w Webie realizowane jest głównie przez wyszukiwarki internetowe [Mansourian 2004]. Nowe serwisy GBC łączą w sobie cechy obu rodzajów zasobów: z jednej strony mają określoną, bogatą strukturę danych, a z drugiej ich zasoby są indeksowane przez wyszukiwarki.

Dzięki integracji w GBC różnego rodzaju serwisów i ich narzędzi wyszukiwawczych możliwe są trzy rodzaje wyszukiwań, które Gary Marchionini nazywa sprawdzaniem (*lookup*), uczeniem się (*learn*) i badaniem (*investigate*), charakteryzowane przez następujące czynności:

- Sprawdzenie: wyszukiwanie faktów, znanych obiektów, nawigacja, weryfikacja, transakcja, odpowiedzi na zapytania.

- **Uczenie się:** gromadzenie wiedzy, rozumienie/interpretacja, porównywanie, agregacja/integracja, socjalizacja.
- **Badanie:** akumulacja, analiza, wykluczenie/negacja, synteza, ocena, odkrycie, planowanie/prognoza, transformacja [Marchionini 2006, s. 42].

Jest to bardzo ogólne spojrzenie na wyszukiwanie informacji, gdyż wymienione czynności zwykle wiązane są z procesem badawczym jako takim. Dwa ostatnie rodzaje Marchionini uważa za wyszukiwanie badawcze (*exploratory search*), w odróżnieniu od pierwszego, służącego jedynie zdobywaniu podstawowych informacji o faktach, lokalizowaniu ludzi, miejsc i obiektów. Wyszukuje się wówczas obiekty znane lub przewidywalne – wystarczy tylko znaleźć odpowiednią kombinację terminów wyszukiwawczych. W dalszej części procesu wyszukiwania użytkownik może uczyć się o nowych obiektach, relacjach międzyludzkich, miejscach i rzeczach. Zakłada się, że logiczną kontynuacją zachowań informacyjnych jest przejście użytkownika od sprawdzania do uczenia się. Badanie jest najtrudniejszą częścią procesów wyszukiwania. Odbywa się ono, gdy osoba posiadająca ogólną wiedzę w danej dziedzinie zamierza odszukać informacje na określony, szczegółowy temat.

Gdy zajmujemy się wyszukiwaniem danych z punktu widzenia **procesów technologicznych**, zachodzących w systemie komputerowym, zawsze musimy mieć na uwadze ich efektywność. Dlatego nie można pomijać faktu, że komputery większość czasu przeznaczono na wyszukiwanie używają oczekując na dostęp do pamięci masowych. Wynika z tego, że każda minimalizacja częstości dostępu do danych dyskowych powoduje znaczne przyspieszenie procesu wyszukiwania.

W celu ograniczenia potrzeby dostępu do dysku powstało kilka podstawowych technik wyszukiwawczych. Należą do nich: wyszukiwanie sekwencyjne, pliki inwersyjne (odwrócone) i obiekty typu tablic mieszania lub haszowania (*hash tables*).

Wyszukiwanie sekwencyjne (pełnotekstowe) rozpoczyna się na początku tekstu, następnie tekst jest odczytywany znak po znaku w poszukiwaniu zadanych wartości (ciągu znaków) i kończy się na końcu tekstu. Wynika z tego, że przetwarzany jest cały plik. Pełne skanowanie tekstu jest czasochłonne, ale można go zastosować do wyszukiwania nawet pojedynczych ciągów znaków. Często stosuje się go do wyszukiwania przy użyciu „wyrażeń regularnych”, które są ograniczonymi rozszerzeniami ciągów. Najprostszą formą wyrażeń regularnych są tak zwane wieloznaczniki¹⁵ (*wildcard*), którymi mogą być różne znaki, w zależności od rodzaju wieloznacznika.

Procedury wyszukiwania ciągów znaków wykorzystywane są w edytorach tekstu lub podczas przeglądania dokumentów tekstowych, w przeszukiwaniach baz danych oraz w pracy z danymi na poziomie systemu operacyjnego (na przykład przy wyszukiwaniu plików, procesów). Bardziej zaawansowane programy pozwalają na stosowanie wyrażeń regularnych, korzystających ze znacznie szerszego zbioru metaznaków o większych możliwościach od zwykłych wieloznaczników.

Wszystkie algorytmy wyszukiwania sekwencyjnego mają pewne cechy wspólne. Podczas ich wykonywania nie ma potrzeby zapewnienia dodatkowego miejsca na dysku, oprócz tego, które zajmuje sam przeszukiwany plik (wymagana jest przestrzeń pamięci wielkości zapytania, a nie pliku). Przeszukiwane pliki mogą znajdować się w postaci, w jakiej zostały wykonane, nie ma potrzeby prowadzenia żadnych prac przygotowawczych. Wyszukiwanie może się odbywać według złożonych wyrażeń, nie tylko ciągów.

Czas pracy wszystkich algorytmów wyszukiwania sekwencyjnego jest proporcjonalny do długości przeszukiwanego zasobu. Jeżeli przeszukiwany plik podwoi swoją objętość,

¹⁵ Wieloznacznik zwany jest też symbolem maski, znakiem globalnym, metaznakiem, symbolem wieloznacznym lub znakiem wieloznacznym. Wyróżnia się zwykle dwa podstawowe rodzaje: wieloznacznik lokalny (?), zastępujący pojedyncze wystąpienie znaku i wieloznacznik ogólny (*), zastępujący dowolną liczbę znaków.

to wyszukiwanie potrwa dwa razy dłużej. Dlatego wyszukiwanie takie można stosować do stosunkowo niedużych zasobów, nie zdaje więc egzaminu w przypadku tekstowych baz danych, zawierających miliardy znaków. W takich przypadkach stosowane są techniki stosujące **pliki odwrócone** (*inverted files*). Można je porównać do indeksu w książce: elementy służące wyszukiwaniu wybierane są z zasobu i odpowiednio sortowane (na przykład alfabetycznie), dzięki czemu są znacznie szybciej dostępne do odszukania. Czynność tę, zwaną indeksowaniem, mogą wykonywać ludzie (głównie bibliotekarze i pracownicy informacji) lub może ona być wykonywana automatycznie. Podczas automatycznego indeksowania dokumentów elektronicznych tekst „cięty” jest na poszczególne słowa, które przenoszone są do oddzielnego pliku wraz z adresem miejsca, w którym słowo znajduje się w pliku pierwotnym. Słowa są następnie sortowane alfabetycznie, co znacznie przyspiesza wyszukiwanie.

Pliki odwrócone są techniką obecnie powszechnie stosowaną dla wielkich baz danych. Nie ma innych sposobów przeszukiwania giga- i terabajtowych baz danych z szybkością możliwą do przyjęcia. Podczas indeksowania tekstów kodowanych, na przykład w XML, jest możliwość określenia przez system funkcji każdej części indeksowanego pliku. Mają one jednak także swoje wady. Nie ma możliwości wyszukiwania według dowolnego wyrażenia. Zazwyczaj systemy z plikami odwróconymi mogą wyszukiwać tylko w oparciu o początek słowa. Jeżeli chcemy mieć możliwość wyszukiwania według końca słów, a nie początku, należy stworzyć dodatkowy plik odwrócony, posortowany według końców słów. Oczywiście plik odwrócony (lub pliki) muszą być wykonane zanim wyszukiwanie może być przeprowadzone, co może być długotrwałym procesem. Po każdej zmianie pliku danych modyfikacji wymaga także plik odwrócony. Pliki odwrócone zajmują dużo miejsca, od 25% aż do 200% więcej niż same pliki danych. Ich rozmiary można zmniejszyć stosując tzw. stop-listy. Ze względu na coraz niższe koszty pamięci operacyjnych i zewnętrznych, ilość dodatkowo zajmowanej pamięci przestaje być problemem.

Inną formą wyszukiwania, która nie wymaga przeszukiwania sekwencyjnego całego zasobu danych, jest technika stosowania **tablic mieszania** (*hash*) lub kodowania mieszanego. Dla danego słowa obliczane jest jego miejsce występowania w pliku. Gdyby w danym języku każda litera występowała z taką samą częstotliwością, można by stwierdzić, że słowa zaczynające się na literę M powinny być w połowie pliku odwróconego. W rzeczywistości tak nie jest. Rozwiązaniem jest obliczenie funkcji hash dla każdego słowa. Funkcja ta podobna jest do generatora liczb losowych: jest próbą utworzenia wartości, których każda możliwa wartość funkcji występuje z podobnym prawdopodobieństwem. Zawsze dla tego samego słowa funkcja wygeneruje ten sam kod.

Ponieważ zadaniem algorytmu mieszania jest rozmieszczenie słów możliwie przypadkowo, słowa podobne, na przykład o tym samym rdzeniu, a różniące się tylko przyrostkiem, mogą znaleźć się daleko od siebie. Oznacza to, że nie nadaje się on do poszukiwania podobnych słów, co często jest potrzebne w zastosowaniach serwisów GBC. Podobnie jak dla plików odwróconych, niezbędne jest wcześniejsze przygotowanie odpowiedniego pliku i jego stałe modyfikacje.

Oprócz przedstawionych, stosowane są także inne techniki zapewnienia szybkiego dostępu do zasobów baz danych przez właściwą organizację plików, takie jak drzewa lub pliki sygnaturowe. Nie mają one jednak większego zastosowania w zakresie masowych danych tekstowych.

Obecnie stosowane są odrębne mechanizmy wyszukiwania informacji dla poszczególnych mediów i zastosowań. Metody wyszukiwania zapisów dźwięku w jednym serwisie GBC są trudne do zastosowania w innym, a jeszcze mniej przydatne do wyszukiwania nagrań wideo czy obiektów graficznych. Zapisy dźwięków mogą być opisywane tekstowo lub za pomocą innych dźwięków; grafika, w tym trykówmiarowa, może być również

opisywana tekstowo, przez narysowanie podobnego obiektu, a także przez przedstawienie lub wybór przykładu. Techniki wyszukiwania, specyficzne dla danego nośnika czy zastosowania, mogą być łączone z wyszukiwaniem tekstowym. Rozwiązaniem w tym zakresie może być rozwój technologii tzw. wyszukiwania uniwersalnego.

W dalszej części rozdziału przedstawię najważniejsze serwisy GBC służące wyszukiwaniu informacji, zarówno wyszukiwarki internetowe, jak i serwisy stosujące metadane ustrukturyzowane, o kontrolowanej jakości.

Wbrew niektórym, potoczny opinii, biblioteki naukowe są jednymi z najbardziej innowacyjnych instytucji, przynajmniej w zakresie wdrażania technologii komputerowych i systemów wyszukiwania informacji. Nie dziwi więc, że jednym z pierwszych zastosowań systemów komputerowych było ich wykorzystanie w systemach informacyjno-wyszukiwawczych. Systemy te od początku można było podzielić na dwa rodzaje: **katalogi biblioteczne (OPAC) oraz bibliograficzne bazy danych**, w których znajdują się metadane tworzone przez człowieka (zob. p. 2.8). Później, wraz z rozwojem Internetu można było dodać trzeci rodzaj zasobów, mianowicie różnego typu portale internetowe, służące porządkowaniu i selekcji jakościowej zasobów internetowych. Każde z tych zastosowań rozwijało się niezależnie, co w efekcie dało produkty o różnych funkcjach [Baeza-Yates, Ribeiro-Neto 1999, s. 397]. Obecnie oprócz wymienionych powstają także serwisy pełnotekstowe, umożliwiające wyszukiwanie informacji według ustrukturyzowanych metadanych. Chodzi głównie o biblioteki cyfrowe i repozytoria różnego rodzaju obiektów cyfrowych, stanowiące przejaw publikowania cyfrowego (i jako takie zostały przedstawione w p. 2.6). Wszystkie te zasoby wraz z ich narzędziami wyszukiwawczymi stanowią serwisy GBC oferujące **informację o określonej jakości**.

Pierwsze systemy bibliograficznych baz danych tworzone były przez ośrodki rządowe, jako narzędzia wspomagające rozwój nauki i techniki. Ich bezpośrednimi użytkownikami mieli być wyszkoleni pośrednicy (pracownicy informacyjni), współpracujący z użytkownikami końcowymi. OPAC był natomiast projektowany do stosowania bezpośrednio przez użytkownika końcowego, co miało istotny wpływ na powstawanie różnic w sposobach organizacji informacji i interfejsów obu grup serwisów. Odrębny rozwój i inne cele spowodowały odmienności obu rodzajów systemów wyszukiwania informacji, co do charakteru dostarczanych usług i możliwości wykorzystania, przez co właściwe jest ich oddzielne traktowanie.

Ważne zjawisko ostatnich lat, serwisy GBC o kontrolowanej jakości, takie jak portale internetowe, biblioteki cyfrowe czy repozytoria, przyczyniają się do większego ujednolicenia sposobów wyszukiwania informacji i stałego zmniejszania różnic pomiędzy tworzeniem i udostępnianiem metadanych o zasobach dostępnych lokalnie i zdalnie. Dawne priorytety zastąpione zostały przez nowe, na przykład jednolitość opisów w wyodrębnionym środowisku, takim jak biblioteczny OPAC, jest mniej istotna niż zdolność do zapewnienia współpracy (współdziałania) pomiędzy środowiskami wielu serwisów GBC: komercyjnym i, instytucji bezpłatnie udostępniających swoje usługi i społecznościowym.

Bibliograficzne i pełnotekstowe bazy danych

Idea bibliograficznych baz danych wywodzi się z czasów, gdy nie funkcjonowały pełne teksty dokumentów w wersji cyfrowej, tworzono więc elektroniczne surogaty zasobów drukowanych (głównie artykułów z czasopism), ułatwiające ich wyszukiwanie, przekazywanie informacji i udostępnianie. Ponieważ wciąż duża część publikacji nie jest dostępna w formie elektronicznej, rekord w bazie danych stanowi ich jedyną reprezentację cyfrową, przynajmniej do czasu zeskanowania oryginału. Rekordy te są z dużym nakładem pracy wykonywane przez wyspecjalizowanych pracowników i zawierają podstawowe metadane o obiekcie fizycznym, zarówno o jego cechach formalnych, jak i treści. Można więc uznać, że każde słowo (z wyjątkiem słów odrzuconych, umieszczanych na stop-liście) zawarte w opisie ma istotne znaczenie i niesie ważną informację o opisywanym obiekcie.

W zakresie tworzenia i udostępniania baz danych współpracują ze sobą twórcy i dostawcy tych baz, chociaż ze względu na postępujące wykupywanie firm na tym rynku coraz częściej pojawiają się przedsiębiorstwa zajmujące się zróżnicowaną działalnością informacyjną. Ogólnie można powiedzieć, że twórcy baz danych przygotowują produkt, a licencję na jego udostępnianie przekazują dostawcom, którzy dostarczają oprogramowanie służące wyszukiwaniu i samą bazę danych (wraz z wieloma innymi) swoim klientom, otrzymującym możliwość wyszukiwania w wielu bazach danych w ujednolicony sposób i w jednym serwisie.

Tworzenie baz danych często określa się mianem przemysłu baz danych, gdyż ich produkcja prowadzona jest przez korporacje i organizacje, w tym rządowe, jako działalność komercyjna lub przynajmniej przynosząca zwrot kosztów. Bazy danych są traktowane jako produkt na sprzedaż lub wynajem, często bibliotekom i innym organizacjom – na przykład wspomnianym dostawcom. Zadaniem dostawców baz danych jest zadbanie o licencjonowany dostęp do baz danych oraz wzbogacenie produktu o swoje usługi, polegające przede wszystkim na udostępnianiu zasobów użytkownikom, ale także rozwijaniu standardyacji formatów danych, tworzeniu indeksów (zwykle jako plików odwróconych) oraz przygotowaniu jednolitego interfejsu służącego wyszukiwaniu w wielu bazach danych. Największe serwisy oferują całe zestawy baz danych, pokrywających swoim zakresem wszystkie dyscypliny¹⁶. Dobrze znanymi dostawcami takich zestawów baz danych są na przykład Thomson, LexisNexis¹⁷, OCLC¹⁸ czy H.W. Wilson¹⁹. Czasem producenci baz danych zajmują się jednocześnie ich dystrybucją, jak ma to miejsce w przypadku firmy H.W. Wilson.

Funkcjonowanie systemów wyszukiwania informacji w bazach danych, dostępnych online, zazwyczaj oparte jest na boole'owskim modelu wyszukiwania informacji. Wybór ten podyktowany został kilkoma czynnikami. Algebra Boole'a stosowana była już wcześniej w systemach wyszukiwania informacji, w tym związanych z małą mechanizacją, natomiast badania nad innymi modelami wyszukiwania informacji dopiero się rozpoczynały. Dużą rolę odegrały istniejące wówczas ograniczenia techniki komputerowej; na ówczesnym sprzęcie można było stosować techniki oparte na plikach odwróconych i operatorach logicznych, w odróżnieniu od bardziej zaawansowanych metod.

Pomimo rozwoju innych sposobów organizacji wyszukiwania, mogących je przepięścić (związanych na przykład z modelem przestrzeni wektorowych), wyszukiwanie boole'owskie pozostało najczęściej stosowaną metodą, oferowaną przez dostawców baz danych, chociaż część z nich proponuje inne rozwiązania, na przykład umożliwia wprowadzanie zapytań w języku naturalnym (zob. interfejs Medline na rys. 8), w połączeniu z przetwarzaniem wyników wyszukiwania przez algorytmy rankingowe. Przyczyną pozostawania przy algebrze Boole'a były głównie finansowe (koszty tworzenia nowych systemów), społeczne (istnienie dużej grupy użytkowników przyzwyczajonych do tradycyjnych metod), brakowało także wyraźnych dowodów na istotne polepszenie procesu wyszukiwania z zastosowaniem nowych metod, gdyż w środowisku danych silnie ustrukturyzowanych stare metody poprawnie spełniają swoją rolę. Zazwyczaj dostawca bazy danych stosuje własne, wyspecjalizowane oprogramowanie wyszukiwawcze, przez co pracownicy informacji naukowej pośredniczący w wyszukiwaniach lub sami użytkownicy, muszą nauczyć się pracy z każdym z nich z osobna.

Coraz częściej bibliograficzne bazy danych przekształcają się w zasoby pełnotekstowe, gdzie od opisu bibliograficznego prowadzi odnośnik na przykład na stronę czasopisma²⁰.

¹⁶ Przykładem może być oferta firmy Thomson, zawierająca m.in. dostęp do 1.200 baz danych, w tym do serwisu Dialog. <http://support.dialog.com/publications/dbcat/>.

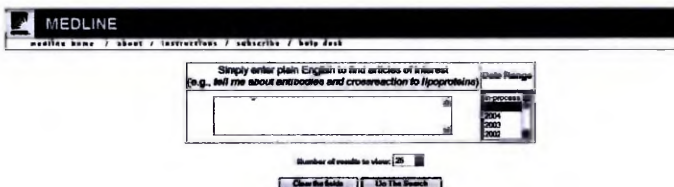
¹⁷ <http://www.lexisnexis.com/>.

¹⁸ <http://www.oclc.org/>.

¹⁹ <http://www.hwwilson.com/>.

²⁰ Serwisy takie funkcjonują m.in. na stronach Web większości wydawców czasopism elektronicznych.

Umożliwia to stosowanie technik wyszukiwania informacji operujących na całym tekście dokumentu, nie tylko rekordzie bibliograficznym. Techniki te obejmują określenie bliskości słów w tekście (na przykład współwystępowanie w odległości nie większej niż pięć słów, w tym samym zdaniu, w tym samym akapicie), określanie częstości wystąpień słów w tekście. Wyszukiwanie może być prowadzone manualnie lub automatycznie; oba rozwiązania mają swoje zalety i wady. Wyszukiwanie manualne pozwala na większą kontrolę procesu wyszukiwania, ale wymaga lepszej znajomości semantycznych i syntaktycznych zasad wyszukiwania, a także specyfiki poszczególnych serwisów GBC. Wyszukiwanie automatyczne nie wymaga posiadania takiej wiedzy, ale może dawać wyniki mniej efektywne.



Rys. 8. Interfejs systemu Medline – zastosowanie języka naturalnego i logiki rozmytej

Bazy danych, zawierające opisy bibliograficzne, przekształcają się także we wspomniane wcześniej indeksy cytowań. Najstarszymi bazami danych tego typu są publikowane pierwotnie przez ISI (Institute for Scientific Information), a obecnie przez Thomson Scientific indeksy cytowań artykułów z czasopism: Science Citation Index, Social Science Citation Index oraz Arts and Humanities Citation Index. Od 1997 r. bazy te dostępne są online jako serwis Web of Science²¹. Najcenniejszą cechą tego typu baz danych jest możliwość docierania do publikacji cytowanych przez dany artykuł oraz określenie dla danego artykułu publikacji go cytujących. W ten sposób istnieje możliwość wyszukiwania publikacji związanych z określonym tematem chronologicznie wstecz i wprzód w stosunku do danej, odszukanej publikacji. Twórcy Web of Science zignorowali jednak fakt, że uczeni coraz częściej umieszczają swoje artykuły w różnych zasobach Internetu, gdzie dostępne są bezpłatnie. Rozwój Webu spowodował, że powstawać zaczęły nowe serwisy, które złamały monopol i hegemonię Web of Science.

Obecnie w GBC funkcjonuje ponad 100 baz danych i innych narzędzi służących do wyszukiwania cytowań. Można je podzielić na trzy kategorie:

- Bazy danych, które umożliwiają użytkownikowi wyszukiwanie pełnotekstowe dla wykrycia cytowań określonych autorów, czasopism lub artykułów. Wśród tych narzędzi wymienić można najbardziej znane: arXiv, CiteSeer, RePEc, getCITED²², Google Books i Google Scholar oraz Scirus Elseviera. Część z tych serwisów, jak CiteSeer i Google Scholar, funkcjonuje jako autonomiczne indeksy cytowań, umożliwiające automatyczny wybór i grupowanie cytowań prac naukowych, dostępnych online.
- Bazy danych pozwalające użytkownikom na wyszukiwanie w bibliografii załącznikowej. Serwisy takie, tworzone często przez wydawców, rozpoczęły działalność pod koniec lat dziewięćdziesiątych XX wieku, gdy dziedziczne bazy danych uzupełnione zostały o bibliografie załącznikowe. Przykładowo wymienić można: MathSciNet²³

²¹ <http://scientific.thomson.com/products/wos/>.

²² getCITED to wielodziedziczna baza danych zawierająca opisy publikacji wszystkich typów (<http://www.getcited.org/>).

²³ <http://www.ams.org/mathscinet/>.

z AMS, ScienceDirect Elsewiera, SciFinder Scholar²⁴ z ACS, Scitation²⁵ z AIP czy Spire-HEP²⁶ ze Stanford Univ.

- Bazy danych funkcjonujące dokładnie tak, jak Web of Science. Tu wymienić należy Scopus²⁷, utworzony w 2004 r. przez Elsevier, w którym także podawane są IF czasopism. Ten ostatni przewyższa produkt Thomsona liczbą opisów, lecz obejmuje piśmiennictwo od 1996 r., podczas gdy Web of Science już od 1900 r. [Meho 2007, s. 30].

Online Public Access Catalog (OPAC)

Katalogi biblioteczne służą jako źródło informacji o zasobach biblioteki, będąc narzędziem umożliwiającym przeszukiwanie ich kolekcji. Przyjmowały one różne formy, od katalogów książkowych, poprzez kartkowe, mikrofilmowe i mikrofiszowe. Katalogi online pojawiły się w latach siedemdziesiątych ubiegłego wieku, lecz pierwsze rozwiązania miały bardzo ograniczoną funkcjonalność i stanowiły uzupełnienie modułu udostępniania, który był początkiem obecnie funkcjonujących zintegrowanych systemów bibliotecznych. Współczesne OPAC używane są od lat osiemdziesiątych XX w. Stanowią rodzaj bibliograficznej bazy danych, choć od baz danych opisanych wcześniej różnią się pod wieloma względami.

Rozwój OPAC, podobnie jak rozwój zintegrowanych systemów bibliotecznych, których OPAC jest częścią, związany jest z rozwojem techniki komputerowej. Wielu autorów (np. [Hildreth 2000; Taylor 2003]) dzieli okres rozwoju na kolejne generacje, takie jak:

- Katalogi pierwszej generacji, tworzone na początku lat osiemdziesiątych XX w. służyły głównie wyszukiwaniu znanych już wcześniej pozycji według ograniczonej liczby kryteriów, takich jak autor, tytuł, sygnatura. Nie było żadnych elementów opisu rzeczowego. Brak było możliwości przeglądania (*browsing*) opisów. Opisy były krótkie i na ogół niestandardowe. Twórcy pierwszych OPAC starali się często upodobnić je funkcjonalnie, a nawet wizualnie, do katalogów kartkowych. Niektóre systemy powiały rozwiązania stosowane w systemach baz danych, głównie serwisu Dialog.
- Druga generacja, która pojawiła się z końcem lat osiemdziesiątych przyniosła wzrost możliwości wyszukiwawczych poprzez dodanie takich kryteriów, jak hasła przedmiotowe (wraz z możliwością przeglądania słownika) lub słowa kluczowe oraz mechanizmów, jak algebra Boole'a. Część katalogów umożliwiała wybór formatu wyświetlenia danych (na przykład krótki, średni, pełny), wspomaganie wykorzystania (na przykład różnie zaprojektowany dialog dla użytkownika nowego i doświadczonego, więcej rozbudowanych komunikatów o błędach). Katalogi stały się dostępne poprzez telnet. Połączenie OPAC z modułem udostępniania umożliwiło poznanie statusu udostępniania wyszukanych materiałów. Wadami były częste błędy w wyszukiwaniu, kłopoty ze stosowaniem słownika JIW oraz słaba organizacja dużych zbiorów odpowiedzi.
- Trzecia generacja OPAC, funkcjonująca od 1996 r., miała zapobiegać wcześniejszym problemom dzięki zastosowaniu takich technik i narzędzi, jak wspomaganie tworzenia strategii wyszukiwawczej, integracja słowników kontrolowanych i wyszukiwanie pełnotekstowe, równoległe wyszukiwanie w wielu bazach danych.
- Czwarta generacja charakteryzuje się zastosowaniem wielu narzędzi znanych z Webu, na przykład graficznego interfejsu użytkownika (GUI), udostępnianiem zasobów przy pomocy protokołu Z39.50, wykorzystaniem hipertekstu a także możliwością przetwarzania danych w różnych formatach (na przykład MARC i Dublin Core). Katalogi zapewniają możliwość dostępu poprzez odnośniki do multimediów oraz pełnych tekstów

²⁴ <http://www.cas.org/SCIFINDER/SCHOLAR/>.

²⁵ <http://scitation.aip.org/>.

²⁶ <http://www.slac.stanford.edu/spires/>.

²⁷ <http://www.scopus.com/scopus/home.url>.

dokumentów, a także zewnętrznych baz danych. Użytkownicy, na wzór wyszukiwarek internetowych, mają do dyspozycji wyszukiwanie proste i złożone. Wiele katalogów pozwala na rozbudowane możliwości pomocy, na przykład w formie czatów (*chat*).

- Jako kolejną generację OPAC traktuje się tzw. *Social OPAC*, w skrócie SOPAC, pozwalający na integrację bibliotecznego OPAC z systemami CMS²⁸ typu Drupal²⁹, co umożliwia użytkownikom dodawanie etykiet, punktowanie i opiniowanie materiałów w zbiorach bibliotecznych. Dane wprowadzane przez użytkowników są wyszukiwalne i dostępne dla wszystkich, więc SOPAC nazwać można narzędziem społecznościowym. Dodatkowo użytkownicy mogą kustomizować interfejs katalogu, zachowywać swoje wyszukiwania, zamawiać, prolongować pozycje i uiszczać opłaty. RSS informuje o nowych nabytkach. Część bibliotek, wzorem Amazon dodaje do opisów widok okładki i urywki treści. Przykładem jest Chamo firmy VTLS³⁰.

W praktyce, jako serwisy GBC, wciąż funkcjonują OPAC wszystkich generacji. Spotykane są we wszystkich rodzajach bibliotek, w bibliotekach naukowych stanowiąc standard. Biblioteka tradycyjna, wraz z udostępnieniem w Webie metadanych o swoich zasobach, staje się serwisem GBC, szczególnie, gdy dane te są dostępne dla wyszukiwarek internetowych oraz funkcjonują w katalogach centralnych, stając się częścią zasobów globalnych. Stanowią one łącznik pomiędzy zbiorami tradycyjnymi, przechowywanymi w bibliotekach, a kolekcjami obiektów cyfrowych, gromadzonymi w serwisach GBC.

Wraz z rozpowszechnieniem OPAC nie poszedł rozwój umiejętności ich użytkowników, tym bardziej, że oczekują oni łatwości i użyteczności porównywalnej z wyszukiwarkami internetowymi. Tymczasem struktury stosowane w OPAC, jak MARC czy LCSH są skomplikowane i trudne do użycia intuicyjnego. Użytkownicy popełniają wiele błędów, których efektem są dwie przeciwstawne sytuacje: albo otrzymują pusty zbiór wyników, albo liczba odpowiedzi uniemożliwia ich wykorzystanie. To szczególnie frustruje użytkowników współczesnych wyszukiwarek internetowych, gdzie bardzo rzadko występuje brak odpowiedzi, a opisy najbardziej użytecznych obiektów najczęściej znajdują się zaraz na początku listy odpowiedzi.

OPAC jest „czarną skrzynką” dla użytkowników, którzy niewiele wiedzą o procesach zachodzących wewnątrz. Nie jest im zazwyczaj znany związek pomiędzy trafnością i kompletnością wyszukiwania. Trudności sprawia także logika algebry Boole’a, której wyrażenia mają inne znaczenie, niż stosowane potocznie. W efekcie użytkownicy starają się budować jak najprostsze wyrażenia wyszukiwawcze. Do tego typu postępowania dostosowały się wyszukiwarki internetowe, których twórcy bardzo dbają, aby użytkownik nie dowiedział się, że do wyszukiwania potrzebna jest znajomość jakiegokolwiek algebry.

Twórcy bibliotecznych OPAC wychodzili z założenia, że użytkownicy winni przychodzić do biblioteki przygotowani – powinni przed rozpoczęciem wyszukiwania znać nazwisko autora, tytuł i/lub szczegółowy zakres swoich zainteresowań. W wielu systemach bibliotecznych opis katalogowy wyświetlany jest w taki sam sposób, w jaki przedstawiany był na karcie katalogowej. Nawet, jeśli umożliwia się wiele sposobów wyświetlania danych, to często nie uwzględniają one możliwości środowiska sieciowego. Katalogi są więc nastawione bardziej na dane niż użytkownika, a przez to rozumiane i chętnie stosowane przez bibliotekarzy, a nie użytkowników. Taki katalog może być przedmiotem podziwu, ale niekoniecznie praktycznego stosowania.

²⁸ CMS – Content Management System, System Zarządzania Treścią to system internetowy pozwalający na łatwe tworzenie i zarządzanie treściami w Sieci, w tym ich udostępnianie. Serwisy Web mogą być tworzone za pomocą prostych interfejsów użytkownika, często zawierających formularze. CMS pozwalają oddzielić treść od wyglądu, układu strony. Przedmiotem zarządzania treścią mogą być obiekty cyfrowe dowolnego rodzaju [Firlej-Buzon 2005, s. 163].

²⁹ <http://drupal.org/>.

³⁰ <http://www.vtls.com/pressrelease/VTLS-Introduces-Chamo,-a-New-Social-OPAC-with-Drupal-Support-62>

Klienci bibliotek naukowych, oceniający je poprzez sposób działania OPAC, posiadają obecnie zupełnie inne umiejętności, niż jeszcze kilkanaście lat temu, gdy powstawały modele współczesnych katalogów bibliotecznych. Żyją oni w stałym dostępie do Sieci, przyzwyczajeni do interaktywnego korzystania z jej zasobów i zaspokajania swoich potrzeb informacyjnych w Web przy pomocy wyszukiwarek [Coyle, Hillmann 2007]. OPAC traktowany jest jak każdy inny zasób GBC; jej użytkownicy przestają być zainteresowani katalogami bibliotecznymi nie nadążającymi za tymi zmianami, co powoduje nieporozumienia między bibliotekarzami a ich użytkownikami; ci drudzy uważają biblioteki za miejsca o przestarzałych technologiach, a ci pierwsi dziwią się lenistwu swoich użytkowników i niewielkim efektem, które dają przeprowadzane z takim wysiłkiem szkolenia biblioteczne³¹.

Mechanizmy wyszukiwawcze tradycyjnego OPAC różnią się wyraźnie od oferowanych przez wyszukiwarki. Te ostatnie oferują takie funkcje, jak: zapytania w języku naturalnym, automatyczne powiązania ze słownikiem kontrolowanym, kontrolę pisowni, dostarczanie stron podobnych, wyniki sortowane według relewancji, wskaźniki popularności i przeglądanie [Yu, Young 2004, s. 194]. Funkcje te niewątpliwie wpływają na oczekiwania użytkowników w stosunku do możliwości wyszukiwawczych OPAC. Spodziewają się oni, że w OPAC, podobnie jak w Internecie, będą mogli pracować samodzielnie, bez potrzeby wcześniejszego odbywania szkoleń oraz w każdym miejscu, gdzie znajduje się komputer podpięty do sieci. W związku z tym powstają propozycje, takie jak artykułowane przez Bradleya Westbrooka, który twierdzi, że nie wszyscy użytkownicy serwisów GBC posiadają takie same potrzeby; nie każdy użytkownik chce i może zapoznać się z rozbudowanymi pomocami wyszukiwawczymi (*finding aids*) lub skomplikowanymi interfejsami i interpretować skomplikowaną organizację informacji. Proponuje pozwolić użytkownikowi na dostęp do obiektów cyfrowych, powiązanych z innymi obiektami, bez potrzeby korzystania z tych narzędzi [Westbrook 2002, s. 79].

Współczesne, zintegrowane systemy biblioteczne nie nadążają za zmieniającymi się potrzebami. Zaimplementowane nieelastyczne opcje wyszukiwawcze utrudniają wykorzystanie katalogu bibliotecznego. Co prawda wielu producentów tego oprogramowania stale zwiększa funkcjonalność sprzedawanych przez siebie systemów, jednak średni ich poziom jest niewystarczający. Jak stwierdziła Kristin Antelman, wszyscy poważni dostawcy zintegrowanych systemów wciąż oferują katalogi drugiej generacji. Poza umożliwieniem tworzenia odnośników pomiędzy rekordami, na co pozwala środowisko Webu, podstawowe struktury danych i zasady wyszukiwania z użyciem algebry Boole'a pozostają niezmienione. Nie można dłużej twierdzić, że bardziej złożone wyszukiwanie jest zbyt kosztowne z punktu widzenia mocy obliczeniowych. Mogą one być raczej kosztowne z punktu widzenia interesów biznesowych producentów [Antelman, Lynema, Pace 2006]. Ze względu na te trudności biblioteki naukowe same próbują wdrażać nowe technologie, tworząc ich aplikacje samodzielnie lub wraz z firmami produkującymi oprogramowanie biblioteczne³². Jedną z inicjatyw jest Catalog Enrichment Initiative³³, mająca za cel uatrakcyjnienie bibliotecznego OPAC przez dodanie do niego takich elementów, jak spisy treści, indeksy, wizerunek okładki, recenzje, streszczenia, urywki treści. Takie podejście zbliża OPAC do tzw. Amazoogle [Swoboda 2009, s. 18], o którym mowa była w p. 1.4.

³¹ W obronie użytkowników staje Roy Tennant formułując wiele błyskotliwych haseł typu: „tylko bibliotekarze lubią wyszukiwać, wszyscy inni chcą znajdować” lub „użytkownicy nie są leniwi, są po prostu ludźmi” [Tennant 2004].

³² Przykładami najbardziej znanych rozwiązań służących wyszukiwaniu informacji są: Endeca (<http://endeca.com/>), Vivisimo (<http://vivisimo.com/>), Aquabrowser (<http://www.medialab.nl/>), Nowe technologie stosują tacy dostawcy oprogramowania bibliotecznego, jak Innovative (OPAC Pro, Encore <http://www.iii.com/>), Sirsi/Dynix (EPS, <http://www.sirsidyndix.com/Solutions/Products/portalsearch.php>), Ex Libris (Primo, <http://www.exlibrisgroup.com/primo.htm>).

³³ <http://www.loc.gov/standards/catenrich/catenrich-home.html>.

Inną pilną potrzebą jest udostępnienie zasobów baz danych, jakimi są OPAC biblioteczne, narzędziom indeksującym wyszukiwarek globalnych oraz użytkownikom mającym bardziej zaawansowane potrzeby, niż wyświetlenie wyszukanych opisów na ekranie. To pierwsze wiąże się z potrzebą stosowania standardowych rozwiązań, typu XML, o czym mowa jest w p. 2.8. Drugie wymaga zastosowania nowych rozwiązań technicznych i prawnych. Obecnie dostęp do OPAC możliwy jest w trybie „rekord po rekordzie”. W środowisku GBC dane bibliograficzne powinny być swobodnie rozpowszechniane, na przykład w oparciu o licencję Open Data Commons – PDDL (zob. p. 5.3), w celu umożliwienia manipulacji na wielkich zbiorach danych. Przykładem jest serwis biblios.net³⁴, oferujący dostęp do 30 mln rekordów bibliograficznych i KHW, pobranych z różnych źródeł. Dodatkowo wyposażenie rekordu OPAC we własny URL, jak w oprogramowaniu Encore³⁵ [Koszowska 2009, s. 84], powoduje, że opis katalogowy staje się pełnoprawnym, samodzielnym obiektem Webu, dostępnym w Sieci jak każda strona Web.

Portale

Portalami zazwyczaj nazywa się wielotematyczne serwisy internetowe, oferujące różnego rodzaju usługi, takie jak zamieszczanie wiadomości (na portalach naukowych: o nowych badaniach, publikacjach, grantach), bezpłatne konta poczty elektronicznej, udostępnianie programów do ściągnięcia na dysk lokalny, grupy dyskusyjne, pogawędki internetowe (*chat*), sprzedaż przez Internet, wyszukiwanie i katalog zasobów internetowych, usługi dla zarejestrowanych użytkowników (personalizacja). Usługi te coraz częściej bywają uzupełniane możliwościami charakterystycznymi dla Webu 2.0. Do niedawna w powszechnym użyciu był także termin „bramka tematyczna” (*subject gateway*). Bez roztrząsania zagadnień terminologicznych, na użytek tego rozdziału można przyjąć, że portal to bramka oferująca większy zakres usług.

Według terminologii RDN (Resource Discovery Network³⁶) portal jest to usługa sieciowa pozwalająca na dostęp do wielu innych, zróżnicowanych usług sieciowych, lokalnych i zdalnych, ustrukturyzowanych i nieustrukturyzowanych.

Mianem portalu określa się serwis GBC, który oferuje dwa rodzaje informacji: własne, przygotowane przez obsługę portalu oraz odnośniki do innych miejsc w Internecie. Zawsze znajduje się tam obszerna baza tematycznie pogrupowanych odsyłaczy, uzupełniona zazwyczaj systemem wyszukiwania informacji (jego podstawą jest duża, znana wyszukiwarka).

Portale oferują najrozmaitsze kombinacje usług i informacji. Sam fakt oferowania usług jednak nie wystarczy; miejsce w Internecie, które ma być traktowane jako portal, musi być bardzo znane w sieci i często odwiedzane. Ideę portalu internetowego przedstawia sama jego nazwa: ma to być swojego rodzaju brama do Internetu, miejsce, do którego kierują swe kroki użytkownicy sieci, gdy chcą coś znaleźć, czegoś się dowiedzieć, może także coś kupić. Nic więc dziwnego, że instytucjami często tworzącymi portale są biblioteki naukowe. W tym przypadku ma to być skrzyżowanie biblioteki z placem targowym. Właśnie portale biblioteczne, czyli takie, które tworzone są przez bibliotekarzy dla zaspokojenia potrzeb informacyjnych ich użytkowników, należy zaliczyć do serwisów GBC i tymi zajmę się w dalszej części rozdziału, chociaż jako serwisy GBC funkcjonują także inne portale, na przykład portale wydawnictw i instytucji naukowych, a nawet poszczególnych uczonych.

Według Marka Dahla i współautorów, portal biblioteczny jest efektem integracji wszelkich udostępnianych przez serwis GBC zasobów informacyjnych oraz usług, w pojedyn-

³⁴ <https://biblios.net>.

³⁵ <http://encoreforlibraries.com/>.

³⁶ <http://www.rdn.ac.uk/>.

czym systemie oferującym je poprzez jeden, wspólny interfejs [Dahl, Banerjee, Spalti 2006, s. 153]. Dla osiągnięcia takiego stanu niezbędne jest sprzężenie różnego rodzaju usług informacyjnych zarówno oferowanych przez serwis GBC, jak i jego otoczenie – najbliższe, jak i najbardziej odległe. W skład tych usług mogą wchodzić wszystkie dotąd wymienione narzędzia wyszukiwania informacji. Celem tworzenia tak skomplikowanego systemu jest zaspokajanie potrzeb użytkowników w środowisku sieciowym.

Jednoznaczne zdefiniowanie portalu bibliotecznego utrudnia duża różnorodność serwisów, które obejmowane są tą nazwą. Portale tworzone są przez różnego rodzaju biblioteki, małe i największe, stawiające sobie różne cele, choć zawsze związane z obsługą użytkowników. Portalem może być po prostu strona Web, wypełniona głównie odnośnikami do zasobów i usług biblioteki; może to być także bardziej rozbudowany serwis składający się z wielu starannie dobranych i zintegrowanych usług; może on stanowić także zintegrowany zbiór usług, dostępnych z wielu miejsc, wybieranych z otoczenia sieciowego. Można więc wyróżnić:

- Portale proste, które w skrajnym przypadku stanowią jedynie zasób odnośników do stron Web i zasobów biblioteki. Zasobami tymi mogą być katalogi online lokalne i centralne, bazy danych dostępne w sieci i zasoby pełnotekstowe. Ich integracja odbywa się głównie na zasadzie tworzenia odnośników do tych i innych usług ze strony domowej biblioteki. Najczęściej takie usługi, jak lokalny katalog biblioteki, system wypożyczeń międzybibliotecznych, zamawianie i rezerwacja pozycji są dla użytkowników dostępne odrębnie i niezależnie od portalu. Możliwa jest także ich częściowa integracja, na przykład poprzez przyjęcie jednolitego układu stron Web, wspólny projekt graficzny i ujednolicony schemat nawigacji.
- Portale zintegrowane przewyższają proste dzięki temu, że przeciwdziałają fragmentacji informacji i usług (typowej dla portali prostych). W tym celu niezbędne jest stosowanie bardziej skomplikowanej technologii, na przykład tzw. *middleware*, do których w tym kontekście zaliczymy systemy zarządzania bazą danych, skrypty, aplikacje CGI i usługi Web. Dla tych portali typowe jest wyszukiwanie sfederowane, o którym będzie mowa w następnej części książki. Może ono być łączone z agregacją metadanych, czyli pobieraniem i scalaniem metadanych z różnych źródeł dla umożliwienia ich łącznego wyszukiwania. W portalach zintegrowanych stosowany jest także coraz częściej RSS, który zwiększa rolę metadanych kosztem wyszukiwania, gdyż pozwala użytkownikowi określić własny obszar zainteresowań i otrzymywać nowe treści wprost na własny komputer [Penny 2008, s. 40]. Zintegrowany portal biblioteczny jest interfejsem użytkownika, scalającym wiele rozproszonych aplikacji w jeden, spersonalizowany serwis.
- Portale organizujące otoczenie sieciowe są częścią ogólniejszej strategii zaspokajania potrzeb użytkowników w środowisku sieciowym. Są one elementem zewnętrznego otoczenia biblioteki, dostosowując się jednocześnie do odmiennych potrzeb różnych jego elementów. Na przykład portal może współdziałać z systemem zdalnego nauczania oferując określone usługi oraz współpracować z systemem administracyjnym uczelni. Jeszcze innej usługi biblioteki naukowej mogą być obecne w obrębie portalu ogólnouczelnianego.

Można więc powiedzieć, że istnieje tyle portali, ile potrzeb użytkowników, co staje się coraz bardziej dosłowne dzięki spersonalizowanemu podejściu Webu 2.0. Podkreślić jednak należy, że wszystkie te punkty widzenia opierają się na jednej, wspólnej bazie, serwisie GBC.

Korzystanie z ogromu zasobów Webu, obsługiwanych przez serwisy GBC, powoduje znaczne uciążliwości. Zasoby te mogą być scalane przez indeksowanie w wyszukiwarkach lub realizację tzw. **wyszukiwania sfederowanego** (*federated search*). Jest ono zwane również metawyszukiwaniem. Pozwala na przygotowanie odpowiedzi na pojedyncze zapytanie użytkownika na podstawie zawartości wielu baz danych. W obecnej sytuacji,

gdy biblioteki naukowe zakupują licencje na dostęp do dziesiątek, a nawet w przypadku większych bibliotek, kilkuset baz danych, do czego dochodzą jeszcze OPAC bibliotek, użytkownicy nie wiedzą, którą z nich powinni wybrać do wyszukiwania. Nawet, jeżeli wiadomo, które bazy danych mogą mieć interesującą zawartość, dla użytkownika uciążliwe jest ich kolejne przeszukiwanie przy pomocy wielu interfejsów. Pojawiające się w bibliotekach naukowych systemy wyszukiwania sfederowanego są próbą rozwiązania tych problemów. Ponieważ użytkownicy preferują wyszukiwanie przy pomocy wyszukiwarek Web, bibliotekarze uznali wyszukiwanie sfederowane za sposób na wykorzystanie funkcjonalności tych wyszukiwarek w zastosowaniu do części zasobów globalnej biblioteki cyfrowej, często stanowiących wcześniej część ukrytego Webu.

W swoim artykule, w którym nawoływał bibliotekarzy do stworzenia własnego serwisu wyszukiwawczego w oparciu o technologie wyszukiwarek, Norbert Lossau w następujący sposób przedstawia zalety wyszukiwania sfederowanego:

- Indeksowanie wyłącznie treści o wysokiej jakości.
- Możliwość obsługi wyszukiwania różnych rodzajów treści w różnych formatach (meta-dane, DLO, grafika, multimedia, dane binarne).
- Wzrost funkcjonalności wyszukiwania poprzez zastosowanie takich narzędzi jak języki deskryptorowe (tezaurusy) i inne JIW.
- Przedstawianie wyników wyszukiwania uporządkowanych według różnych sposobów, zarówno prostego sortowania według takich kryteriów, jak autor, tytuł, rok, opis rzeczowy, jak i według rankingu relewancji.
- Automatyczne tworzenie metadanych, przynajmniej na poziomie minimum opisu [Lossau 2004].

Wyszukiwanie sfederowane daje wiele korzyści, ale jednocześnie ma pewne wrodzone ograniczenia. Jednym z nich jest niekonsekwencja w sposobie myślenia jego zwolenników. Z jednej strony narzekają oni na ograniczenia technologii wyszukiwarek komercyjnych, z drugiej jednak sami chcą ją stosować, często na skalę globalną, mając nadzieję na ominięcie wskazanych przez siebie wad.

Biblioteki prenumerują bazy danych zawierające różne rodzaje zasobów, wykorzystywanych do różnych celów, co utrudnia formułowanie zapytań. Podczas wyszukiwania sfederowanego zapytania użytkownika muszą być tłumaczone przez uproszczony interfejs wyszukiwawczy na wiele syntaktyk tworzenia zapytań, stosowanych w zróżnicowanych zasobach, podlegających wyszukiwaniu. Scalanie w jeden, sensownie skomponowany zbiór wielu zestawów wyników, uzyskanych z wyszukiwania pełnotekstowych baz danych, indeksów cytowań, OPAC, wyszukiwarek internetowych, repozytoriów i archiwów cyfrowych jest zadaniem zarówno metodologicznie jak i technicznie bardzo trudnym. Metody określania relewancji wyników, wykorzystywane przez wyszukiwarki, takie jak Google lub Yahoo, nie dają się łatwo przenosić do wyszukiwania sfederowanego. Największe wyszukiwarki Web przechowują informacje we własnych bazach danych i obliczają rankingi według własnych, nie znanych ogólnie kryteriów. Wyszukiwarki sfederowane natomiast konwertują dane wejściowe użytkownika na strukturę zapytań poszczególnych baz danych, tworzonych przez wielu dostawców. W bazach tych używane są różne struktury danych, dostosowane do odmiennych algorytmów wyszukiwawczych. Z tego powodu systemy wyszukiwania sfederowanego nie mogą zapewnić tworzenia rankingów wyników w sposób podobny do wyszukiwarek Web.

Oczywistym efektem stosowania wyszukiwania sfederowanego jest wzrost częstotliwości wyszukiwania we wszystkich dostępnych bazach danych. Utrudnia to prowadzenie wiarygodnych analiz wykorzystania baz danych, często służących do modyfikacji polityki ich zakupu, a dla dostawców baz danych może być sygnałem do podniesienia cen. Z drugiej strony często zdarza się, że użytkownicy, przyzwyczajony się do konkretnej bazy danych, nie korzystają z innych, równie przydatnych serwisów. W takim przypadku wy-

szukiwanie sfederowane może zwiększyć użyteczność rzadziej wykorzystywanych baz danych, z korzyścią dla wyników wyszukiwania.

Metawyszukiwarki krytykowane są także z powodu niskich wymagań stawianych użytkownikom co do ich umiejętności informacyjnych. William Frost twierdzi, że z punktu widzenia wysiłków podejmowanych dla szkolenia użytkowników, metawyszukiwanie jest krokiem wstecz i sposobem na uniknięcie potrzeby uczenia się technik informacyjnych. Wybór narzędzia wyszukiwawczego powinien być podstawową umiejętnością zdobywaną przez użytkownika informacji. Metawyszukiwanie ogranicza potrzebę uczenia użytkowników zaawansowanych technik wyszukiwania informacji w bibliotekach [Frost 2004, s. 38].

Można doszukać się podobieństw pomiędzy możliwościami metawyszukiwarek a usługami Google Scholar. Obie usługi kierują swoją ofertę do tej samej grupy użytkowników. Metawyszukiwanie jest bardziej przydatne dla użytkowników mających dostęp do zasobów licencjonowanych; Google Scholar daje większe możliwości wyszukiwania i dostępu do zasobów Open Access [Herrera 2007, s. 51].

Gdyby chcieć szczegółowo opisać sposób działania **wyszukiwarki**, w wyniku którego użytkownik może wyszukać przydatne mu informacje, byłby to bardzo obszerny tekst, a w dodatku w szczegółach różniłby się nieco od rzeczywistości, ze względu na konieczne uproszczenia. Wyszukiwarki są bowiem obecnie bardzo skomplikowanymi i rozbudowanymi systemami wyspecjalizowanego oprogramowania, którego najważniejsze zasady działania okryte są ścisłą tajemnicą. Cały czas trwają prace nad udoskonalaniem sposobu ich działania (czyli nad dalszym komplikowaniem ich funkcjonowania), co w przyszłości doprowadzić może do sytuacji, że przy ich pomocy wyszukać będzie można nie tylko informacje o wszystkim i każdym, ale także wszystko i każdego (tzn. nie tylko informacje, ale także obiekty fizyczne, w tym osoby).

Początkowo wyszukiwarki projektowane były w oparciu o tradycyjne metody wyszukiwania informacji. Niektóre z nich, jak AltaVista, Lycos czy Excite tworzyły olbrzymie zcentralizowane indeksy stron Web. Odpowiadając na zapytanie przeglądały one swoje indeksowane bazy danych i wyświetlały je z pamięci podręcznej (*cache*) w oparciu o wystąpienia słów kluczowych i ich współwystępowanie. Zasady tradycyjnego indeksowania były przydatne w przeszukiwaniu baz danych, jednak w Internecie, maksymalnie nieustrukturyzowanym i zmiennym, stały się niewystarczające. W takim środowisku zmienia się znaczenie tradycyjnych miar trafności (liczy się kilka pierwszych wyników) i szczególnie kompletności (często miarom stron Web jako wynik wyszukiwania) na rzecz perzynencji (użytkownik powinien na początku listy wyników dostać to, czego potrzebuje). W efekcie pomiędzy wyszukiwarkami a innymi serwisami opartymi na bazach danych powstało w tym zakresie wyraźne zróżnicowanie: naukowe serwisy informacyjne nadal za priorytet uważają zapewnienie maksymalnej kompletności wyszukiwania, podczas gdy wyszukiwarki pozwalają osiągnąć wysoką trafność rankingowanych wyników.

Wyszukiwanie w Internecie z punktu widzenia użytkownika musi być możliwie proste. Wiąże się to z wielką demokratyzacją procesów wyszukiwania informacji, które stało się codzienną aktywnością większości osób [Hartley 2009, s. 151]. Znakomita większość zapytań do wyszukiwarki składa się z jednego lub dwóch słów kluczowych³⁷. Zapytanie zazwyczaj zadawane jest tylko raz, nie jest ono następnie reformułowane. Rozszerzone możliwości wyszukiwawcze, na przykład z wykorzystaniem algebry Boole'a są wykorzystywane sporadycznie. Siłą wyszukiwarek jest ich powszechne stosowanie. Według badań przeprowadzonych w OCLC 84% użytkowników decydujących się na poszukiwania

³⁷ Badania wskazują na stały wzrost stopnia skomplikowania zapytań, sformułowanych przez użytkowników wyszukiwarek. W 1998 roku zapytanie składało się przeciętnie z 1,2 słowa; w 2004 r. z 2,5 słowa, a w 2006 z 3,3 słowa. W 2006 r. wyszukiwania wg jednego słowa stanowiły 22% wyszukiwań, według dwóch – 30% i według trzech – 24% [Yang, Yang, Yuan 2007, s. 236].

informacji w źródłach elektronicznych rozpoczyna je od wyszukiwarki, natomiast tylko 2% od strony Web biblioteki [Rosa 2005]. Na wysoką pozycję wyszukiwarek globalnych (i innych źródeł elektronicznych) wskazują także polskie badania [Jaskowska, Korycińska-Huras, Próchnicka 2009, s. 240]. Wyszukiwarka, jako narzędzie wyszukiwawcze, nie ma więc praktycznie konkurencji. Do słów wykorzystywanych w zapytaniach użytkowników stosuje się model *long tail*³⁸: istnieje niewielka liczba słów kluczowych, wykorzystywanych bardzo często i wielka liczba słów, stosowana znacznie rzadziej. Użytkownicy zwykle poprzestają na przeglądnięciu pierwszej strony (ekranu) wyników przedstawionych przez wyszukiwarkę, czyli na ogół pierwszych dziesięciu wyszukanych stron Web, w tym średnio pięć stron jest przeglądanych [Spink, Jansen 2004, s. 101].

Podobne wyniki uzyskała w swoich badaniach Julie Hallmark, wskazując przy okazji na szybki wzrost wykorzystania cyfrowych zasobów dostępnych online [Hallmark 2004]. W 1998 r. 83% badanych przez nią specjalistów w dziedzinie chemii stosowało wyszukiwarki do przeszukiwania zasobów, jednak tylko 5% z nich korzystało także z publikacji w formie elektronicznej, raczej wybierając materiały drukowane. W 2002 r. wykorzystanie wyszukiwarek pozostało na podobnym poziomie (85%), jednak aż 96% odszukanych materiałów było czytanych w formie elektronicznej lub drukowanych z tej formy do późniejszego czytania.

Dirk Lewandowski zwraca uwagę na przyczyny niepowodzeń podczas wyszukiwania w Webie [Lewandowski 2005, s. 138-139]. Podstawową jest olbrzymi rozmiar Webu i konieczność indeksowania miliardów jego stron, o czym była mowa w p. 2.6. Liczba stron do indeksowania oraz ich dynamika powoduje, że indeksy wyszukiwarek są zwykle nieaktualne; wielkie wyszukiwarki częściej odwiedzają strony Web, ale również tu odstęp czasu pomiędzy kolejnymi indeksowaniami może dochodzić do 30 dni. Efektywne indeksowanie utrudnione jest też przez funkcjonowanie z jednej strony spamu, czyli obiektów niepotrzebnie indeksowanych, bo niechcianych, zaśmiecających indeksy wyszukiwarek, a z drugiej strony ukrytego Webu, czyli obiektów niemożliwych do zindeksowania; są to strony, do których dostęp jest zablokowany przez ich autorów (na przykład hasłem dostępu³⁹), którzy nie życzą sobie indeksowania (umieszczając odpowiednią etykietę w części Meta kodu), tworzone w locie (na przykład przez CGI), zbyt głęboko umieszczone w strukturze strony Web, a w końcu strony nie zawierające odnośników do innych stron i nie skupiające odnośników zewnętrznych.

Z punktu widzenia celu wyszukiwania informacji w Webie można wyróżnić dwa przypadki:

- Wyszukiwanie określonej pojedynczej, dobrze zdefiniowanej strony Web (pojedynczego zasobu sieciowego) odbywa się w trakcie poszukiwania na przykład strony określonej instytucji, projektu naukowego lub publikacji, której autor i tytuł jest znany użytkownikowi

³⁸ Long tail („długi ogon”) jest określeniem zjawisk występujących w biznesie elektronicznym, związanych z różnicami pomiędzy działalnością na rynku tradycyjnym i elektronicznym. Główna różnica polega na opłacalności sprzedaży (dostarczenia) użytkownikowi nawet pojedynczego egzemplarza interesującego go przedmiotu (np. dokumentu). W tradycyjnym handlu należało nastawić się na masową sprzedaż niewielkiej liczby produktów; obecnie, w handlu prowadzonym elektronicznie, gdzie zniknęły (lub znacznie zmalały) koszty magazynowania i dystrybucji jest obojętne, czy zarabia się na sprzedaży wielkiej liczby egzemplarzy pojedynczych produktów (np. tytułów książek), czy też na sprzedaży pojedynczych egzemplarzy wielkiej liczby produktów. Dzięki temu przedsiębiorstwa typu Amazon (księgarnia) czy Google (reklamy) mogą znacznie lepiej dostosować swoją ofertę do potrzeb użytkowników. Po raz pierwszy na to zjawisko uwagę zwrócił Chris Anderson [Anderson 2004]. Występuje ono także w bibliotekach oferujących swoim użytkownikom zdalne usługi [Dempsey 2006].

³⁹ Dotyczy to większości komercyjnych stron zawierających treści o dostępie płatnym, np. stron wydawców czasopism komercyjnych. Twórcy tych stron jednak coraz częściej dbają o ich właściwe indeksowanie przez wyszukiwarki, dzięki czemu dostęp do dokumentów jest odpłatny, ale istnieje możliwość ich wyszukania poprzez wyszukiwarkę internetową, co zazwyczaj daje bezpłatny dostęp do metadanych, abstraktu i tp. informacji.

kowi. W tym przypadku właściwe jest zastosowanie wyszukiwarki indeksującej całość zasobów Web, typu Google. Dzięki algorytmom wyszukiwarek, po poprawnym sformułowaniu zapytania, poszukiwana strona znajduje się zazwyczaj wśród kilku pierwszych adresów, wyświetlonych jako wynik wyszukiwania;

- Wyszukiwanie informacji na wybrany temat, bez względu na źródło tej informacji i najczęściej formę zawierającego ją dokumentu. Wówczas, dla zapewnienia odpowiedniej jakości wyszukiwanych obiektów, lepiej jest posłużyć się wyspecjalizowaną wyszukiwarką typu Google Scholar, Scirus⁴⁰, CiteSeer⁴¹. W dalszej kolejności przydatne może być lokalne przeszukiwanie zasobów poszczególnych serwisów GBC. W odpowiedzi otrzymujemy zazwyczaj bardzo obszerną grupę obiektów, zawierających bardzo zróżnicowane informacje, zarówno co do treści, jak i jakości, konieczna jest więc ich dalsza selekcja.

Główna różnica pomiędzy dwoma powyższymi rodzajami procesów wyszukiwania polega na tym, iż w drugim przypadku wyszukiwania sieć Web jest traktowana jako „czarna skrzynka”, czy też nawet autonomiczny system ekspertowy, natomiast w pierwszym przypadku wyszukiwania pojedynczych stron, struktura sieci jest istotna dla użytkownika.

Procesy wyszukiwania Web składają się z dwóch części:

- Pierwsza, realizowana offline, jest wykonywana okresowo przez wyszukiwarkę i polega na ściąganiu części zawartości Webu w celu stworzenia zestawu stron, które są następnie transformowane w indeks wyszukiwawczy.
- Druga, realizowana online, jest wykonywana za każdym razem, gdy tworzona jest odpowiedź na zapytanie użytkownika; wówczas indeks wykorzystywany jest do wyselekcjonowania zestawu dokumentów, sortowanych następnie według reguł określających stopień ich relewancji z zapytaniem użytkownika.

Współczesne wyszukiwarki łączą różne sposoby wyszukiwania: wyszukiwanie proste, złożone, wspomagane, na podstawie katalogu stron, zapytania o strony podobne (*query-by-example*). Mają one zapewnić wzrost trafności wyszukiwania, gdyż właśnie ten element efektywności wyszukiwania jest według twórców Google głównym celem konstruktorów wyszukiwarki [Brin, Page 1998].

W wyniku starań o zapewnienie maksymalnej trafności wyszukiwania nastąpił rozwój wyszukiwarek wyspecjalizowanych. Web jest wykorzystywany przez bardzo różne osoby, z różnym przygotowaniem i potrzebami. Rozmiar Webu i setki tysięcy stron w odpowiedzi na każde zapytanie może powodować u nich dezorientację. Z tego powodu powstały specjalne wyszukiwarki tematyczne, specjalizujące się w informacjach lokalnych, muzyce i grafice (jak już wspomniałem, wyszukiwanie multimediów zwykle sprawia znacznie więcej kłopotów niż wyszukiwanie tekstowe), nazwach geograficznych czy osobowych. Również tworzenie metawyszukiwarek można uznać za działania w podobnym kierunku. Zbliżony efekt można uzyskać kastomizując sposób wyszukiwania wyszukiwarki globalnej, takiej jak Yahoo lub Google, na przykład przez dodanie listy stron służących do przeszukiwania [Bradley 2008].

Stwierdzenie, że wyszukiwarka po prostu przechodzi od strony do strony pobierając z nich słowa i robi z nich indeks, który jest wykorzystywany do udzielania odpowiedzi na zapytania, jest zgodne z prawdą, jednak tak ogólne, że aż wprowadza w błąd. Po pierwsze, istnieje tzw. ukryty Web, o czym była mowa. Po drugie, dobra wyszukiwarka nie jest po prostu sprytnie ułożonym algorytmem wyszukiwawczym. Jest to wieloczęściowy system działający dla przekazywania wysokiej jakości wyników wyszukiwania. Stosowane są wielopiętrowe struktury danych, zaprojektowane dla realizacji określonych celów.

Pierwszy etap indeksowania dokumentu Web polega na wydobyciu z niego standardowej, logicznej reprezentacji treści. Najczęściej stosowaną przez wyszukiwarki reprezenta-

⁴⁰ <http://www.scirus.com/>.

⁴¹ <http://citeseer.ist.psu.edu/>.

cją logiczną dokumentu jest zestaw słów, przez co indeksowany dokument jest postrzegany jako ich nieuporządkowany zestaw. We współczesnych wyszukiwarkach reprezentacja jest rozbudowana o dodatkowe dane, wskazujące na częstotliwości słów, atrybuty formatowania tekstu, oraz metadane tworzone zgodnie z zasadami HTML [Baeza-Yates, Castillo 2006, s. 530].

Sortowanie dokumentów odszukanych odbywa się według algorytmu PageRank, który decyduje o miejscu wyszukanej strony na liście stron przedstawionej użytkownikowi. Algorytm ten daje zadowalające rezultaty nawet w przypadku, gdy dostępnych jest bardzo mało informacji o rzeczywistych potrzebach użytkowników. Jakość strony Web w żadnym stopniu nie zależy od jej zawartości. Wykorzystuje się wyłącznie strukturę odnośników w zasobie. Każdy odnośnik to głos oddany na stronę, do której prowadzi. Odnośniki nie są jednak traktowane równorzędnie. Odnośnik ze strony wysoko postawionej w rankingu jest bardziej wartościowy, niż odnośnik z innej strony. Całkowity PageRank strony jest sumą wszystkich wartości odnośników wskazujących na nią. Wartość odnośnika zależy od cech strony, od której odnośnik przychodzi: PageRank jest dzielony przez liczbę odnośników wychodzących ze strony. W efekcie możliwe jest porównywanie dowolnych dwóch stron i określenie ich miejsca w rankingu. Autorzy PageRank wykorzystali więc ideę tzw. mądrości lub inteligencji zbiorowej⁴², nazywanej też „mądrością tłumów” (będącej także cechą serwisów Webu 2.0), uznając, że najlepsze są te strony, które uzyskały w powszechnym głosowaniu użytkowników Internetu najwięcej głosów. Innym przejawem mądrości zbiorowej, skrzętnie rejestrowanym przez twórców Google są tzw. strumienie kliknięć (*clickstream*), czyli zjawisko polegające na tym, że subiektywne oceny samodzielnie działających jednostek sumują się we wskazania najwartościowszych miejsc w Internecie⁴³. Kierowanie się opinią i doświadczeniami innych, a więc mądrością zbiorową, jest obecnie jedną z najbardziej użytecznych strategii wyszukiwania informacji [Juza 2007, s. 47]. Jest ona jednak również krytykowana za tworzenie tzw. Google Effect, czyli mylenie popularności z jakością [Brabazon 2006, s. 160].

Szczególnie istotna dla globalnej biblioteki cyfrowej wydaje się inicjatywa Google pod nazwą Google Scholar. Wyszukiwarka ta, udostępniona jak dotąd w wersji beta, przedstawia użytkownikom wyniki wyszukiwań ograniczone do obiektów o treści naukowej, co jak się wydaje nie stanowiło wcześniej oferty żadnej innej wyszukiwarki globalnej. Jest to bezpłatnie dostępna usługa internetowa o przyjaznym interfejsie, podobnym do standardowego interfejsu Google. Część obiektów indeksowanych przez Google Scholar znajduje się na serwerach wydawców komercyjnych, skąd dostępne mogą być za opłatą, jednak nawet w takich sytuacjach użytkownik ma do dyspozycji przynajmniej opis, często z abstraktem. Jednak znaczną część zasobów Google Scholar stanowią obiekty publikowane w trybie Open Access. Często zdarza się, że wyszukiwarka kieruje do dwóch wersji publikacji: płatnej i OA.

Oprócz dostępu do pełnych tekstów dokumentów, interesującą ofertą dla użytkowników stanowią analizy wykonywane przez wyszukiwarkę i wynikające z nich rankingi relewancji. Jednym z wielu kryteriów służących rankingowi jest liczba cytowań dokumentu, które zastępują odnośniki do stron Web, uwzględniane w standardowych Google. W efekcie częściej cytowane dokumenty są umieszczane wyżej w rankingu, co ułatwia ich odszukanie przez użytkowników. Dzięki istnieniu indeksu cytowań użytkownik może także śledzić wszystkie cytowania danego dokumentu, co nie jest dostępne w Google – tam wyszukiwarka nie pokazuje odnośników do wyszukanych stron. Automatyczna realizacja indeksu cytowań wymaga, aby odsyłacze bibliograficzne w tekście były dostępne

⁴² Henry Jenkins, za Pierre Levym, twórcą pojęcia inteligencji zbiorowej, określa ją jako zdolność wirtualnych społeczności do wykorzystywania połączonych uzdolnień wszystkich ich członków [Jenkins 2007, s. 31].

⁴³ Na podstawie strumieni kliknięć poznawane są preferencje użytkowników – informacji o tym, co użytkownik odwiedził i co wybierał przeglądając Web. Strumienie te są formą metadanych [Battelle 2006, s. 194].

do analizy; jest to ułatwione w przypadku indeksowania pełnotekstowego. Zastosowane procedury przypominają rozwiązania wykorzystywane w takich indeksach cytowań, jak CiteSeer, z tą różnicą, że indeks ten dotyczy tylko wybranego zakresu dokumentów (informatyka ogólnie pojęta), natomiast Google Scholar jest wyszukiwarką globalną i ogólnym indeksem cytowań.

Wielu badaczy podejmuje próby oceny jakości wyszukiwania Google Scholar stosując metody przeniesione ze środowiska tradycyjnych baz danych. Jak wynika z badań z 2005 r., przedstawionych przez Philippa Mayra i Anne-Kathrin Walter, w stosunku do wielu abstraktowych baz danych Google Scholar nie oferuje porównywalnej kompletności wyszukiwanych zasobów, w związku z czym proponują stosować go jako narzędzie uzupełniające wyszukiwanie w bazach danych [Mayr, Walter 2007]. Wynika to z indeksowania głównie dokumentów dostępnych bezpłatnie online. Może się to zmienić po zrealizowaniu projektów masowej digitalizacji zbiorów największych bibliotek naukowych, prowadzonych m.in. przez Google. Inne badania mogą dawać bardziej optymistyczne wyniki. Lokman Meho i Kiduk Yang porównywali wyniki wyszukiwań uzyskane w Google Scholar oraz w Scopus i ISI Web of Science, a więc bardziej porównywalnych serwisów [Meho, Yang 2007]. Google pozwalał na lepsze wyszukiwanie w materiałach konferencyjnych (czterokrotnie lepsze wyniki), źródłach nie-anglojęzycznych (ponad sześciokrotnie) oraz materiałach archiwizowanych na osobistych lub instytucjonalnych stronach Web (w ogóle nie uwzględnianych przez pozostałe zasoby). Scopus oraz Web of Science lepiej natomiast selekcjonują publikacje o wysokiej cytowlności, podczas gdy wyniki Google Scholar obciążane są dużą masą słabo cytowanych publikacji. Porównanie jakości wyszukiwania w Google i systemach bibliotecznych wskazuje na uzyskiwanie prawie identycznych wskaźników trafności (w obu przypadkach około 60%), przy czym wyszukiwarka góruje nad OPAC dostępnością wyszukiwanych materiałów, a katalog biblioteczny ich jakością [Brophy, Bawden 2005, s. 510]. Wśród wyszukiwarek Google Scholar daje najlepszą trafność wyszukiwania treści naukowych [Robinson, Wusteman 2007, s. 76]. Wyszukiwarki globalne, takie jak Google i Yahoo są przy tym wyraźnie bardziej efektywne od mniejszych wyszukiwarek lokalnych [Lewandowski 2008, s. 927]. Badania wskazują także na lepsze wyniki prezentowane przez Google w niektórych dyscyplinach, jak informatyka czy niektóre nauki społeczne [Kousha, Thelwall 2007].

W wyniku realizacji takich inicjatyw, jak Google Scholar i Google Books, nieaktualny staje się dychotomiczny podział na wyszukiwanie w Google (dające ogólne pojęcie o tematyce) i wyszukiwanie w zasobach bibliotek (dające wyniki wiarygodne) [Lack 2006, s. 77]. Dzięki wyszukiwarce typu Google możliwe jest zbadanie różnego rodzaju zasobów informacji – cyfrowej i tradycyjnej, wysokiej i niezanej jakości, darmowej i udostępnianej komercyjnie⁴⁴. Dzięki temu, że digitalizacja w ramach Google Books opiera się na drukowanych zbiorach bibliotek akademickich, powstający zasób cyfrowy znajduje głównie zainteresowanie w środowiskach naukowych⁴⁵. Wyszukiwarki tego typu służą także do zapewnienia możliwości przeszukiwania zasobów o ograniczonym dostępie (ukrytych), na przykład zasobów określonej instytucji czy organizacji. Wówczas użytkownik ma możliwość wyboru pomiędzy wyszukiwaniem ograniczonym do określonego zasobu i przeszukiwaniem całego Webu. Wyszukiwarki stają się więc narzędziem pozwalającym na wyszukiwanie badawcze, o którym mowa była na początku rozdziału.

⁴⁴ Często zwraca się uwagę, że wielkie projekty digitalizacji zbiorów bibliotek, realizowane przez Google mają na celu nie tyle udostępnienie digitalizowanych publikacji w wersji cyfrowej, lecz „umożliwienie wyszukiwania w trybie online informacji dostępnej offline” [Johnson 2007, s. 3]. Chodzi o wyszukiwanie pełnotekstowe publikacji drukowanych. Skanowanie to ma głównie zapewnić możliwość indeksowania pełnotekstowego kopiowanych zasobów; ich udostępnianie w formie cyfrowej jest jakby efektem ubocznym. Dlatego też pojawiająca się krytyka dotycząca jakości skanów (np. [Duguid 2007]) nie jest najistotniejsza.

⁴⁵ Można szacować, że ponad 90% pozycji (materializacji) digitalizowanych przez Google Books nie należy do literatury pięknej, a z tego 78% to publikacje na poziomie naukowym [Lavoie, Dempsey 2009].

Google jest obecnie zestawem produktów, które ulegają bardzo dynamicznemu rozwojowi. Może on prowadzić do realizacji idei, które od wielu lat znane są pod nazwą Semantycznego Webu. Nadzieję na to dają nowe rozwiązania, opublikowane w 2007 r., pod nazwą Programmable Search Engine (PSE). Przy jej pomocy strony Web mogą być lepiej strukturyzowane, a ich twórcy mają zapewnioną dwustronną komunikację z Google. Strony Web są efektywniej indeksowane, m.in. dzięki temu, że ich właściciele mają możliwość decydowania o sposobie postępowania wyszukiwarki podczas indeksowania. Dzięki tym narzędziom Google będzie mógł bardziej efektywnie indeksować zasoby Internetu, stosując takie nowe możliwości, jak:

- Dostarczanie bardziej szczegółowych informacji w wynikach wyszukiwania (na przykład wykaz dostępnych zasobów biblioteki, a nie tylko jej adres), wraz z możliwością porównywania zawartości stron (cen i innych warunków).
- Zwiększenie kustomizacji wyników w oparciu o preferencje użytkowników i wzorce wcześniejszych zachowań, m.in. na podstawie badania strumieni kliknięć. Google, ucząc się, będzie zwiększał swoją efektywność wraz z każdym wykonanym wyszukiwaniem. Na podstawie zdobytej wiedzy możliwa jest modyfikacja zapytania użytkownika w taki sposób, aby uzyskać on odpowiedź najlepiej spełniającą jego potrzeby;
- Ograniczenie liczby spamu i stron niechcianych.
- Indeksowanie zasobów chronionych hasłem (ukryty Web) za zgodą właściciela, co umożliwi tworzenie informacji o zasobach płatnych.
- Indeksowanie zasobów udostępnianych jako bazy danych.
- Indeksowanie stron dynamicznych (stron zmieniających się pod wpływem żądań użytkowników), w tym rozpoznawanie ich kolejnych wersji.
- Znacznie lepsze indeksowanie stron zawierających obiekty nietekstowe (głównie pliki video i audio).
- Integracja informacji z różnych stron Web w celu dostarczania pełniejszej odpowiedzi na zapytanie. Można będzie zadawać pytania w celu przeszukania całego Webu dla osiągnięcia tylko jednego wyniku, na przykład opisu książki w określonej bibliotece (na przykład nie dalej niż kilka kilometrów od domu użytkownika) – Google wyświetli najbliższe biblioteki posiadające poszukiwaną pozycję i pozwoli ją zamówić.
- Zastosowanie nowych technik wyszukiwawczych dla uszczegółowienia informacji, służącej doborowi grupy celowej dla reklamodawców, w wyniku czego reklamy trafią będą dokładnie do osób ich potrzebujących.

Modyfikacje te spowodują, że wyszukiwarka będzie w stanie reagować na zmiany zachodzące w jej otoczeniu. Stanie się narzędziem otwartym, programowalnym i bardziej interaktywnym. Użytkownicy uzyskają możliwość komunikowania się z wyszukiwarką i ujawniania swoich preferencji, do których program dostosowywać będzie swój sposób pracy⁴⁶. Wyszukiwarka będzie posiadać nie tylko informacje o tym, co wcześniej zostało wyszukane (czym dysponuje już obecnie), ale także będzie decydować, czy ujawnić te informacje użytkownikom z uwzględnieniem kontekstu ich wyszukiwań, wzorców wyszukiwań a także historii dodawania i usuwania terminów wyszukiwawczych podczas modyfikacji zapytania na podstawie uzyskanych wyników [Nolan 2008, s. 40]. Taki sposób postępowania przypomina działanie systemów uczących się.

Google, z płaskiego indeksu stanie się bazą danych innych baz danych, czyli rodzajem super lub meta bazy danych. Zawierać będzie informacje o stronach Web oraz ich zawartości, oferowanych dobrach (w tym książkach i innych publikacjach) i inne bardzo szczegółowe informacje. Spowoduje to konieczność gromadzenia jeszcze bardziej szczegółowych metadanych o indeksowanych stronach w postaci kodów i etykiet dotyczących ich

⁴⁶ W tym kierunku idą też prace wydawców, czego przejawem jest ACAP (Automatem Control Access Protocol – <http://www.the-acap.org/>), który pozwala wydawcom wskazywać wyszukiwarkom treści dozwolone do indeksowania i wyświetlania, okresy dozwolenia i sposób ewentualnej płatności.

treści. W oparciu o tę bazę danych możliwe stanie się połączenie treści z inteligentnymi aplikacjami przetwarzania informacji. Wyszukiwarka nie utraci prostoty swego interfejsu, a więc z punktu widzenia użytkownika nic się nie zmieni.

W zasobach pełnotekstowych, powiązanych wzajemnie hiperlinkami, klasyczne metody wyszukiwania stają się mniej efektywne i mogą być zastępowane nowymi technikami. Dla wyszukiwań pewnego rodzaju nadal właściwe jest tworzenie ustrukturyzowanych zapytań, jednak techniką bardziej wydajną, szczególnie w środowisku rozproszonym, staje się **przeglądanie** (*browsing*). Jest ono uważane za rodzaj nawigacji⁴⁷. Technika ta bywa uożsamiana z przeglądaniem zawartości regałów w bibliotece tradycyjnej [Borgman 1996, s. 494].

Maria Próchnicka przeglądanie uznaje za specyficzny dla Internetu sposób poszukiwania informacji. Internet pojmowany jest wówczas w kategoriach przestrzeni (zob. p. 3.1), a użytkownik porusza się w tej przestrzeni pomiędzy umieszczonymi w niej obiektami. Podczas przeglądania użytkownik dokonuje selekcji informacji na podstawie rozpoznania jej cech, które nie były w momencie rozpoczęcia poszukiwania dobrze (precyzyjnie) zdefiniowane. Rozpoznanie to ma często charakter niesystematyczny i przypadkowy [Próchnicka 2006, s. 47].

Niesystematyczne i przypadkowe poszukiwania dają jednak dobre rezultaty dzięki specyficznej strukturze Webu, opartej na hipertekście. Odległość pomiędzy dowolnymi dwoma dokumentami Web jest prawie zawsze mniejsza niż 20 odnośników⁴⁸ (kliknięć) [Borgman, Furner 2002, s. 15]. Oznacza to, że Web jest co prawda olbrzymim zasobem, jednak w wysokim stopniu nasyconym odnośnikami, pozwalającymi na postrzeganie Webu jako jednolitej, logicznej struktury. Wynika z tego, że struktura odnośników w Webie odgrywa równie istotną rolę, co same dokumenty.

Użytkownicy GBC często rozpoczynają poszukiwania nie mając ściśle określonych potrzeb informacyjnych. Odczuwają oni raczej trudny do opisanego stan braku wiedzy w jakimś zakresie. Potrzeby informacyjne zazwyczaj trudne są do jednoznacznego opisanego i zdefiniowania, co w oczywisty sposób utrudnia sformułowanie zapytania. W takiej sytuacji właściwą techniką jest przeglądanie, gdyż umożliwia stwierdzenie relewantności informacji dopiero po dotarciu do niej. Przeglądanie i nawigacja umożliwiają użytkownikowi podążanie za odnośnikami od jednego źródła do innego, powiązanego z nim relacjami; podążać można zarówno od metadanych do danych, od jednej wersji obiektu do innej, od jednego obiektu do innego, połączonego relacjami: wspólnego przedmiotu, autora, właściciela, dostępności czy innymi atrybutami. Powiązania te mogą być eksplorowane przez automatycznych agentów, działających w imieniu użytkowników, badających zasoby serwisów GBC oraz przedstawiających użytkownikowi do oceny zestaw wybranych zasobów.

Przeglądanie jako technika wyszukiwania informacji pozwala na:

- poznanie pytań, które można zadawać na dalszych etapach wyszukiwania;
- uszczegółowienie wyników wyszukiwania;
- dotarcie do informacji, o których istnieniu użytkownik nie wiedział.

⁴⁷ Termin nawigacja jest tu stosowany w znaczeniu techniki polegającej na badaniu zasobów informacji przy pomocy wyboru i podążania ścieżkami wytyczonymi wewnątrz poszczególnych zasobów (systemów) i pomiędzy nimi.

⁴⁸ Jest to tzw. efekt małego świata (*small world effect*), znany także jako „sześć stopni oddzielenia”. Pojęcie to wprowadził w 1967 r. Stanley Milgram, psycholog społeczny z Uniw. Harvarda i Yale, który stwierdził, że dowolna osoba na świecie może być połączona z dowolną inną za pomocą łańcucha pięciu lub sześciu znajomych. Tego typu opis małego świata jest oparty na większej strukturze społeczeństwa ludzkiego, powstałej na zasadzie relacji funkcjonujących pomiędzy ludźmi. Małe światy są rozpowszechnione w przyrodzie i społeczeństwie strukturą sieciową, w której powiązane elementy albo łączą się bezpośrednio, albo występują między nimi krótkie łańcuchy pośredników [Castells 2008, s. 83].

Proces przeglądania zawsze rozpoczyna się od jakiegoś, wyznaczonego punktu, więc użytkownik musi wstępnie opisać swoje potrzeby za pomocą wybranych terminów czy innych cech. Niezbędny jest też wybór serwisu początkowego, od którego rozpoczynane będzie przeglądanie. Właściwy wybór punktu startowego ma decydujące znaczenie dla efektywności wyszukiwania.

Nawigacja i przeglądanie wydają się naturalnymi sposobami zachowania człowieka, opierającymi się na dwóch podstawowych zasadach poznania: ludzie uczą się przez tworzenie asocjacji z wcześniej nabytą wiedzą oraz łatwiej rozpoznają informację prezentowaną, niż wydobywaną z pamięci. Systemy wymagające od użytkownika wprowadzania słów kluczowych opierają się na przywoływaniu wiedzy, gdyż wymagają od użytkownika tworzenia terminów. Systemy, które przedstawiają użytkownikowi uporządkowany zestaw terminów do wyboru opierają się na rozpoznawaniu wiedzy. Są one szczególnie efektywne w zakresach, wychodzących poza codzienną wiedzę użytkownika [Borgman i in. 1995]. Przeglądanie i nawigacja nie są substytutami wyszukiwania przez zapytanie. Te pierwsze są bardziej przydatne w małych zasobach dziedzinowych, te drugie w wielkich, różnicowanych zbiorach; nie przeszkadza to jednak, aby przeglądanie było najpopularniejszą techniką w Internecie. Przeglądanie jest także bardziej efektywne w przypadku zapytań o zagadnienia słabo zdefiniowane.

W praktyce występuje połączenie obu technik wyszukiwania. Dzięki temu, że użytkownicy postępują ścieżkami definiowanymi przez własne wybory, przeglądanie pozwala przynajmniej częściowo ominąć konieczność powtórnego sformułowania zapytania, w przypadku, gdy w odpowiedzi uzyskuje się wynik nieadekwatny do potrzeb (zbyt wiele, zbyt mało lub nierelevantne odpowiedzi). Hipertekst, który jest wykorzystywany w tym postępowaniu, może być kolejną barierą, gdyż użytkownicy mogą czuć się „zagubieni” w sieci powiązań. Dla uniknięcia tej wady tworzone są mapy serwisów i inne narzędzia ułatwiające nawigację.

Dobre efekty daje połączenie technik przeglądania i nawigacji z klasyfikacją fasetową, czyli tzw. **przeglądanie fasetowe** (*faceted browsing*), zwane też aspektowym. Dzięki temu połączeniu umożliwia się użytkownikowi przeglądanie i wyszukiwanie przy użyciu pojedynczego interfejsu. Wyniki wyszukiwania grupowane są według ustalonych kategorii (faset). Poszczególne fasety wyświetlane są tylko w przypadku, gdy zawierają wyniki wyszukiwania, więc użytkownik ma możliwość szybkiej orientacji w układzie wyników i nie trafia na puste linki podczas przeglądania. Wyniki wyszukiwania mogą także być w obrębie faset pogrupowane przy pomocy rankingu wyników.

Wyszukiwanie zarówno za pomocą słów kluczowych tworzonych automatycznie, jak i metadanych tworzonych przez człowieka polega, w pewnym uproszczeniu, na wyszukiwaniu (porównywaniu) prostych ciągów znaków, składających się na przykład na nazwiska autorów. Typowy system wyszukiwania informacji użytkownikowi (człowiekowi) pozostawia decyzję wyboru odpowiedniego ciągu znaków służących wyszukiwaniu, a także takie przeprowadzenie wyszukiwania (oraz interpretację wyników), aby odróżnić nazwę osobową, traktowaną jako hasło autorskie, od tej samej nazwy, stosowanej jako hasło przedmiotowe.

Wyszukiwanie wspomagane ontologią (zob. p. 2.7) daje możliwości nie tylko mechanicznego porównywania ciągów znaków, ale także wyszukiwania obiektów powiązanych relacjami semantycznymi. Dzięki temu możliwe jest tworzenie inteligentnych serwisów GBC, które pozwalają traktować Web jako kolekcję różnorodnych zasobów, posiadających dobrze zdefiniowane interfejsy, służące udostępnianiu usług informacyjnych. Inaczej mówiąc, inteligentne serwisy GBC w Webie wykorzystują techniki systemów inteligentnych do efektywnej realizacji powtarzalnych zadań użytkowników Webu. W tym celu należy posiadać opis serwisu GBC: możliwości, funkcji interfejsów i wyników pracy, zapisanych w sposób jednoznaczny, czytelny dla komputera i kodowany, dzięki czemu

umożliwia się współdziałanie tych serwisów [Gašević, Djurić, Devedžić 2006, s. 101]. Opisy te zawarte są w ontologiach serwisów Web. Dzięki integracji ontologii i agentów Web wzrasta inteligencja serwisów. Agent, realizujący wyszukiwanie według słów kluczowych, może na przykład wywoływać usługi, takie jak słowniki kontrolowane, pozwalające na wyszukiwanie rozmyte. Ontologie mogą także zastąpić fizyczną agregację danych w jedną bazę danych, stosowane w wyszukiwaniu sfederowanym. Stanowią one wówczas system mediacyjny, funkcjonujący jako pojedynczy interfejs użytkownika do wielu zasobów [Lourdi, Papatheodorou, Doerr 2009].

Usługi inteligentnego Webu znacznie ułatwiają pracę użytkownika informacji. Pozwalają na automatyczne i natychmiastowe wyszukiwanie serwisów dzięki semantycznemu kodowaniu stron Web oraz zastosowaniu wyszukiwarek, wykorzystujących ontologie. Wykonanie usługi realizowane jest automatycznie przez agenta działającego w imieniu użytkownika, dzięki temu, że agent, na podstawie semantycznego kodu wie, co powinno być na wejściu systemu, jakie dane zostaną zwrócone i w jaki sposób należy je przetwarzać oraz w jaki sposób można automatycznie z tymi usługami współpracować. Możliwe jest też automatyczne łączenie i współdziałanie serwisów, gdyż semantyczne kodowanie usług dostarcza wszystkich informacji niezbędnych do selekcji, kompozycji i komunikowania się z usługami.

Według Deborah McGuinness zastosowanie ontologii do wyszukiwania informacji może dotyczyć:

- Współdziałania: dzięki posiadaniu zestawu terminów oraz zdefiniowanych relacji pomiędzy terminami można zastosować aksjomaty lub mapowanie służące porównywaniu terminów w celu bardziej precyzyjnego wyrażenia zakresu terminu na tle relacji z innymi terminami, co pozwala na osiągnięcie bardziej „inteligentnego” współdziałania;
- Wyszukiwania ustrukturyzowanego, porównawczego i kustomizowanego: jeżeli ontologia zawiera informacje kodowane, może być stosowana do wyszukiwania porównawczego (pozwalającego na porównanie wybranych cech wyszukiwanych obiektów) i wskazania cech najbardziej przydatnych do prezentacji w analizie porównawczej, dzięki czemu użytkownicy mogą dysponować zwięzłym opisem źródeł, zawierającym cechy najbardziej ich interesujące, zamiast otrzymywać pełne opisy na zbyt szczegółowym poziomie. Zestawy cech istotnych z punktu widzenia użytkownika mogą być także podstawą do tworzenia formularzy, służących formułowaniu zapytań, a więc wpływać na wygląd interfejsu.
- Uogólniania/uszczegółowiania: jeżeli narzędzie wyszukiwawcze zauważy, że zapytanie użytkownika generuje zbyt wiele/za mało odpowiedzi, możliwe jest sprawdzenie terminów wyszukiwawczych w ontologii i zaproponowanie terminu pozwalającego na zawężenie/rozszerzenie wyszukiwania [McGuinness 2002, s. 180].

Ontologie w wyszukiwaniu mogą być stosowane do wnioskowania o ukrytych zależnościach pomiędzy pojęciami. Na ich podstawie modyfikacji mogą ulegać zapytania użytkowników, po czym zasoby obiektów są powtórnie przetwarzane [Yang, Yang, Yuan 2007, s. 240]. W wyniku do zasobów dołączane są bardziej formalnie zdefiniowane informacje semantyczne.

Podjęcie takie zastosowano w funkcjonujących serwisach Semantycznego Webu (zob. p. 4.9), łączących ontologie z obszernymi bazami danych o znanej strukturze, jak na przykład Wikipedia. Jednym z nich jest Freebase⁴⁹, który przetworzoną informację udostępni zarówno dla ludzi, jak i wyspecjalizowanego oprogramowania. Na tych danych działają wyszukiwarki semantyczne typu Powerset⁵⁰, umożliwiające tworzenie zapytań w języku naturalnym i dostarczające konkretnej odpowiedzi, a nie wykazu stron, mogących zawierać elementy tej odpowiedzi.

⁴⁹ <http://www.freebase.com/>.

⁵⁰ <http://www.powerset.com/>.

Z opisu skomplikowanych procesów wyszukiwania informacji, realizowanych w GBC, wydaje się wynikać tendencja do **integracji narzędzi i technik wyszukiwania informacji** wokół wyszukiwarek internetowych [Hull, Pettifer, Kell 2008]. Odgrywają one coraz większą rolę, stając się głównym narzędziem dostarczania usług globalnej biblioteki cyfrowej. Integracja polega na tym, że wyszukiwarki internetowe umożliwiają użytkownikom dostęp do zróżnicowanych usług i zasobów z jednego, przyjaznego im miejsca – interfejsu wyszukiwarki globalnej. Wyszukiwarki zaspokajają różne potrzeby informacyjne, zarówno użytkowników poszukujących konkretnej informacji o znanych sobie faktach, jak i tych, którzy szukają informacji na wybrany temat, w tym przypadku wyszukiwarka często skierowuje do odpowiedniego serwisu GBC. Im większa część potrzeb informacyjnych może być zaspokojona w Webie za pomocą wyszukiwarek, tym bardziej zaciera się różnica pomiędzy zasobami i usługami, a użytkownicy przestają interesować się odrębnościami serwisów i ich zindywidualizowanych interfejsów, traktując je wszystkie jako jednolite środowisko informacyjne – globalną bibliotekę cyfrową.

Opisane procesy i kierunki rozwoju, pomimo wciąż występujących zjawisk negatywnych [Cisek, Sapa 2007, s. 45], sugerują dalszy wzrost znaczenia wyszukiwarek w dostarczaniu usług globalnej biblioteki cyfrowej. Wyszukiwarki indeksują coraz więcej treści, które wcześniej pozostawały pod wyłącznym nadzorem katalogów bibliotecznych oraz licencjonowanych baz danych. Zasoby metadanych nabierają nowego znaczenia: stają się cennymi zbiorami danych i informacji zarówno o obiektach cyfrowych, jak i dokumentach tradycyjnych (posiadających cyfrowe metadane), bezpłatnych i komercyjnych. Dzięki nim, w przypadku braku kopii cyfrowej, wyszukiwarka może wskazać miejsca przechowywania egzemplarzy drukowanych.

Najbardziej zaawansowanym projektem, idącym w tym kierunku wydaje się Google Scholar, specjalizujący się w indeksowaniu materiałów naukowych w Webie, po raz pierwszy oferujący jednolity dostęp do zasobów zarówno bezpłatnych, takich jak obiekty cyfrowe w repozytoriach i artykuły w czasopismach naukowych Open Access, jak i zasobów komercyjnych, umieszczanych na stronach wydawców i dostawców [Ross, Pongracz 2008, s. 148]. Indeksowane są także cytowania odnalezione w tych dokumentach, co zwiększa dostępność materiałów cytowanych. Dzięki porozumieniu między Google Scholar a wydawcami publikacji naukowych i organizacją nadzorującą projekt CrossRef, indeksowane są spisy treści publikacji komercyjnych. W wyniku powstaje indeks cytowań wszystkich publikacji dostępnych w Internecie. Bardzo istotną rolę odegrać może porozumienie zawarte z OCLC, dzięki któremu indeksowana jest także zawartość światowego katalogu WorldCat, z którym współpracuje polski Nukat. Pozwala to na umieszczanie opisów z katalogów bibliotecznych (w formie charakterystycznej dla Google) wśród wyników wyszukiwania.

Google Scholar umożliwia powiązanie wyników wyszukiwania z resolverem OpenURL, więc użytkownicy mogą uzyskiwać dostęp do opisów wyszukanych przez siebie książek i artykułów poprzez resolver ich biblioteki uczelnianej. W ten sposób Google Scholar umożliwia umiejscowienie zasobów bibliotecznych przy użyciu wyszukiwarki internetowej. Wyszukiwarka stała się alternatywą dla prenumerowanych, naukowych baz danych, dotychczas udostępnianych przez te biblioteki, chociaż istnieją do niej pewne zastrzeżenia, jak mniejsza aktualność i możliwości wyszukiwawcze Google Scholar w stosunku do tradycyjnych baz danych [Dahl, Banerjee, Spali 2006, s. 147]. Część autorów wiąże ten drugi problem z brakiem ustrukturyzowanych metadanych dla indeksowanych obiektów [McGregor 2008] lub błędami podczas automatycznego tworzenia metadanych [Nunberg 2009]. Popołniają oni jednak błąd metodologiczny: udowadniają, że wyrażenie wyszukiwawcze, odpowiednie dla określonej bazy danych, nie da się wprost przenieść do wyszukiwarki, co jest prawdą (nie da się go też zwykle przenieść do innej bazy danych). W wyszukiwarce należy utworzyć inne wyrażenie wyszukiwawcze i nie

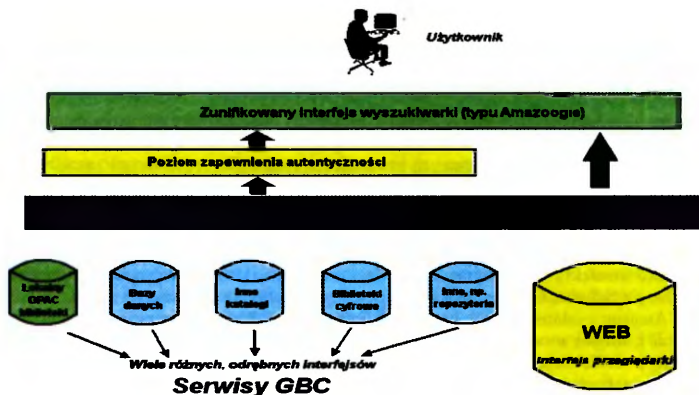
należy oczekiwać takich samych wyników, gdyż wyszukiwarka przeszukuje środowisko otwarte. Nie oznacza to jednak, że wyszukiwanie w wyszukiwarce jest gorsze, a tylko inne [Woźniak-Kaspepek 2008, s. 123]. W czasie korzystania z bazy danych użytkownik musi wejść w rolę indeksującego, aby domyślić się, jakich użyłby on terminów dla opisu zapytania. Podczas korzystania z wyszukiwarki użytkownik w ten sam sposób musi wejść w rolę autora, aby domyślić się, jakich słów użył on w swoim dziele.

Dotychczas istniało wyraźne rozróżnienie pomiędzy serwisami wyszukiwawczymi działającymi w oparciu o zasoby metadanych ustrukturyzowanych, a serwisami opartymi o indeksowanie pełnotekstowe (Google, Yahoo i innych). Wraz ze wzrostem liczby materiałów dostępnych w formie cyfrowej, użytkownicy oczekują systemów łączących oba podejścia: wyszukujących metadane zarówno o dokumentach analogowych, jak i obiektach cyfrowych, dając w tym drugim przypadku także dostęp do samego obiektu [Slone 2005, s. 658]. Dla realizacji tych oczekiwań różnicowane systemy wyszukiwania informacji o dokumentach tradycyjnych i cyfrowych muszą zostać zintegrowane (przynajmniej do czasu pełnej digitalizacji publikacji drukowanych); tego zadania podjęli się twórcy wyszukiwarek, indeksujących i tworzących rankingi rozproszonych zasobów. Po odszukaniu potrzebnych metadanych o dokumencie w wyszukiwarce, użytkownicy mogą być przeniesieni w odpowiednie miejsce (katalog centralny i/lub biblioteki lokalnej, księgarnia, inny serwis GBC), pozwalające na dostęp do wyszukanego dokumentu w każdej dostępnej formie (drukowanej lub cyfrowej).

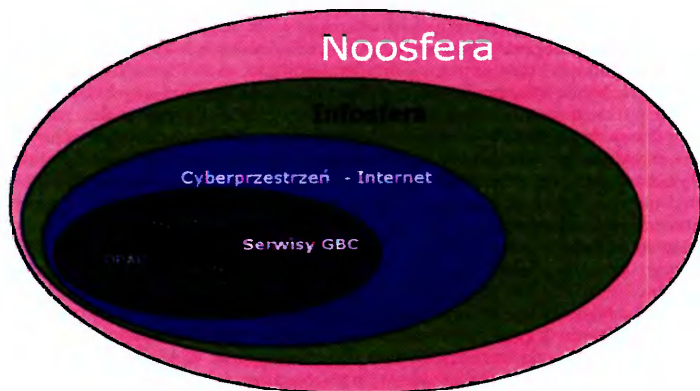
Właściwe efekty daje integracja narzędzi i serwisów takich, jak wyszukiwarki globalne, Google Scholar, Google Books i inne zasoby materiałów zdigitalizowanych, katalogi w Amazon i odnośniki Wikipedii z bibliotecznymi OPAC, bibliograficznymi bazami danych i innymi serwisami GBC. Dzięki współpracy twórców wyszukiwarek internetowych z właścicielami najpotężniejszych światowych zasobów metadanych (głównie OCLC i wielkimi bibliotekami naukowymi) oraz stosowaniu usług identyfikacji, powstaje nowy sposób korzystania z serwisów GBC w Internecie, jako jednolitym środowisku informacyjnym (rys. 9). Dzięki integracji wielu zestawów metadanych, pochodzących z różnych miejsc, użytkownik otrzymuje pełny obraz interesujących go zasobów [Library of Congress 2008]. Zasoby te gromadzone i opracowywane są lokalnie, jednak informacja o nich dostępna jest globalnie. Na wielu poziomach następuje agregacja treści obiektów cyfrowych i metadanych o tych obiektach, w połączeniu z metadanymi dokumentów tradycyjnych. Jednym z takich interfejsów jest katalog WorldCat, który w sposób typowy dla wyszukiwarek i księgarni internetowych przedstawia światowe zasoby bibliotek tradycyjnych. Ponad tym katalogiem stoi jednak wyszukiwarka, która scala wszystkie zasoby w jednym indeksie, będącym wielkim zasobem różnego rodzaju metadanych. Tak więc to wyszukiwarka udostępnia uniwersalny interfejs, przy pomocy którego użytkownik uzyskuje dostęp do wszystkich interesujących go zasobów, tradycyjnych (tu ustrukturyzowane metadane i informacja o warunkach dostępu) i cyfrowych (tu bezpośredni dostęp do obiektu, czasem po wniesieniu opłaty). Użytkownik posiadający biblioteczne hasło lub IP może przeszukiwać i uzyskiwać dostęp do zasobów nie zdejmuje nawet sprawy z istnienia biblioteki i jej roli w tych procesach.

Użytkownik nie rozpoczyna już wyszukiwania od OPAC, nawet od portalu bibliotecznego, lecz od wyszukiwarki. Właśnie ona pozwala mu na bezpośredni dostęp do poszukiwanego obiektu lub w razie potrzeby kieruje go do odpowiedniego serwisu GBC, na przykład biblioteki – cyfrowej lub tradycyjnej (do jej OPAC) – albo strony wydawcy, czy innego. Potwierdzają to badania użytkowników polskich bibliotek cyfrowych, którzy w 44% przypadków do serwisu kierowali byli przez wyszukiwarkę Google [Mazurek 2009]. Biblioteka wraz ze swoim systemem bibliotecznym i OPAC „zanurzona” jest w globalnym środowisku informacyjnym, zamiast stać obok niego, tak jak działo się to do tej pory; staje się częścią obszernej platformy Web, gdzie udostępniane dane stają się

podstawą tworzenia nowych usług i serwisów. Sytuacja ta przedstawiona jest na rys. 10: Serwisy GBC są częścią cyberprzestrzeni, którą w dużym stopniu utożsamiać można z Internetem: jest to globalny system komputerów spiętych siecią, urządzeń infrastruktury komunikacyjnej i zasobów informacyjnych. W cyberprzestrzeni znajdziemy także takie infrastruktury, jak intranety i inne zasoby „ukryte”. Ma ona zazwyczaj większe znaczenie technologiczne, niż kolejne warstwy: infosfery i noosfery, jednak cyberprzestrzeń to nie tylko miejsce funkcjonowania technologii; zauważana jest jej rola społeczna.



Rys. 9. Integracja wielu zasobów informacyjnych



Rys. 10. Serwisy GBC w obrębie kolejnych sfer infrastruktury informacji

Termin infosfera bywa stosowany zamiennie z cyberprzestrzenią, jednak zwykle infosfera rozumiana jest jako coś znacznie większego niż Internet, gdyż oprócz niego obejmuje systemy informacyjne pozostające poza Siecią i nie stanowiące jej części. Są to między innymi tradycyjne biblioteki naukowe i ich zasoby publikacji drukowanych. Infosfera wraz z cyberprzestrzenią mogą być traktowane jako globalna infrastruktura informacyjna [Ronfeldt, Arquilla 2007], stanowiąca podstawę funkcjonowania GBC.

Noosfera jest najbardziej abstrakcyjnym pojęciem. Termin ten pochodzi od greckiego słowa *noos*, co oznacza „rozum, umysł”. Jest to pojęcie z zakresu inteligencji zbiorowej. Dzięki możliwości komunikacji w skali globalnej, powstaje globalny obszar umysłu, sfera myśli i ducha, niezwykła myśląca maszyna, jak sferę tę nazywał Pierre Teilhard de Chardin w wielu swoich książkach (np. [Teilhard 1993]). Julian Huxley noosferę nazywał „siecią żywej myśli” [Ronfeldt, Arquilla 2007]. Zgodnie z tymi ideami spotkanie pojedynczych umysłów ludzkich, które dokonuje się za pomocą Internetu, prowadzi do wykształcenia się Sieci, czyli zbiorowej inteligencji oplatającej świat [Juza 2007, s. 45]. Idea inteligencji zbiorowej jest obecnie wykorzystywana praktycznie w funkcjonowaniu serwisów Web 2.0 (zob. p. 4.9).

Dzięki zintegrowanemu systemowi wyszukiwania informacji, bazującemu na wyszukiwarce internetowej, możliwe jest już obecnie:

- Wyszukiwanie informacji o publikacjach tradycyjnych: poprzez Google Scholar w połączeniu z WorldCat (oraz katalogami centralnymi współpracującymi z nim, na przykład Nukat) użytkownik otrzymuje informacje o dostępności publikacji tradycyjnej we wskazanej przez niego bibliotece naukowej. W zależności od sposobu funkcjonowania systemu biblioteki lokalnej istnieje możliwość zamówienia wyszukanej pozycji.
- Wyszukiwanie informacji o stronach Web poprzez Google, wraz z bezpośrednim dostępem do tych stron.
- Wyszukiwanie informacji o zdigitalizowanych wersjach publikacji, udostępnionych w trybie Open Access, wraz z bezpośrednim dostępem do tych publikacji.
- Wyszukiwanie informacji o cyfrowych wersjach publikacji komercyjnych, wraz z dostępem do nich poprzez strony wydawców komercyjnych, po spełnieniu odpowiednich warunków, na przykład wykupieniu dostępu typu „pay per view”.
- Wyszukiwanie w Google Scholar przy użyciu CrossRef artykułów wraz z artykułami cytowanymi (poprzez bibliografię załącznikową), co pozwala na stworzenie indeksu cytowań artykułów bez względu na miejsce ich przechowywania w Internecie.
- Google Scholar daje możliwość połączenia wyników wyszukiwania z resolverem OpenURL. Dzięki temu użytkownicy mogą uzyskiwać dostęp do obiektów przedstawionych w wynikach wyszukiwania poprzez resolver OpenURL własnej biblioteki naukowej.

Wyszukiwarki typu Google Scholar można traktować jako implementację wyszukiwania sfederowanego. Wyszukiwarka ta umożliwia dostęp do treści stanowiących typowy zasób dla tego wyszukiwania, takich jak pełne teksty artykułów, metadane i abstrakty, poprzez jeden indeks i jeden interfejs wyszukiwawczy. Prezentacja wyników, posortowanych za pomocą algorytmu rankingowego, przez wyszukiwarkę internetową jest przy tym szybsza niż przez metawyszukiwarkę. W ten sposób Google Scholar stał się konkurencją zarówno dla systemów wyszukiwania informacji w repozytoriach, jak i wyszukiwania sfederowanego realizowanego przez biblioteki akademickie. Łączy on indeksowanie i wyszukiwanie pełnotekstowe z wyszukiwaniem na podstawie standardowych metadanych, tworzonych w repozytoriach zgodnie ze standardem OAI. Jednym z obiecujących sposobów wykorzystania metadanych OAI jest wyszukiwanie nie tylko na poziomie poszczególnych obiektów cyfrowych, ale głównie na poziomie kolekcji [Mischo 2005], co ułatwia odszukanie właściwych serwisów GBC.

Już obecnie odczuwalny jest wpływ sposobu funkcjonowania takich wyszukiwarek, jak Google i Google Scholar na sposób funkcjonowania serwisów globalnej biblioteki cyfrowej. Znaczna część ruchu w serwisach GBC generowana jest właśnie przez te wyszukiwarki. Skierowują one użytkownika wprost do pełnego tekstu z pominięciem metadanych (zarówno tekst, jak i metadane dostępne są do indeksowania). Powoduje to pewne konsekwencje. Ponieważ użytkownicy stosują głównie Google jako narzędzie wyszukiwawcze, metadane tworzone przez bibliotekarzy nie są przez nich bezpośrednio wykorzystywane do wyszukiwania. Oznacza to, że tworzenie metadanych, w szczególności metadanych o tradycyjnej, bibliotecznej jakości, w inny sposób niż dotychczas służy temu procesowi; są one indeksowane, tak jak każdy inny tekst i stają się częścią zasobów indeksowych wyszukiwarki. Tego typu metadane mają coraz większy udział w wyszukiwaniu informacji.

Metadane o wysokiej jakości mają obecnie inne niż dotąd zastosowanie, gdyż są stosowane przez inne grupy użytkowników. Ich użytkownikami są bibliotekarze, którzy po pierwsze wiedzą o ich istnieniu, a po drugie rozumieją ich strukturę. Można więc stwierdzić, że metadane, tworzone przez bibliotekarzy są bezpośrednio wykorzystywane głównie przez nich samych³¹. Dotyczy to zarówno metadanych tradycyjnych, typu rekordów MARC, jak i nowych, tworzonych w nowopowstających standardach.

W tym kontekście rozumiale jest znaczenie możliwości dostępu w Internecie do pełnych tekstów dokumentów. Jednym ze sposobów na zwiększenie liczby dostępnych tekstów jest realizacja idei Open Access. Inną grupą projektów realizujących podobne cele, w których biorą udział największe światowe wyszukiwarki komercyjne, jest digitalizacja materiałów drukowanych, przechowywanych w największych bibliotekach świata. Pomimo różnic w oczekiwaniach i planowanych rezultatach projektów digitalizacyjnych (biblioteki naukowe chcą zwiększyć dostępność swoich zbiorów jako dobra publicznego, firmy chcą zwiększyć zasób dostępny poprzez oferowane przez nie narzędzia – wyszukiwarki), wydaje się, że korzyści z realizacji tych projektów mogą być obopólne. Co znamienne, bardzo ostro przeciw tym inicjatywom wystąpili wydawcy komercyjni. Równoległe do tych przedsięwzięć wymieniać można jeszcze dwa rodzaje inicjatyw: ekspansję księgarni internetowych oraz rozwijającą się sieć bibliotecznych wypożyczalni cyfrowych [Franke 2007, s. 124].

Możliwości techniczne indeksowania i wyszukiwania informacji zbliżają nas do idealnego celu pełnej dostępności informacji, formułowanej już przez twórcę OCLC, Fredericka Kilgoura w latach osiemdziesiątych XX wieku [Pomerantz 2006, s. 52]. Rozumienie pełnej dostępności szybko uległo jednak zmianie: początkowo oznaczało starania o stałą dostępność publikacji tradycyjnych na półce, w oczekiwaniu na żądanie użytkownika, później zapewnienie możliwości sprawdzenia stanu posiadania biblioteki za pomocą OPAC i katalogów centralnych bez potrzeby jej odwiedzania, obecnie raczej zdalną dostępność do obiektów cyfrowych, tekstowych i multimedialnych.

Paradoksalnie jednak Google, poprzez zastosowanie nowych technologii, może spowodować wzrost częstotliwości odwiedzin użytkowników w bibliotece (tradycyjnej). W bliskiej przyszłości, dzięki masowej digitalizacji zasobów tradycyjnych, Google umożliwi użytkownikom poznanie zasobów, dostarczanych przez biblioteki naukowe. Przynosi to korzyści zarówno dla Google, jak i dla bibliotek i ich użytkowników: biblioteki zyskują, przyjazną dla użytkownika, drogę przedstawienia swojej oferty. Możliwe są także inne obszary współpracy, na przykład wspólna archiwizacja digitalizowanych zasobów [LIBER 2009].

Z jednej więc strony Google i podobne narzędzia dostępne online, wywołują obawy odebrania bibliotekom użytkowników. Jednakże z drugiej strony, te same narzędzia stwarzają perspektywę powstawania nowych usług bibliotecznych; wzrostu dostępności

³¹ Nie oznacza to oczywiście, że nie są one lub nie mogą być stosowane przez innych użytkowników; chodzi tu raczej o statystyczną tendencję.

usług dla nowych użytkowników i ich grup; bliższego współdziałania z użytkownikiem w procesie wyszukiwania; udostępniania w formie cyfrowej zasobów wysokiej jakości wszystkim użytkownikom, zaspokajając wszystkie potrzeby informacyjne; wzrostu zainteresowania ofertą biblioteki, również w tradycyjnej formie.

Trwają dalsze prace nad doskonaleniem technik wyszukiwania, zarówno sposobu funkcjonowania wyszukiwarek, jak również łączenia funkcjonalności narzędzi. W przyszłości dla technik wyszukiwania istotne znaczenie może mieć połączenie kilku elementów, już obecnie istniejących i stosowanych, ale rozwijanych odrębnie. Są to:

- Nowe wyszukiwarki, pozwalające na wyszukiwanie otwarte (*open search*)⁵².
- Wyszukiwanie sfederowane.
- Przetwarzanie rozproszone.
- Przetwarzanie Gridowe.
- Środowisko społecznościowe, znane z Web 2.0.

Połączenie inteligencji zbiorowej z nowoczesną technologią wyszukiwania i współpracą w niewielkich społecznościach może spowodować trwałe zmiany w technice wyszukiwania informacji. W efekcie mogą powstać nowe narzędzia wyszukiwawcze, rozwijają się też już istniejące, typu Google, czego przejawem są obecnie prowadzone prace w kierunku wyszukiwania uniwersalnego, czyli jednolitego wyszukiwania obiektów różnego rodzaju: tekstu, dźwięku i obrazu [Evans, Woody 2008, s. 14].

4.4. Zapewnienie integralności i autentyczności obiektów cyfrowych

Chociaż potrzeba tworzenia architektury zaufania, dającej możliwość bezpiecznych i prywatnych transakcji, wiązana jest z rozwojem e-handlu [Barney 2008, s.70], to również w nauce odgrywa istotną rolę. Architektura taka polega na budowaniu protokołów autentyczności i autoryzacji (zob. nast. punkt), co z kolei oznacza nadzór nad siecią.

Jak stwierdził Clifford Lynch, autentyczność i integralność obiektów cyfrowych są kontrowersyjnymi pojęciami filozoficznymi, które w skomplikowany sposób dotyczą konceptualizacji dokumentów oraz ich znaczenia i roli prawnej, społecznej, kulturalnej i historycznej [Lynch 2000, s. 32]. Uwierzytelnienie i zapewnienie integralności są dwiema dziedzinami, w których technologia cyfrowa styka się z zagadnieniami filozoficznymi i społecznymi. Część działań w tym zakresie może być wykonywana automatycznie, inne powinny pozostać domeną ludzkiego osądu.

W najszerszym znaczeniu autentyczność może być definiowana nie tylko jako uwierzytelnienie oznaczające weryfikację autorstwa, ale także szerzej, gdyż zawiera zagadnienia integralności, kompletności, poprawności, walidacji, podobieństwa do oryginału, a także odpowiedniości dla założonych celów. Autentyczność obiektów cyfrowych ma zasadnicze znaczenie dla wymogów technicznych stawianych przed systemami archiwizacji tych obiektów [Rothenberg 2000, s. 52].

Integralność obiektu cyfrowego oznacza, że w trakcie istnienia nie był on uszkodzony, na przykład podczas transmisji. Dysponujemy więc tym samym zestawem i sekwencją bitów, które powstały, gdy obiekt był tworzony. Algorytmy potwierdzające integralność obiektu badają jego reprezentację, a nie ciągi bitów, z których obiekty są tworzone. Jed-

⁵² Wyszukiwanie otwarte opiera się na otwartej i transparentnej technologii (*open source*), co jest zaprzeczeniem zasad obowiązujących w Google. Przykładem realizacji tej idei jest wyszukiwarka Wikia Search (<http://re.search.wikia.com/index.html>), w której obowiązują cztery zasady: przezroczystość (pełny dostęp do opisu algorytmów), społeczność (współpraca internautów na zasadach stosowanych w Web 2.0), jakość (wysoka efektywność wyszukiwania), prywatność (ochrona danych poufnych).

nym ze sposobów badania integralności jest porównanie posiadanego obiektu z jego egzemplarzem, który uznawany jest za „prawdziwy”. Jednak, jeżeli posiadamy bezpieczne łącze do znanego, prawdziwego egzemplarza, możemy po prostu pobrać jego kopię, nie martwiąc się o poprawność posiadanego egzemplarza. Szczególna sytuacja powstaje, gdy właśnie chodzi o stwierdzenie, że posiadany egzemplarz jest poprawny, na przykład dla wykrycia prób oszustwa.

W przypadku braku bezpiecznego dostępu do niezależnie zarządzanego egzemplarza obiektu, znanego jako prawdziwy (lub przynajmniej jego surogatu), test integralności ograniczany jest do kontroli wewnętrznej zgodności. Jeżeli obiekt jest wyposażony w autoryzowany skrót (z podpisem cyfrowym), można sprawdzić zgodność obiektu z tym skrótem, poprzez powtórne utworzenie skrótu z posiadanego obiektu i porównanie obu. Wówczas jednak pewność co do integralności obiektu jest tym samym, co pewność autentyczności i integralności skrótu. Nastąpiła tylko zmiana punktu widzenia: teraz możemy ufać integralności obiektu, jeżeli skrót jest autentyczny i poprawny.

W środowisku cyfrowym istotną rolę odgrywa także zjawisko, które nazwać można integralnością sytuacyjną, czyli integralnością dzieł pochodnych. Wówczas badane jest, czy posiadany obiekt zawiera właściwą transkrypcję, tłumaczenie lub rozdzielczość w stosunku do oryginalnego dzieła. Tutaj należy zająć się dwoma zagadnieniami: jedno dotyczy integralności obiektu podstawowego, a drugie poprawności przeprowadzonych transformacji na obiekcie⁵³. Sprowadza się to zazwyczaj do określenia wiarygodności źródła lub dostawcy transformacji, a więc do zagadnień dotyczących uwierzytelnienia źródła lub do walidacji, integralności i poprawności kodu.

Pojęcie autentyczności wywołuje wiele dyskusji w gronie specjalistów w zakresie informacji naukowej. Dla archiwistów dokument „autentyczny” to albo dokument oryginalny, niezmienny lub jego dokładna, certyfikowana kopia. Dla autora dokument jest wyrażeniem jego poglądów, opinii i założeń. W takim znaczeniu dokument jest autentyczny, ponieważ dotyczy cech identyfikujących daną osobę jako autora⁵⁴. Dokumentem autentycznym jest w tym przypadku ten, który odzwierciedla aktualne opinie autora, wyznawane idee i tworzone założenia, nie zaś ten, w którym zawarte zostały dawne, nieaktualne poglądy, od których być może sam autor się dystansuje.

Pracę umysłową można traktować jako działalność ciągłą, czyli proces. Dokumenty wówczas są traktowane jako odzwierciedlenie realizacji tego procesu i mogą być tworzone oraz zmieniane wraz z postęпами w jego realizacji. Z takiego punktu widzenia nie można właściwie wskazać wersji „ostatecznej” czy „autentycznej”. Z drugiej strony jednym z dogmatów publikowania naukowego jest twierdzenie o publikacji jako „oficjalnym zapisie” umieszczonym w „kanonicznym archiwum” nauki. Sposób rozumienia pojęcia autentyczności i kierunek zmian w tym zakresie może mieć duży wpływ na rolę publikacji w procesie naukowym oraz na metodę zapisu i archiwizacji wyników badań naukowych.

Dla obiektów umieszczanych w zasobach serwisów GBC określenie autentyczności może oznaczać weryfikację twierdzeń dotyczących obiektu, a w efekcie weryfikację tego, że obiekt w rzeczywistości jest tym, czym wydaje się być. Obiekt może na przykład zawierać stwierdzenia, dotyczące daty powstania lub autorstwa albo też posiadania nazwy

⁵³ Dla ścisłości zauważmy także, że należy wziąć również pod uwagę integralność wyników przeprowadzonych transformacji po ich wykonaniu. Technika pozwalająca na stwierdzenie, czy podstawowe charakterystyki obiektu w wyniku transformacji nie zmieniły się, nazywa się kanonizacją. W jej efekcie powstaje forma kanoniczna obiektu, niezmienna (zawierająca ten sam ciąg bitów) bez względu na wykonywane transformacje [Lynch 1999].

⁵⁴ Według Johna Mackenzie Owena na tym właśnie polega współczesne rozumienie pojęcia autorstwa, które tłumaczy, dlaczego publikacje nie są wydawane anonimowo [Owen 2006, s. 69]. Tłumaczy również powody, dla których plagiaty są tak surowo ścigane: nie polegają one na kradzieży informacji, lecz na zaprzeczeniu identyfikacji autora. Jeśli jednak dokument wyraża autora, to ten może chcieć go zmienić, jeżeli sam się zmieni. Takie postępowanie jest w środowisku elektronicznym dość łatwe, może być na przykład częścią autoarchiwizacji.

lub identyfikatora, nadanych przez określoną organizację. Testy potwierdzające autentyczność dotyczą tylko określonych stwierdzeń (typu: czy X jest autorem tego dokumentu), a nie realizują wyszukiwań na pytania otwarte (typu: kto jest autorem dokumentu). Walidacja autentyczności obiektu jest bardziej ograniczona niż otwarte zapytania. Jest to tym bardziej zrozumiałe, że obecnie wiele danych generowanych jest przez zautomatyzowane procesy i systemy (sensory, czujniki, kamery). W takim przypadku trudno o określenie autorstwa w tradycyjnym znaczeniu [Owen 2006, s. 66].

Istnieją dwie podstawowe strategie testowania stwierdzeń. Pierwszą jest zaufanie stwierdzeniu, gdy można zweryfikować jego integralność i uwierzytelnić jego źródło, które uznane zostało za godne zaufania. Innymi słowy waliduje się stwierdzenie typu „A jest autorem obiektu ze skrótem X” najpierw poprzez weryfikację integralności obiektu, którego dotyczy stwierdzenie (że posiada skrót X), a następnie przez skontrolowanie czy stwierdzenie jest uwierzytelnione (tzn. posiada podpis cyfrowy) przez zaufaną jednostkę. Tutaj zasadnicze znaczenie ma posiadanie pewności tego, czym w rzeczywistości jest zaufana jednostka, która tworzy lub gwarantuje stwierdzenie. Druga strategia może być nazwana niezależną weryfikacją stwierdzeń. Jeżeli na przykład istnieje wiarygodny, narodowy wykaz (kartoteka) autorów, możliwe jest sprawdzenie, czy dane z wykazu autorów zgodne są ze stwierdzeniem o autorstwie. Jednakże w obu przypadkach walidacja stwierdzenia związanego z obiektem ostatecznie oznacza podjęcie decyzji o zaufaniu jakiejś jednostce, tworzącej lub gwarantującej stwierdzenia.

W środowisku cyfrowym kontrola integralności może być uważana za szczególnie przypadki i etap oceny autentyczności. Częściowo wynika to z poglądu, że kontrola integralności obiektu oznacza porównywanie go z czymś precyzyjnie zidentyfikowanym i rygorystycznie zweryfikowanym, co nazwać można wersją oryginalną lub egzemplarzem autorytarnym. W świecie fizycznym natomiast kontrola integralności na ogół nie polega na wyszukiwaniu oszustw, ale raczej na akceptowaniu przedmiotów według pobieżnego oglądu, który wskazuje na to, czym wydają się one być, a następnie na poszukiwaniu dowodów uszkodzeń (na przykład wyrwanych kartek). Dla tego typu kontroli integralności efektywnym mechanizmem zapewnienia, że obiekt nie został uszkodzony, jest użycie skrótu komunikatu, który towarzyszy obiektowi cyfrowemu w formie metadanych. Jest to właściwa metoda nawet w przypadku, gdy skrót komunikatu nie jest wyposażony w podpis czy ocenę wiarygodności, lecz jedynie w ogólny poziom poufności środowiska cyfrowego, w którym obiekty są przechowywane i transmitowane. W środowisku cyfrowym istnieje tendencja do niezauważania potrzeby tego rodzaju kontroli integralności na korzyść silniejszych metod, łączących badanie autentyczności z kontrolą integralności.

W dyskusji nad autentycznością i integralnością obiektów cyfrowych zazwyczaj pojawia się także zagadnienie proveniencji: jest ona udokumentowaniem pochodzenia, charakterystyk oraz historii obiektu, jego kolejnych zarządców oraz relacji z innymi obiektami. To ostatnie jest szczególnie ważne. Obiekt cyfrowy, utworzony poprzez zmianę formatu innego, starszego obiektu, który był walidowany za pomocą pewnego określonego algorytmu, może być postrzegany w dwojaki sposób. Możemy uważać go za pojedynczy obiekt, którego proveniencja zawiera określoną transformację lub też za wiele obiektów, które są powiązane ze sobą relacjami opisanymi w dokumentacji proveniencji. Stąd proveniencja to nie tylko metadane dotyczące obiektu, ale także metadane opisujące relacje pomiędzy obiektami. Ponieważ proveniencja zawiera również stwierdzenia o obiektach, jest ona częścią uwierzytelnienia oraz infrastruktury zapewniającej wiarygodność.

Jak uważa Clifford Lynch, bardzo trudno jest wskazać właściwe miejsce przechowywania danych o proveniencji i organizację do zarządzania nimi [Lynch 2000, s. 42]. Nie jest nawet całkiem jasne, czy rekord proveniencji ma być tworzony niezależnie i na stałe, czy też ma powstawać w razie potrzeby z różnych elementów metadanych, obsługiwanych przez wiele systemów, stosowanych w serwisach GBC. Brak jest także schematu metada-

nych i struktur wymiany danych służących proveniencji. Pewne działania w tym kierunku umożliwia Dublin Core, dzięki istnieniu elementu metadanych *Provenance*. Element ten zastosowano między innymi w *Collecions Application Profile*, gdzie funkcjonuje element *Custodial History*, w którym można zawrzeć dane dotyczące zmian właściciela obiektu od czasu jego utworzenia, ważne z punktu widzenia jego autentyczności, integralności i interpretacji [Dublin Core Collection Description Task Group 2007].

Na szczęście istnieje stosunkowo niewiele zastosowań, dla których autentyczność ma znaczenie krytyczne. Są to głównie dane medyczne, dokumenty prawne, dotyczące bezpieczeństwa narodowego, informacja komercyjna i handlowa, dane osobowe. W tych przypadkach informacja jest zazwyczaj tworzona i zarządzana w sposób kontrolowany, dzięki czemu ryzyko zamierzonych czy przypadkowych uszkodzeń jest ograniczone do akceptowalnego poziomu. Do wielu zasobów obiektów cyfrowych mamy zaufanie, że są tym, na co wyglądają: dokumenty historyczne w Web, poczta elektroniczna otrzymywana od kolegów, fotografie umieszczane w serwisach społecznościowych czy wiadomości z portali informacyjnych. Nie chcielibyśmy żyć w świecie, w którym stale musielibyśmy podejrzewać, że to, co widzimy, nie jest tym, czym powinno być. Aby do tego nie dopuścić, określone środowiska, na przykład pracownicy nauki, dziennikarze, muszą podjąć decyzję, jakie informacje powinny mieć najwyższy poziom zaufania oraz stworzyć protokoły zapewniające integralność takiej informacji. Określenie prawdziwej wartości informacji zawsze jednak pozostanie domeną jej użytkowników, zarówno w środowisku informacji analogowej, jak i cyfrowej.

4.5. Dostępność informacji i autoryzacja dostępu

Funkcjonowanie GBC znacznie zwiększa dostępność publikacji naukowych, m.in. dzięki obniżeniu kosztów dostępu. W najprostszym przypadku dostęp do informacji w postaci cyfrowej zapewnić może użytkownikowi wykorzystanie skanera oraz poczty elektronicznej, za pomocą której przesyłany jest zeskanowany obraz dokumentu. Taki sposób udostępniania bardzo przypomina przesyłanie kserokopii za pomocą tradycyjnej poczty lub faksu: pozwala zaspokoić potrzeby tylko jednego użytkownika w danym czasie, jest jednak szybszy i często zapewnia lepsze parametry techniczne kopii [Holz 2006, s. 32]. Wystarczy jednak umieścić pliki powstałe w wyniku digitalizacji na stronie Web, aby zapewnić do nich dostęp każdemu zainteresowanemu o dowolnym czasie. Opublikowanie skanów jednego dokumentu nie zmienia funkcjonalności serwisu GBC. Znacznie poważniej wygląda sytuacja, gdy umieszczanie obiektów cyfrowych w Sieci jest planowe i systematyczne. Wówczas niezbędne jest podejmowanie istotnych decyzji, dotyczących wielu zagadnień, takich jak metadane, standaryzacja, jakość skanowania, współdziałanie, przygotowanie użytkowników, metody archiwizacji; zagadnienia te przedstawiłem w innych rozdziałach książki.

Christine Borgman wyróżniła kilka elementów, które decydują o dostępności informacji w globalnej bibliotece cyfrowej [Borgman 2003a, s. 56]. Podstawowym warunkiem jest zapewnienie łączności, niezbędnej dla pracy z komputerem w środowisku sieciowym, umożliwiającej wykorzystywanie zasobów i usług oferowanych przez to środowisko; przejawem dbałości o ten aspekt była realizacja projektów „autostrad informacyjnych” (*information superhighways*) w krajach rozwiniętych w latach dziewięćdziesiątych XX w. Kolejnym jest istnienie treści i usług gotowych do udostępnienia, bez których sama możliwość połączenia nie miałaby sensu. W dalszej kolejności należy uwzględnić użyteczność zasobów, o której mówić można na trzech poziomach. Po pierwsze, użyteczność dotyczy

sieci komputerowej, czyli infrastruktury informatycznej; ważnym aspektem jest umożliwienie wykorzystania sieci przez każdego, nie tylko przez specjalistów-informatyków. Obecnie ważne jest nie tylko uzyskanie połączenia, ale także tworzenie, a przynajmniej współtworzenie zasobów (Web 2.0). W tym celu powstają wyspecjalizowane aplikacje komercyjne. Potrzeby nauki w tym zakresie są nieco inne; od użytkowników spodziewać można się wyższych umiejętności, a typowa dla komercyjnych produktów elegancja i dopracowanie nie muszą być najważniejsze. Jest to aspekt związany z poziomem umiejętności, które każdy użytkownik wnosi do sieci. Dostęp może być ograniczony przez brak możliwości zastosowania niezbędnej technologii oraz przekształcenia dostępnej informacji w wiedzę. W GBC jest to o tyle istotne, że dostęp do niej, przynajmniej do istotnej części, jest powszechny (nie ograniczany). Ostatni element użyteczności dotyczy udostępnianych treści. Aby mogły one zostać uznane za „dostępne” muszą zostać odszukane w formie umożliwiającej ich odczyt, przeglądanie czy inne wykorzystanie. Obiekt cyfrowy, którego nie można otworzyć, wyświetlić na ekranie, uruchomić za pomocą posiadanego oprogramowania i sprzętu lub odczytać w znanym języku, nie może być uważany za w pełni dostępny.

Język jest jednym z najbardziej krytycznych czynników, wpływających na dostępność informacji. Jest ona użyteczna jedynie, gdy zarejestrowano ją w znanym użytkownikowi języku. Poważnym wyzwaniem dla projektantów GBC jest umożliwienie dostępu do jej zasobów bez względu na język treści i język, jakim posługuje się użytkownik, a więc stworzenie możliwości wielojęzycznego dostępu do informacji. Jest to problem, który można rozpatrywać z dwóch punktów widzenia: wielojęzyczne rozpoznawanie, manipulacje i wyświetlanie oraz wielojęzyczne wyszukiwanie. Pierwszy punkt widzenia dotyczy takich zagadnień, jak kodowanie znaków i symboli w sposób umożliwiający poprawną wymianę, wyświetlanie i drukowanie, przy jednoczesnym zapewnieniu możliwości manipulacji (sortowania, wyszukiwania). Drugi dotyczy wyszukiwania treści w obcych językach, nazywanego wielojęzycznym wyszukiwaniem informacji.

Po uwzględnieniu wszystkich tych przesłanek Christine Borgman zdefiniowała dostępność informacji jako możliwość uzyskania połączenia z siecią komputerową i dostępnymi w niej treściami, w taki sposób, że stosowana technologia jest przyjazna, użytkownik posiada odpowiednie umiejętności i wiedzę, a forma zasobów ułatwia ich wykorzystanie [Borgman 2003a, s. 57]. Każdy z tych czynników cechują rozległe implikacje dla projektowania i realizacji serwisów GBC. Ponieważ definicja jest dość ogólna, nie obejmuje części czynników mogących mieć istotny wpływ na dostępność. Nie uwzględniono na przykład możliwości ekonomicznych, które mają duże znaczenie w przypadku usług o zasięgu uniwersalnym; wyrazem tego są na przykład akcje wspomagające rozwój zasobów GBC w krajach rozwijających się. Nie ma także mowy o roli dostawców, pomimo istnienia licznych doświadczeń dotyczących Open Access. Według tych zasad dostawcą może być sam użytkownik informacji, gdyż w cyklu życia naukowej informacji cyfrowej granica pomiędzy twórcą, dostawcą i użytkownikiem staje się niewyraźna i zmienna.

Jedną z ról, zawsze przypisywanych bibliotekom, jest selekcja materiałów, w wyniku której kształtowana jest kolekcja optymalnie realizująca potrzeby użytkowników. W globalnej bibliotece cyfrowej selekcja ta nadal odbywa się w zakresie poszczególnych serwisów GBC i jest wykonywana przez bibliotekarzy, decydujących o tym, co i w jakiej kolejności digitalizować i publikować w Sieci (z uwzględnieniem praw autorskich). Użytkownicy korzystający z określonych serwisów szukają więc w zasobie preselekcjonowanym, na tych samych zasadach, jak w bibliotece tradycyjnej. Pozostaje problem korzystania z GBC na wyższym poziomie zasobów światowych, z natury rzeczy w żaden sposób nie selekcjonowanych, gdyż obejmujących całość dostępnego piśmiennictwa. Na tym poziomie selekcja odbywa się jednocześnie z wyszukiwaniem, w trakcie komponowania instrukcji wyszukiwawczej dla wyszukiwarki internetowej.

Nowe sposoby komunikowania naukowego, związane z narzędziami Nauki 2.0 (zob. p. 4.9) powodują zmiany w sposobach selekcji informacji w GBC. Zmienia się w szczególności sposób, w jaki użytkownicy identyfikują wiarygodne i wartościowe informacje i serwisy w Sieci. Do niedawna możliwe były dwa podejścia: albo wykorzystywano pośredników (specjalistów, bibliotekarzy, portale specjalistyczne), którzy wskazywali informację relewantną albo całkowicie rezygnowano z jakiegokolwiek pośrednictwa, na przykład samodzielnie wyszukując informację poprzez wyszukiwarki. Obecnie w Nauce 2.0 powstaje specyficzna forma pośrednictwa, zwana apomediacją⁵⁵ [Eysenbach 2008]. Użytkownicy mniej polegają na tradycyjnych ekspertach i ich autorytecie, a częściej używają narzędzi służących wspólnemu, społecznościowemu filtrowaniu informacji, co oznacza, że agenci (ludzie, oprogramowanie) „stoją obok” użytkownika, wskazując drogę ku informacji wysokiej jakości. Przykładami serwisów apomediacyjnych są Digg⁵⁶, CiteULike i Connotea⁵⁷. Serwisy te łącznie tworzą elementy tzw. otwartej nauki. W praktyce wszystkie trzy modele postępowania użytkownika mogą funkcjonować jednocześnie, przenikając się.

Serwisy globalnej biblioteki cyfrowej muszą posiadać silne systemy **autoryzacji i identyfikacji**, co pozwala na dostarczanie przyjaznych i zintegrowanych usług. Identyfikacja służy rozpoznawaniu użytkowników, czyli upewnianiu się, że ten, kto podaje się za jakąś osobę, jest nią w rzeczywistości. Autoryzacja jest procesem służącym określeniu, czy zidentyfikowany użytkownik posiada prawa dostępu do określonych zasobów lub realizacji określonych działań. Procesy te umożliwiają zindywidualizowanie funkcjonalności serwisu GBC i jego usług na poziomie poszczególnych użytkowników, co czyni je bardziej przyjaznymi i użytecznymi. Zazwyczaj wyróżnia się trzy poziomy zabezpieczeń:

- Tworzenie list użytkowników wraz z bezpośrednio przypisanymi im elementami danych, do których posiadają dostęp.
- Wyznaczenie poziomów bezpieczeństwa zarówno dla użytkowników, jak i danych. Poziomy tworzą strukturę kratową, w której zakodowane są relacje nadrzędności między nimi.
- Przypisanie ról użytkownikom – rolem przyporządkowuje się uprawnienia [Kodali, Farakas, Wijesekera 2004, s. 140].

Wraz ze wzrostem liczby zasobów sieciowych, udostępnianych przez serwisy globalnej biblioteki cyfrowej, szczególnego znaczenia nabiera stosowanie standardów umożliwiających tym serwisom określenie właściwego **poziomu dostępu** dla danego użytkownika. Zasady dotyczące autoryzacji, identyfikacji i zapewnienia bezpieczeństwa są uniwersalne w całym środowisku Web.

Użytkownicy żądają możliwości dostępu do usług z dowolnego miejsca. Wszystko, czego można od nich żądać, to podanie hasła i nazwy użytkownika, najlepiej jednego do dużej grupy usług, na przykład z zastosowaniem SSO⁵⁸. Każda zmiana nazwiska, adresu, hasła i innych danych identyfikujących powinna być dokonywana tylko w jednym, godnym zaufania miejscu. Należy zapewnić możliwość tworzenia profili użytkownika w różnych aplikacjach; profile te powinny uruchamiać się w każdym miejscu, w którym użytkownik zdecyduje się użyć tej aplikacji. Z jednej strony serwisy powinny być niezawodne i proste w użytkowaniu, a z drugiej autoryzowani użytkownicy powinni mieć pełną możliwość korzystania z ich usług. Systemy, w których przechowywana jest informacja

⁵⁵ Przyrostek apo- w łacinie oznacza oddzielnie, z dala, w oderwaniu.

⁵⁶ <http://digg.com/>.

⁵⁷ Wciąż przybywa nowych serwisów tego typu; przykładami mogą być Friendfeed (wymiana zasobów między znajomymi), Scilink (powiązania między uczonymi), MyExperiment (opisy eksperymentów badawczych), Scivee (filmy wideo: filmowe abstrakty artykułów, referaty, prelekcje) i wiele innych.

⁵⁸ SSO – Single sign on, usługa umożliwiająca dostęp do wielu różnych usług i systemów za pomocą jednego hasła (autoryzacji).

o użytkownikach, obiekty podlegające ochronie prawnej oraz dane finansowe, muszą być chronione przed włamaniami od wewnątrz i z zewnątrz systemu.

Większość oprogramowania stosowanego w serwisach GBC, oferującego personalizowane usługi, w tym OPAC biblioteczne, systemy wypożyczeń międzybibliotecznych, portale Web, została wyposażona w zdolność do przechowywania danych tego typu we własnym, wyodrębnionym zbiorze. Typowy system biblioteczny przechowuje dane dotyczące adresów i telefonów swoich użytkowników, często wraz z wykazem wcześniej wypożyczanych publikacji. Przechowywane są nazwy i hasła służące identyfikacji. Trudności może stwarzać synchronizacja tych elementów danych w wielu systemach. Ręczne dopisywanie użytkownika do wszystkich możliwych systemów jest uciążliwe i może być powodem dezintegracji danych w przyszłości. Rozwiązaniem może być dopuszczenie zewnętrznej identyfikacji i autoryzacji, wykonywanej różnymi metodami, na przykład zastosowanie odpowiedniego protokołu (LDAP, Shibboleth, Athens i in.).

Jak wynika z rozważań przedstawionych powyżej, jakość funkcjonowania poszczególnych usług serwisów GBC oraz wysokość kosztów i ilość czasu wydatkowanego na realizację tych usług, zależy w znacznym stopniu od efektywności zastosowanych systemów identyfikacji, autoryzacji i zarządzania informacjami na temat użytkowników. Dla zapewnienia bezpieczeństwa posiadanym zasobom, niezbędna jest znajomość ogólnych zasad funkcjonowania systemów oraz edukacja użytkowników. Jeżeli oprogramowanie jest oryginalne i na bieżąco aktualizowane, a użytkownicy świadomi niebezpieczeństw związanych z kopiowaniem niepewnych materiałów i podobnymi nieodpowiedzialnymi praktykami, wykonywane są standardowe procedury identyfikacji, a funkcjonowanie systemu regularnie monitorowane, to można oczekiwać poprawnej i stabilnej jego pracy (choć zapewne nie uda się całkowicie wyeliminować wszystkich negatywnych zjawisk).

4.6. Współdziałanie

Zdolność do współdziałania rozproszonych zasobów i systemów globalnej biblioteki cyfrowej ma dla jej funkcjonowania podstawowe znaczenie. Zagadnienia związane ze współdziałaniem dotyczą prawie wszystkich aspektów GBC, będąc podstawowym celem projektowania jej serwisów, zawierających zbiory niezależnie gromadzone, lecz uzupełniające się nawzajem.

Zgodnie z definicją Institute of Electrical and Electronics Engineers (IEEE), który jako stowarzyszenie fachowe uczestniczy w działalności normalizacyjnej, współdziałanie (*interoperability*) jest zdolnością dwóch lub więcej systemów lub składników do wymiany informacji i wykorzystywania wymienianej informacji⁵⁹. Można uogólnić tę definicję, mówiąc nie tylko o wymianie informacji, ale także wykorzystywaniu części systemu lub jego wyposażenia przez inny system. Podstawowe znaczenie dla utrzymania współdziałania ma szeroko pojęta standaryzacja. William Arms określa współdziałanie, w kontekście serwisów GBC, jako tworzenie usług spójnych dla użytkowników z elementów różnych pod względem technicznym i odrębnie zarządzanych [Arms 2000]. Współpraca różnych systemów informatycznych jest poważnym wyzwaniem. Web Service Interoperability Organization⁶⁰ (WS-I) podchodzi do tego zagadnienia z bardziej informatycznego punktu widzenia, zamierając promować współdziałanie serwisów GBC bez względu na stosowane platformy, systemy operacyjne i języki programowania. Organizacja ta definiuje termin „współdziałający” jako odpowiedni i zdolny do zaimplementowania w sposób neutralny w wielu systemach operacyjnych i językach programowania.

⁵⁹ <http://www.sei.cmu.edu/str/indexes/glossary/>.

⁶⁰ <http://www.ws-i.org>.

Christine Borgman przedstawia trzy główne zagadnienia, związane ze współdziałaniem:

- Posiadanie systemów współpracujących ze sobą w czasie rzeczywistym.
- Umożliwienie oprogramowaniu pracę w różnych systemach (przenoszalność).
- Umożliwienie wymiany danych pomiędzy różnymi systemami [Borgman 2003a, s. 213].

Wszystkie wymienione przez autorkę zagadnienia bezpośrednio wiążą się z funkcjonowaniem globalnej biblioteki cyfrowej, decydując o jej postrzeganiu przez użytkowników jako jednolitej struktury i zasobu.

Paul Miller w swoim rozumieniu tego terminu idzie nawet dalej: aby współdziałać należy być aktywnie zaangażowanym w realizację bieżących procesów, zapewniających zarządzanie systemami, procedurami i funkcjonowaniem organizacji w taki sposób, aby zmaksymalizować korzyści z wymiany i wielokrotnego wykorzystania informacji, zarówno wewnątrz, jak i pomiędzy serwisami [Miller 2000]. Oprócz ważnego zagadnienia kompatybilności sprzętu i oprogramowania komputerowego zwraca on także uwagę na współdziałanie w sensie sposobu funkcjonowania serwisów GBC, głównie z punktu widzenia przepływów informacyjnych. Przedstawia on współdziałanie w sześciu aspektach:

- Współdziałanie techniczne, związane ze stałym rozwojem normalizacji komunikacji, transportu, pamięci masowych i reprezentacji danych.
- Współdziałanie semantyczne, dotyczące nazewnictwa, na przykład stosowania w różnych kontekstach (zastosowaniach) terminów typu autor, twórca czy kompozytor, które mogą oznaczać to samo lub bardzo podobne pojęcie.
- Współdziałanie polityczne/społeczne – decyzja o zwiększeniu dostępności zasobów ma implikacje dla serwisu, którego dotyczy (utrata kontroli lub poczucia własności), dla jego pracowników (umiejętność obsługi bardziej skomplikowanych systemów i nowych, rozproszonych grup użytkowników) oraz obsługiwanych użytkowników.
- Współdziałanie pomiędzy społecznościami, ich instytucjami i poszczególnymi dyscyplinami, wynikające z potrzeb docierania przez uczonych do różnorodnych zasobów, zawierających informacje zarówno z ich głównych, jak i pobocznych obszarów zainteresowań.
- Współdziałanie prawne pomiędzy unormowaniami poszczególnych krajów dotyczących uprawnień do dostępu i własności.
- Współdziałanie międzynarodowe – wszystkie wymienione problemy jeszcze narastają, gdy rozpatrywane są w skali międzynarodowej, gdzie różnice technologiczne, w stosowanych zasadach pracy i organizacji narastały przez wiele lat. Z drugiej strony rzadko można znaleźć przykład tak rozbudowanej współpracy, jaka realizowana jest między serwisami GBC na wielu płaszczyznach i od wielu lat.

Można wyróżnić wiele obszarów, w których mogą występować utrudnienia w zakresie osiągnięcia współdziałania, wpływające na rozwój GBC. Wiele z nich opisałem szczegółowo w innych rozdziałach tej książki. Współdziałaniu sprzyjają:

- Jednolite interfejsy użytkownika.
- Jednolite systemy (przestrzenie) nazw i identyfikacji.
- Standaryzacja formatów dla zasobów informacyjnych.
- Jednolite systemy semantyczne, w tym ontologie.
- Standardowe schematy metadanych.
- Standardowe protokoły sieciowe.
- Standardowe protokoły wyszukiwania informacji.
- Standardowe zasady autoryzacji i zapewnienia bezpieczeństwa [Tedd, Large 2005, s. 88].

Jak wynika z tego wyczerpanego wyliczenia, współdziałanie i standaryzacja rozwiązań każdego rodzaju idą ze sobą w parze; są dwiema stronami tego samego zagadnienia. Na podstawie

innych rozdziałów tej książki można powiedzieć, że w zakresie współdziałania (standaryzacji) dokonano już bardzo wiele. Wszystkie działania związane z GBC powinny być realizowane w taki sposób, aby zapewniać możliwe wszechstronne współdziałanie wybranych rozwiązań. Sprzyja temu tendencja, widoczna w rozwoju IIN, tworzenia raczej systemów (publikowania, oprogramowania, technologii) otwartych, ułatwiających współdziałanie.

4.7. Archiwizacja (długotrwałe przechowywanie)

Długotrwała archiwizacja, obok zapewnienia właściwej liczności udostępnianych zbiorów, współdziałania oraz identyfikowalności treści, usług i osób, uznawana jest za najważniejszą cechę właściwie zorganizowanego środowiska informacji cyfrowej [Lavoie, Dempsey 2004]. Jest ona bardzo istotnym zagadnieniem: braki w jej realizacji powodowały w początkowym okresie funkcjonowania Internetu formułowanie opinii o Sieci jako źródle immanentnego kryzysu w zakresie zachowania dziedzictwa kulturowego i naukowego ludzkości⁶¹. Powstawały określenia typu „ciemny wiek cyfryzacji”, oznaczające obawę przed całkowitą utratą piśmiennictwa z zakresu kultury i nauki [Kuny 1997]. Zwracano jednak także uwagę na pozytywne cechy Webu z punktu widzenia archiwizacji – obiektu w Webie nie da się ukraść czy zniszczyć, tak jak obiektu fizycznego [Koehler 2006]. Obecnie na szczęście już nie padają pytania o to, czy należy archiwizować zawartość Web, ale raczej w jaki sposób zarządzać zasobami Webu w sposób zapewniający ich archiwizację i stały dostęp. O wadze tego typu zadań świadczą międzynarodowe inicjatywy, takie jak International Internet Preservation Consortium (IIPC)⁶² [Hakala 2004, s. 182] oraz amerykańska National Digital Information Infrastructure and Preservation Program (NDIIPP)⁶³.

O ile prawdą jest, że obiekty cyfrowe są trudniejsze do archiwizowania niż analogowe, to należy stwierdzić, że stopień ryzyka utraty materiałów różni się znacznie w zależności od rodzaju zbiorów. Istnieje na przykład spora obawa przed utratą osobistej strony Web uczonego, nawet z dnia na dzień, jednak znacznie mniejsze niebezpieczeństwo takiego zdarzenia istnieje w przypadku zasobów czasopisma komercyjnego, udostępnianego w Webie czy repozytorium instytucji naukowej. Dość nieoczekiwanie pomocne w zakresie archiwizacji okazały się Google, które wiecześnie archiwizują strony Web, indeksowane przez wyszukiwarkę⁶⁴. Innym zasobem, ukazującym historię Webu, jest Internet Archive Brewstera Kahle, który jest organizacją non profit, utworzoną w 1996 r. w celu archiwizacji informacji cyfrowej i zachowania jej dla przyszłych pokoleń. Średnio co dwa miesiące do archiwum dodawane są strony indeksowanego Webu. Jest to najbardziej kompletny zapis Webu dostępny dla użytkowników, a jego najstarsza zawartość bywa unikalna. Internet

⁶¹ Jednym z pierwszych autorów, którzy zwracali uwagę na problemy archiwizacji był Jeff Rothenberg, który w 1995 r., w swoim artykule przedstawił hipotetyczną sytuację: zostawił swoim wnukom CD-ROM i list, w którym wyjaśnił, że na CD-ROM-ie znajduje się instrukcja odnalezienia pozostawionego wnukom skarbu. CD-ROM został jednak znaleziony w 2045 r. Czy będzie możliwe odczytanie CD-ROM, jeśli nośnik ten nie jest już używany? [Rothenberg 1995, s. 42]. Artykuł ten zapoczątkował szeroką dyskusję.

⁶² <http://netpreserve.org/about/index.php>.

⁶³ <http://www.digitalpreservation.gov/>.

⁶⁴ W związku z archiwizacją przez Google wszystkich indeksowanych przez nie treści, w czasopismach komputerowych pojawiają się nawet porady, w jaki sposób usunąć niechciane materiały z pamięci Google. Może się bowiem na przykład okazać, że ktoś chce usunąć lekkomyślnie wysłany post na listę dyskusyjną podpisany własnym nazwiskiem i zawierający niecenzuralne treści. Wówczas jedynym sposobem jest indywidualna prośba do przedstawicieli wyszukiwarki [Dren 2008, s. 22].

Archive dostępny jest poprzez Wayback Machine⁶⁵. Po podaniu przez użytkownika URL poszukiwanej strony, Wayback Machine zwraca listę stron o danym URL ze wszystkich dostępnych okresów, co pozwala użytkownikowi na wybór wersji do przeglądania [Arms i in. 2006].

Archiwizacja obiektów cyfrowych definiowana jest jako działania zarządcze, niezbędne dla: 1) długoterminowego zarządzania strumieniami bajtów (włącznie z metadanymi), wystarczającego do reprodukcji wyglądu dokumentu oryginalnego, oraz 2) ciągłego udostępniania treści dokumentu w czasie, bez względu na zachodzące zmiany technologiczne [Jantz, Giarlo 2006, s. 195]. Celem systemu archiwizacji jest więc zapewnienie użytkownikom stałej dostępności do informacji, zawartej w obiektach cyfrowych, poprzez ich właściwe przechowywanie [Rosenthal i in. 2005]. Oznacza to, że chodzi nie tylko o zapisywanie bitów i regularne tworzenie kopii, ale wymagane są dodatkowe działania zapewniające możliwość interpretacji danych⁶⁶ w długim okresie. Obejmują one szereg czynności, od zapisu na nośnikach po transformację, realizowanych w zależności od rodzaju zasobów. Nie są to działania jednorazowe, wymagają zaplanowania stałe wykonywanych czynności i istnienia organizacji, które podjęłyby się ich wykonywania. Niezbędne jest więc utworzenie systemu – serwisu GBC, który zapewniałby archiwizację w znacznie dłuższym okresie, niż czas życia poszczególnych nośników danych, elementów sprzętu i oprogramowania oraz formatów, w jakich informacja jest zapisywana.

Archiwizacja jest kosztownym i długotrwałym przedsięwzięciem. Pliki muszą być właściwie rozmieszczane, sprzęt konserwowany, oprogramowanie aktualizowane, dane należy kopiować i przenosić do różnych formatów, a wszystko to muszą robić właściwie opłacani ludzie, nawet, gdy żadne nowe dane nie trafiają już do bazy danych [Borgman 2007, s. 95]. Największą szansę na właściwą archiwizację mają materiały, za które odpowiedzialność wzięła jakaś instytucja.

Przedsięwzięcia takie są zadaniem bibliotek od początku ich istnienia; obecnie zasadniczą zmianę stanowią jedynie warunki technologiczne archiwizacji obiektów cyfrowych. Stąd częste poglądy, że właśnie na bibliotekach spocznie obowiązek zapewnienia długotrwałej dostępności obiektów cyfrowych. Istnieją obawy przed przekazaniem tych zadań firmom komercyjnym, takim jak wydawcy czy wyszukiwarki internetowe⁶⁷. Wydawcy mogą bowiem zakończyć działalność bez pozostawienia następcy, odpowiedzialnego za obsługę archiwum. Zasoby archiwalne, nie przynoszące dochodów, mogą być pozostawiane bez opieki lub nawet likwidowane. Natomiast biblioteki od stuleci odgrywają podstawową, na ogół obiektywną i apolityczną, rolę w zapewnieniu stałego dostępu do zasobów dokumentów naukowych [Nentwich 2003; Waters 2006]. Obecnie częściej wskazuje się na repozytoria, jako miejsca służące długotrwałej archiwizacji. Warto jednak zaznaczyć, że archiwizacja informacji często stoi w sprzeczności z udostępnianiem; paradoksalnie najlepsze metody służące długotrwałej archiwizacji utrudniają udostępnianie (i na odwrot). W związku z tym nie wszystkie serwisy GBC muszą i mogą archiwizować udostępniane zasoby. W takich przypadkach podejmowana jest współpraca, na przykład z biblioteką narodową [Hakala 2004, s. 178], centrum obliczeniowym czy dostawcą komercyjnym, którzy zapewniają usługi archiwizacyjne na odpowiednim poziomie⁶⁸ orga-

⁶⁵ <http://www.archive.org/web/web.php>.

⁶⁶ Archiwizacja nie dotyczy tylko zachowania ciągu bitów, ale także sposobu interpretacji tych bitów. Bez interpretacji ciąg bitów jest wyłącznie nic nie znaczącym zestawem zer i jedynek.

⁶⁷ Wspomniane już projekty masowej digitalizacji, takie jak Google Books i inne, uruchomione na ich wzór, mają na celu nie tylko umożliwienie wyszukiwania dzięki stworzeniu wersji cyfrowych, ale także archiwizację tych wersji wraz z innymi zindeksowanymi obiektami cyfrowymi.

⁶⁸ „Poziom archiwizacji” oznacza zakres odpowiedzialności, jaką instytucja podejmuje w stosunku do użytkowników, za przechowywanie obiektów cyfrowych. Dzięki temu instytucja może oczekiwać na moment, w którym oferowane będą technologie, umożliwiające zastosowanie wyższego poziomu archiwizacji. Nie istnieje ustalona lista poziomów archiwizacji, ale można wskazać kilka, ogólnie akceptowanych.

nizacyjnym i technicznym. Może zdarzać się, że instytucja zarządzająca serwisem GBC rozpoczyna tworzenie kolekcji zasobów cyfrowych, archiwizuje ją na minimalnym poziomie, a po określonym czasie przekazuje te zasoby organizacji lepiej przygotowanej do długotrwałej archiwizacji.

Zasady długotrwałej archiwizacji wciąż nie zostały do końca dopracowane. Nie istnieją jednoznaczne przepisy ani podręczniki, zawierające sprawdzone rozwiązania na każdą sytuację. Należy przyjąć, że nie istnieją nośniki, sprzęt ani oprogramowanie, co do których można mieć pewność niezawodności w długim okresie. W związku z tym długotrwała archiwizacja musi uwzględniać następujące założenia:

- System archiwizacji jako całość musi być odporny na błędy poszczególnych jego elementów, w taki sposób, aby był w stanie funkcjonować nawet w przypadku wystąpienia kilku błędów jednocześnie.
- Pomimo błędu lub dezaktualizacji nośnika, sprzętu lub oprogramowania, dane muszą być nadal dostępne w systemie. Nieodpowiedni element powinien być w takim przypadku wymieniany. System musi zapewniać różnorodność elementów, aby zapobiec problemom, wynikającym ze stosowania tylko wybranych rozwiązań. Pozwala to uniknąć wymiany wszystkich elementów jednocześnie, a także wiązania się z ofertą jednego dostawcy.
- Większość danych, przechowywanych w archiwum, jest używana rzadko lub sporadycznie. System, który rozpoznaje błędy tylko podczas próby dostępu, będzie miał tendencję do narastającego gromadzenia „uspionych” błędów. W związku z tym trzeba zapewnić regularne kontrole, w odstępach czasu wystarczających do utrzymania prawdopodobieństwa istnienia błędów na akceptowalnym poziomie.

Zapobieganie efektom błędów ułatwiają strategię archiwizacji. Ich wybór zależy między innymi od rodzaju archiwizowanych materiałów. Obecnie biblioteki naukowe zazwyczaj zakupują licencję na dostęp do treści czasopism elektronicznych, które pozostają w bazach danych wydawców, co uniemożliwia bibliotekom realizację archiwizacji. Wynika z tego, że zagadnienia stałego dostępu i długotrwałej archiwizacji nie są tożsame. Potrzeba zapewnienia stałego dostępu jest związana z sytuacją, powstałą w wyniku licencjonowanego dostępu do czasopism elektronicznych; walka o stały dostęp ma na celu przywrócenie sytuacji z czasów czasopism drukowanych, gdy biblioteki posiadały zakupione przez siebie materiały bez względu na to, czy kontynuowały prenumeratę. Długotrwała archiwizacja natomiast polega raczej na zapewnieniu, że do określonych materiałów, przyszłe generacje uczonych i ludzkości w ogóle, będą miały dostęp bez względu na to, czy jakaś biblioteka je wcześniej zakupiła.

Archiwizacja najczęściej realizowana jest przez takie strategie, jak:

- Druk i muzeum. Ze względu na nietrwałość nośnika elektronicznego czasem proponuje się archiwizację treści, szczególnie w formie tekstowej, w postaci drukowanej lub na mikroformach. Urządzenia (komputery, czytniki) są jednocześnie przechowywane w muzeach, gdzie są one konserwowane i utrzymywane w ruchu. Strategia ta na ogół nie jest traktowana poważnie, chociaż zdarzają się przypadki jej stosowania. Może ona raczej służyć w sytuacjach awaryjnych.
- Usługi biblioteki narodowej, tworzącej odpowiedni serwis. Przykładem może być e-Depot w Koninklijke Bibliotheek, współpracujący z takimi wydawcami, jak Elsevier, Springer, Blackwell, OUP i Sage.
- Replikacja, w której wykorzystuje się łatwość kopiowania obiektów cyfrowych bez

Podstawowy poziom archiwizacji polega na zachowaniu „ciągu bitów”; zachowywane są surowe dane w dokładnie takim stanie, w jakim zostały dostarczone. Pomimo tego, że wymaga ona dobrze zorganizowanego środowiska technologii informacyjnej, przyszli użytkownicy nie mają gwarancji że obiekt będzie przedstawiony w taki sam sposób jak oryginał, gdyż brak jest informacji o interpretacji ciągu bitów. Innymi poziomami archiwizacji są na przykład: „archiwizacja dostępu” lub „archiwizacja reprezentacji”.

utruty informacji, co umożliwia tworzenie wielu kopii dokumentu, ułatwiając jego zachowanie. Jeżeli weźmiemy pod uwagę błędy wymienione wcześniej, pojedynczy egzemplarz ma niewielkie szanse na długotrwałe istnienie, więc replikacja jest ważnym warunkiem archiwizacji, ale nie jedynym i wystarczającym. Najbardziej znanym projektem wykorzystującym tę strategię jest LOCKSS⁶⁹, w którym każda biblioteka uczestnicząca gromadzi własną kopię interesującego ją obiektu. Ta metoda jest więc najbardziej zbliżona do sposobów archiwizacji dokumentów tradycyjnych.

- Emulacja. Polega na utrzymywaniu przechowywanych dokumentów w pierwotnej formie oraz emulacji (naśladowaniu) sposobu działania oryginalnego systemu informatycznego (sprzętu i oprogramowania) w przyszłym, nieznanym środowisku. Dzięki temu oryginalne oprogramowanie, przy pomocy którego archiwizowany dokument powstał, może być uruchomione na nowym sprzęcie i pod nowym oprogramowaniem systemowym.
- Niezbędna jest znajomość stosowanych formatów i oprogramowania, przez co do celów archiwizacji najlepiej nadają się formaty otwarte, których specyfikacja jest powszechnie dostępna. Ta strategia wykorzystana jest w wielu ważnych projektach archiwów cyfrowych, takich jak CEDARS⁷⁰ i CAMILEON⁷¹.
- Migracja. Polega na konwersji dokumentu ze starego formatu do innego, używanego przez nowe oprogramowanie. W tym przypadku nie wystarczy samo skopiowanie fizycznego ciągu bitów na nowy nośnik, ale należy także przetłumaczyć go na nowy format logiczny. Trudność tych działań zależy od cech modyfikowanego oprogramowania, ale zazwyczaj nie sprawia kłopotów. Funkcje te bywają zaszyte w nowe wersje oprogramowania; wówczas możliwe jest wczytanie pliku starej wersji i zapisania go w formacie nowej wersji oprogramowania.
- Filmowanie. W przypadku plików o skomplikowanym formacie, dla których istotny jest sposób interakcji użytkownika, można zapisywać wykorzystanie pliku i jego wygląd przez filmowanie użytkownika korzystającego z pliku [Clausen 2004, s. 19]. Samo „filmowanie” może mieć postać kolejnych zrzutów ekranów lub fizycznego filmowania za pomocą kamery, czasem wraz z komentarzami. Ta metoda może być szczególnie przydatna w przypadku systemów składających się z wielu, wzajemnie zależnych elementów, które nie mogą być łatwo konwertowane czy emulowane lub też, gdy sposób wykorzystania jest ważniejszy od samego programu (jak np. w przypadku chatów). Jest to metoda kosztowna i pracochłonna.

Długotrwała archiwizacja wiąże się z zaangażowaniem sił i środków w przedsięwzięcie, realizowane w niezdefiniowanym okresie. Organizacje, które się jej podejmują, powinny mieć świadomość zadań i odpowiedzialności, związanych z tą działalnością. Pomimo euforii i pośpiechu, które często towarzyszą tworzeniu serwisów globalnej biblioteki cyfrowej, należy zastanowić się, kto będzie zajmował się zgromadzonym zasobem, jeśli pierwotnie zaangażowana instytucja temu nie podola. Dlatego też większe szanse mają na długotrwały sukces instytucje, osadzone w solidnym otoczeniu instytucjonalnym.

Model systemu długotrwałej archiwizacji OAIS (Open Archival Information Systems) utworzony został przez NASA, w 2003 r. zatwierdzony jako norma ISO⁷². Stwarza on podstawy dla zapewnienia efektywności i trwałości tych systemów przez określenie modeli: funkcjonalnego oraz informacji⁷³. Według OAIS archiwum jest organizacją ludzi i systemów, która przyjęła odpowiedzialność za archiwizację informacji i jej udostępnia-

⁶⁹ LOCKSS – Lots Of Copie Keep Staff Safe (<http://www.lockss.org/lockss/Home>).

⁷⁰ <http://www.leeds.ac.uk/cedars/>.

⁷¹ <http://www.si.umich.edu/CAMILEON/>.

⁷² ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model.

⁷³ Termin OAIS stosowany jest w kilku znaczeniach: jako nazwa publikacji modelu informacyjnego OAIS, jako sam model informacyjny oraz jako repozytorium spełniające wymagania modelu.

nie określonej społeczności. Model opisuje systemy definiując pakiety informacji, które zawierają treść informacji (wraz z metadanymi) – SIP, opis archiwizacji (AIP) i zasady udostępniania informacji (DIP) [Brindley, Muir, Proberts 2004, s. 242].

Model może być podstawą dla analiz funkcjonowania systemu archiwizacyjnego. Serwis zgodny z modelem OAIS, musi posiadać właściwą strukturę organizacyjną oraz wykazać, że realizuje sześć obowiązkowych funkcji związanych z gromadzeniem, archiwizacją i udostępnianiem obiektów cyfrowych:

1. Przyjmowanie odpowiedniej informacji⁷⁴ od jej producenta (z selekcją).
2. Uzyskiwanie wystarczającej kontroli nad informacją do realizacji długoterminowych celów archiwizacji.
3. Określenie społeczności użytkowników.
4. Zapewnienie dostępności do informacji bez potrzeby pośrednictwa producenta informacji.
5. Stworzenie udokumentowanej polityki i procedur dla zapewnienia ochrony na wypadek przewidywalnych błędów oraz zapewnienia udostępniania autentycznej kopii archiwizowanej informacji.
6. Udostępnianie archiwizowanej informacji użytkownikom.

Zastosowanie archiwizacji w skali globalnej biblioteki cyfrowej oznacza konieczność zdecentralizowanego podejmowania decyzji, a więc współpraca wielu serwisów tej biblioteki musi odbywać się na zasadzie porozumienia, a nie dyktatu. Ponieważ nie ma możliwości dokładnego przewidywania przyszłości ani panowania nad kierunkami rozwoju, ambicja działania na skalę globalną, pomimo braku możliwości kontroli wszelkich działań prowadzonych na świecie, wymusza elastyczność przyjmowanych rozwiązań. Wymaga to takiej realizacji funkcjonujących zastosowań, aby w przyszłości możliwe było łatwe korygowanie podejmowanych działań.

4.8. Ocena (kontrola jakości)

W globalnej bibliotece cyfrowej ocenie podlegać mogą i powinny wszystkie jej elementy. W szczególności przydatna wydaje się kontrola jakości dwóch elementów cyklu komunikacji naukowej: ocena jakości publikacji naukowych, a przez to również ich autorów – uczonych i instytucji ich zatrudniających oraz ocena jakości funkcjonowania systemów informacyjno-wyszukiwawczych, stanowiących serwisy GBC. Obie oceny mogą się łączyć, na przykład w sytuacji, gdy dokonuje się oceny wyszukanych publikacji zarówno z punktu widzenia ich jakości, jak i stopnia relewancji wyników wyszukiwania. Ocena jakości publikacji, w tym stron Web, oznacza określenie poziomu zaufania i wiarygodności; jest obecnie często wykonywana automatycznie. Ocena serwisów GBC dotyczy głównie efektywności wyszukiwania (relewancji, pertynencji). Oba te aspekty oceny jakościowej przedstawione zostaną w dalszej części rozdziału. Miary służące kwantyfikacji tych ocen zaprezentowałem już w p. 3.3.

Publikowanie elektroniczne wpływa na modyfikację, wcześniej stosowanych i dobrze ukształtowanych, metod i procedur **oceny jakości publikacji** w systemie komunikacji naukowej. Podstawą tych procedur było recenzowanie publikacji, przyjmujące różne formy.

⁷⁴ W odróżnieniu „odpowiedniej” informacji od pozostałej przydatne mogą być odpowiedzi na pytania przedstawione przez Rossa Harveya: (1) dla kogo materiał będzie przechowywany? (2) dlaczego materiał jest wart archiwizacji? (3) czy istotny jest wygląd materiału (czy wraz ze zmianą wyglądu materiał utraci przydatność)? (4) czy istotny jest sposób funkcjonowania materiału (czy utraci on wartość, jeżeli odnośniki w jego wnętrzu przestaną działać)? (5) czy możliwe jest wyróżnienie poszczególnych elementów lub obszarów (np. czy banery reklamowe stanowią integralną całość z pozostałymi elementami)? (6) jeżeli wyróżnienie elementów, które trzeba obsługiwać sprawia problemy, może łatwiej będzie określić efekt braku obsługi elementu i odszukać funkcje lub elementy, które z pewnością nie są potrzebne [Harvey 2007, s. 17].

W tym zakresie cyfryzacja spowodowała z jednej strony powstanie nowych form oceny jakości, wcześniej nie występujących lub rzadko stosowanych (takich jak komentarze po opublikowaniu lub punktowanie), a z drugiej strony udostępnianie wielkiej liczby materiałów, których jakość w żaden sposób nie była oceniana. W środowisku sieciowym zanika lub ulega zmianie wiele tradycyjnych sposobów zapewnienia jakości publikacji, stąd użytkownicy mogą mieć do czynienia, jeśli nie z obniżoną, to przynajmniej zróżnicowaną jakością udostępnionych materiałów. Powoduje to konieczność dokonywania starannej selekcji obiektów cyfrowych, szczególnie, jeśli wiąże się ona z długotrwałą archiwizacją w serwisach globalnej biblioteki cyfrowej.

W procesach komunikacji naukowej kontrola jakości publikacji obejmuje dwie podstawowe kategorie, często mające części wspólne. Są to: jakość treści naukowych oraz jakość samych uczonych [Borgman 2007, s. 58]. Treści naukowe, rozpowszechniane w formie publikacji, badane są pod względem poprawności oraz tego, czy są one wystarczająco ważne (nowe, oryginalne), aby warte były opublikowania. Podstawowym mechanizmem w tym przypadku jest recenzowanie. Pomiędzy poszczególnymi dyscyplinami naukowymi istnieją poważne różnice w zakresie rozumienia właściwego poziomu poprawności i ważności tekstu, jak również sposobów realizacji recenzowania.

Pracownicy nauki (czy raczej ich osiągnięcia) oceniane są na potrzeby decyzji o zatrudnieniu, promocji oraz konkursów, na przykład służących uzyskiwaniu funduszy na badania. Ocena w tym zakresie może być holistyczna, uwzględniająca jakość poszczególnych publikacji wraz z innymi osiągnięciami, takimi jak uzyskane granty, promocje na stopnie naukowe, osiągnięcia dydaktyczne, prace edytorskie i administracyjne. Duże znaczenie ma tzw. reputacja uczonego i instytucji, jaką reprezentuje. Reputacja ma istotne znaczenie także w działalności wydawniczej (reputacja czasopisma, redakcji, wydawcy).

Uczni i czasopisma, które uważane są za wiarygodne i posiadają wysoką reputację, mogą być traktowane jako tzw. autorytety kognitywne (*cognitive authority*). Wydawnictwa i czasopisma, uważane przez określoną grupę uczonych za tego typu autorytety, mogą być identyfikowane przez badanie poglądów osób tworzących sieci praktyk. Helena Francke wskazuje na autorytet kognitywny jako główne kryterium wyboru źródeł informacji przez osoby pozyskujące wiedzę poza własnym, bezpośrednim doświadczeniem. Autorytetami kognitywnymi są osoby, których twierdzeniom ufamy w konkretnej sytuacji lub sprawie; mogą to być także teksty⁷⁵ lub instrumenty realizujące podobne funkcje. Czynniki, wpływające na bycie autorytetem kognitywnym w znacznym stopniu zależą od rodzaju odbiorców (sieci praktyk). Na decyzję o uznaniu dokumentu za autorytet kognitywny wpływa, między innymi, jego nośnik, co jest powodem stosunkowo powolnego odchodzenia od publikacji drukowanych do elektronicznych, w szczególności Open Access [Francke 2008, s. 120].

Jako wyznaczniki poziomu jakości działalności naukowej można stosować także bardziej sformalizowane metody, w tym wskaźniki publikowania. Sprowadza to jednak niebezpieczeństwo łatwego wypaczenia idei kontroli jakości treści naukowych [Monastersky 2006; Revheim 2002]. Można także mieć wątpliwości co do tego, komu w rzeczywistości w tradycyjnym systemie publikowania naukowego była potrzebna kontrola jakości – uczonemu, których chronić miała przed otrzymywaniem publikacji niskiej jakości czy po prostu błędnych, czy może raczej wydawcom, którzy musieli stworzyć system eliminacji publikacji ze względu na niewydolny system publikowania, gdzie chętnych do ogłaszania swych dzieł jest znacznie więcej, niż pozwalają na to techniczne i ekonomiczne możliwości wydawców. Kontrola jakości ma dla uczonych różne znaczenie, w zależności od

⁷⁵ Helena Francke proponuje trzy sposoby określenia poziomu autorytetu tekstu: określenie osób związanych z jego powstaniem (autorów, redaktorów, recenzentów); badanie historii tekstu; przydatność dla odbiorcy: przekazywanie ważnych treści odpowiednio sformatowanych i sformułowanych [Francke 2008, s. 316]. Sposoby te nie dają jednak jednoznacznych wyników i mogą być traktowane pomocniczo.

tego, do czego służy ma oceniany akt komunikacji – na przykład w trakcie komunikacji nieformalnej pomiędzy specjalistami podobnego poziomu i specjalizacji, kontrola jakości komunikatów w ogóle nie jest potrzebna.

Ocena jakości publikacji może być wykonywana w dwojaki sposób: nowy i tradycyjny. Nowe sposoby to takie, które nie mogłyby istnieć bez funkcjonowania środowiska elektronicznego. Można tu także zaliczyć metody oceny jakości, które mogły co prawda istnieć wcześniej, jednak uległy poważnym modyfikacjom. Przykładem jest recenzowanie, znane od czasu pojawienia się pierwszych czasopism naukowych, ale w środowisku cyfrowym bardzo zmodyfikowane. Metody tradycyjne obejmują sposoby zapewnienia jakości istniejące przed Internetem, które nie uległy poważnym zmianom pod wpływem zastosowania nowych technologii.

Dla publikacji drukowanych kontrolę jakości przeprowadza się w różny sposób:

- Brak formalnej oceny, sytuacja stosunkowo rzadka, gdyż materiały przekazywane przez autora są przynajmniej przeglądane przez redaktora czasopisma. Pomijanie kontroli zdarza się dla materiałów konferencyjnych, gdzie referaty są kwalifikowane na podstawie abstraktów oraz dla szarej literatury.
- Przegląd materiałów wpływających do redakcji wykonywany przez redaktora, najczęściej uczonego doświadczonego i dysponującego odpowiednim prestiżem. Ma on też wpływ na formę publikacji. Czasem wpływ redaktora jest tak duży, że staje się on nieformalnym współautorem publikacji.
- Przegląd materiałów przez zespół redaktorów. Praca może wówczas przebiegać dwuetapowo, najpierw odbywa się ocena przez redaktora odpowiedzialnego, a następnie przez członka redakcji.
- Recenzowanie stanowi rozwinięcie poprzednich sposobów. Istnieje kilka wariantów tego systemu. Różnice dotyczą liczby recenzentów, tego czy nazwiska recenzentów i autorów są im odpowiednio nawzajem znane⁷⁶, stosowania powtórnych recenzji po poprawkach autora [Nentwich 2003, s. 370].

Przedstawione sposoby oceny jakości publikacji pozostają w „długim cieniu Oldenburga”, jak to określił Jean-Claude Guédon [Guédon 2001]. Oznacza to, że pomimo zmian, powodowanych cyfryzacją środowiska nauki, wciąż realizowany jest model Henry’ego Oldenburga, twórcy pierwszego czasopisma naukowego, oparty na ścisłym powiązaniu oceny jakości i rozpowszechniania wiedzy, którego efektem są tzw. czasopisma recenzowane. Nowe środowisko pozwala jednak, a może nawet wymusza, rozdzielenie certyfikacji od rozpowszechniania [Armbruster 2008b]. Stąd, oprócz stosowania wcześniejszych metod, powstają nowe formy kontroli jakości, typowe dla publikacji elektronicznych. Można je podzielić na dwie grupy – ocena wykonywana przed opublikowaniem i później⁷⁷:

1. Ocena przed formalnym opublikowaniem:

- Recenzowanie otwarte (komentarze): teksty są umieszczane na stronie Web czasopisma i każdy zainteresowany może przekazywać komentarze na ich temat. Redaktor nie wybiera recenzentów, a raczej otwiera dyskusję. Dane zarówno autora, jak i komentujących nie są ujawniane, chociaż powinni oni być zarejestrowani (komentarze nie powinny być całkowicie anonimowe). Środowisko cyfrowe daje tu nowe możliwości, takie jak na

⁷⁶ W tym zakresie wyróżnić można recenzje „ślepe” w jedną stronę – gdy recenzenci znają nazwisko autora, oraz w obie strony, gdy nazwiska autorów ani recenzentów nie są im nawzajem znane. To drugie rozwiązanie jest stosowane częściej ze względu na korzyści: pozwala recenzentowi na swobodną krytykę bez obawę o reakcję autora. Dzięki temu, że redaktor zna obu, działa on jak filtr, wylapując ewentualne opinie niesprawiedliwe. Brak jakiegokolwiek anonimowości może wpływać na sposób oceniania prac.

⁷⁷ Pomimo tego, że obie formy oceny znane są także w publikowaniu tradycyjnym, funkcjonowanie tekstu publikacji w środowisku cyfrowym znacznie ułatwia ocenę po opublikowaniu, z czego wynika różnorodność form takiej oceny. Poza tym środowisko cyfrowe umożliwia ocenę poprzez bardziej bezpośrednią wymianę opinii, co upodabnia ją do pierwotnych form oceny jakości pracy naukowej, dokonywanej podczas dyskusji po odczytaniu tekstu przez autora w trakcie zebrania towarzystwa naukowego.

przykład ocena oceniających; nie tylko artykuły, ale także komentarze i komentujący mogą podlegać ocenie i uzyskiwać reputację [Whitworth, Friedman 2009].

- Systemy mieszane, łączące rozwiązania recenzowania tradycyjnego i otwartego komentowania. Recenzje udostępniane są do otwartego komentowania; powoduje to uczestnictwo różnych osób na kolejnych etapach oceny. Na podstawie recenzji, opinii i modyfikacji autora, przygotowywana jest ostateczna wersja publikacji.

2. Ocena po opublikowaniu:

- Komentarze i adnotacje, dodawane przez użytkowników po opublikowaniu materiałów. Można je zapisywać wraz z ocenianym tekstem, jako uzupełnienie wszystkich wcześniej omówionych form oceny. W tym kierunku idą prace realizowane przez W3C, takie jak projekt Annotea⁷⁸, które umożliwić mają tworzenie adnotacji do stron Web. Założeniem jest, że interesujące prace uzyskają dużo dobrych komentarzy, a nieciekawe – negatywne lub żadne. Odmianą tej metody jest m.in. Faculty of 1000⁷⁹, organizowany przez BioMed Central, gdzie najwyżsi rangą uczeni opiniują świeżo opublikowane artykuły [Armbruster 2007, s. 265]. Członkostwo w Faculty of 1000 daje wysoki prestiż, co zachęca do częstego komentowania.
- Wykorzystanie wspomnianych już zakładek (*bookmark*) – zob. p. 3.3. Zakładki te, oprócz kierowania do zaznaczonej strony Web (na której umieszczono na przykład artykuł naukowy), zawierają także opisujące ją słowa kluczowe. Badanie tych zakładek może pozwolić na określenie popularności strony oraz jej aktualności (krótkotrwałej ważności) [Taraborelli 2008].
- Punktowanie przez czytelników. Przyznawane punkty służą do obliczania średnich, a wyniki ujawniane są użytkownikom. Przydatne są narzędzia Webu 2.0, takie jak Digg. Stosowanie tej metody może prowadzić do fałszowania wyników.
- Ocena stopnia wykorzystania publikacji, znacznie łatwiejsza w środowisku elektronicznym. Używane są metody webometryczne. Najczęściej wykorzystuje się dwie miary: liczbaostępów (otwarcia pliku, żądania wydruku) i analiza cytowań (traktowanych tradycyjnie, jako umieszczenie opisu w bibliografii albo jako liczba hiperlinków podłączonych do obiektu lub zakładki – *bookmark*ów go dotyczących). Pierwsza metoda znajduje zastosowanie w repozytoriach, takich, jak ArXiv, SSRN⁸⁰ i RePEc, druga w indeksach cytowań typu CiteSeer i w wyszukiwarkach. Obie mają swoje zalety i wady⁸¹. Zasadniczym elementem tej oceny jest założenie o pełnej dostępności publikacji naukowych poprzez serwisy globalnej biblioteki cyfrowej (głównie repozytoria). Wówczas lista możliwych badań i wskaźników webometrycznych jest imponująca [Shadbolt i in. 2006, s. 195-202].

Z bardzo pobieżnego przedstawienia nowych sposobów i metod oceny jakości publikacji elektronicznych, dostępnych w globalnej bibliotece cyfrowej, wynika kilka ogólnych wniosków. Po pierwsze, jakość publikacji naukowych nie zależy od medium, w jakim są publikowane. Bez względu na to, czy publikacja jest drukowana, czy elektroniczna, jej jakość może być wysoka lub niska, a standardy jakościowe recenzowanych czasopism, publikowanych w obu formach, są jednakowo rygorystyczne. System kontroli wpływa na podniesienie jakości przez odrzucenie najbardziej ewidentnych pomyłek, jednak w systemie tradycyjnym ma on nie tylko wspierać pracowników nauki w ich wyborach (bo to właśnie pracownicy nauki są w stanie odróżnić publikacje wysokiej jakości od innych,

⁷⁸ <http://www.w3.org/2001/Annotea/>.

⁷⁹ <http://www.f1000biology.com/>.

⁸⁰ Social Science Research Network (<http://www.ssrn.com/>).

⁸¹ Badania webometryczne w odróżnieniu od indeksów cytowań Thomson Scientific uwzględniają odnośniki do bardzo zróżnicowanej grupy publikacji, w dodatku bardzo zmiennej w czasie. Rzadko uwzględniają one publikacje funkcjonujące w zamkniętych, komercyjnych bazach danych. Powszechnie znane są także nieustające próby oszukiwania algorytmów wyszukiwarek, na przykład przez tworzenie sztucznych odnośników.

więc często nie potrzebują takiego doradztwa), ale także służyć wydawcom do wstępnej selekcji dostarczanych materiałów, niezbędnej ze względu na ograniczoną ilość miejsca. Z tego względu wydawcy muszą negatywnie selekcjonować nie tylko materiały zawierające błędy, ale także takie, które zawierają informacje interesujące dla zbyt małej grupy odbiorców (znajdujące się w tzw. długim ogonie) [Weinberger 2007, s. 103], co biorąc pod uwagę wysoką specjalizację nauki jest zjawiskiem szkodliwym. Również to, czy czasopiśmo jest komercyjne czy Open Access, nie musi wpływać na jego jakość, o czym świadczą przykład czasopiśm OA z grupy Public Library of Science (PLOS), które osiągnęły bardzo wysoki poziom naukowy.

Po drugie, nowe formy oceny jakości zmieniają ten system z preselekcji materiałów do publikacji w kierunku uwidocznienia i sformalizowania opinii środowiska (pozytywnych i negatywnych) dotyczących materiałów już opublikowanych (dostępnych w Sieci). Każdy więc może swobodnie publikować w Sieci, jednak za swoje publikacje bierze znacznie większą odpowiedzialność, gdyż nie ma nikogo, kto zazwyczaj wytknie błędy, a publikacja zostanie oceniona przez najlepszych recenzentów – innych autorów w danej dyscyplinie i w razie błędów w najlepszym przypadku pominięta milczeniem. Propozycje tego typu oceny jakości pochodzą już z pierwszej połowy lat dziewięćdziesiątych XX wieku, gdy Stevan Harnad przedstawił swoją „wywrotową propozycję” (*subversive proposal*) polegającą na promocji modelu zdecentralizowanego, gdzie uczone sam publikuje swoje prace, które stają się przedmiotem dyskusji środowiska (*scholarly skywriting*) [Nahotko 2007b, s. 22].

Po trzecie można odnieść wrażenie, że różnica pomiędzy tradycyjnymi i nowymi sposobami oceny polega nie tyle na dużych zmianach w metodach oceny, ale w poziomie jawności tego procesu. Nowe serwisy GBC, szczególnie związane z Nauką 2.0, powodują, że procesy tworzenia i oceny treści stają się jawne, bardziej widoczne „na zewnątrz” w porównaniu z tradycyjnym recenzowaniem, zlecanym przez redaktora anonimowemu specjalistce. Publikując w czasopiśmie drukowanym autor dochodzi do porozumienia z recenzentem w sposób niewidoczny dla odbiorców; produkt końcowy jest skończony i nie podlega dalszym modyfikacjom⁸². Podczas publikowania elektronicznego ostateczna wersja tekstu może powstawać podczas interakcji autora z czytelnikami, którzy jednocześnie są oceniającymi i sami podlegają ocenie. W obu przypadkach chodzi jednak o to samo – o uzgodnienie konsensusu w sprawie przedstawionych tez pomiędzy autorem a środowiskiem specjalistów w danej dziedzinie.

Ze względu na wzrost poziomu interaktywności komunikacji naukowej, również w procesach kontroli jakości jednym z najważniejszych czynników staje się czas, niezbędny do realizacji takiej rozbudowanej komunikacji. Tradycyjne recenzowanie znacznie wydłuża czas udostępnienia tekstu użytkownikom. W środowisku elektronicznym materiały dostarczane są natychmiast, a ich ocena nie ma wpływu na zwłokę w dostępności. Trudno obecnie przewidzieć, czy publikowanie naukowe nadal będzie sterowane wymogiem posiadania jak największej liczby publikacji, czy raczej istotniejsza będzie ich jakość. W pierwszym przypadku pozostanie bardzo niewiele, jeśli w ogóle, czasu na dodatkową komunikację. W efekcie, w zakresie oceny jakości, nadal ważniejsza będzie selekcja ilościowa, a główną korzyścią przyspieszenie wszystkich procesów. W drugim przypadku więcej czasu można będzie poświęcić na komunikację, w tym dyskusje na forach i komentowanie. Wówczas lepiej wykorzystywany jest potencjał, znajdujący się w nowym medium, a nowe technologie przyspieszą zmiany w działalności naukowej, w tym w procesie publikowania.

Kierunek i tempo zmian w dużym stopniu zależy od dyscypliny nauki i jej tradycji w zakresie publikowania. Rozwój nowych form oceny jakości nie jest jednoczesny ani

⁸² Pomijam fakt, że autor w kolejnych publikacjach może doskonalić przedstawione wcześniej tezy, może także się z nich wycofać.

jednokierunkowy we wszystkich dyscyplinach. W jednych powstają wielkie, pojedyncze repozytoria wszystkich publikacji, w innych tworzone są rozproszone repozytoria instytucjonalne, w jeszcze innych publikowanie tradycyjne wciąż jest wyżej cenione. We wszystkich tych przypadkach muszą być wypracowane metody oceny jakości, nie utrudniające transformacji ku globalnej bibliotece cyfrowej.

Od początku lat osiemdziesiątych XX wieku, wraz z pojawieniem się technologii osobistych, których symbolem jest komputer osobisty (PC), rozpoczęto zwracać większą uwagę na dostosowywanie rozwijającej się technologii do możliwości i potrzeb przeciętnego człowieka. Technologie informacyjne zaczęły być tak projektowane, aby mógł z nich bezpośrednio korzystać tzw. użytkownik końcowy, bez potrzeby pośrednictwa specjalistów – informatyków czy bibliotekarzy. Zapoczątkowało to badania w zakresie interakcji człowiek-komputer (*human-computer interaction* – HCI), które doprowadziły do sformułowania podstawowych zasad projektowania technologii informacyjnych i kryteriów oceny ich funkcjonowania.

Ocena serwisów GBC ma spełniać określone wymogi. Musi ona odzwierciedlać wybory i decyzje podejmowane w zakresie:

1. Przedmiotu oceny: tego, co jest oceniane, co to jest (globalna) biblioteka cyfrowa, z jakich składa się części, składników, procesów. Ze względu na podobieństwo podstawowych funkcji (wyszukiwanie informacji), GBC w pewnym zakresie może być modelowana i oceniana na zasadach podobnych, jak biblioteki tradycyjne. Z drugiej strony pewne funkcje w bibliotece cyfrowej wykonywane są całkowicie odmiennie, więc również ich ocena musi odbywać się na innych zasadach. Ze względu na to, że globalna biblioteka cyfrowa nie jest instytucją, nie można stosować do niej kryteriów oceny dla instytucji, takich jak na przykład poziom przygotowania zawodowego pracowników. Tego typu kryteria mogą być wykorzystywane do oceny biblioteki cyfrowej na poziomie poszczególnych serwisów, ich usług i aplikacji.

2. Zakresu oceny: wybór celów, metod, punktu widzenia, poziomów oceny, określenie celów dla wybranych poziomów. Globalna biblioteka cyfrowa, podobnie jak każdy system, może być opisywana i oceniana z różnych punktów widzenia i poziomów. Dużą trudność podczas oceny sprawia zazwyczaj wybór odpowiedniego poziomu celów. Tefko Saracevic podzielił cele, a co za tym idzie również oceny biblioteki cyfrowej, na kilka ogólnych klas (poziomów) [Saracevic 2000]. Trzy pierwsze ukierunkowane są bardziej na użytkownika, trzy kolejne uwzględniają cechy systemów (szczególnie istotne ze względu na zapewnienie współdziałania), natomiast poziom interfejsu umieszczony został pomiędzy tymi poziomami. Poziomy te, z uwzględnieniem potrzeb oceny globalnej biblioteki cyfrowej, można opisać jako:

– Poziom społeczności. Możemy mówić tu o globalnej społeczności pracowników nauki i ich potrzebach, możliwa jest także ocena przydatności globalnej biblioteki cyfrowej i jej poszczególnych serwisów dla wybranych grup użytkowników. Badany jest poziom realizacji potrzeb i żądań, ról i zasad praktycznego postępowania społeczności.

– Poziom instytucjonalny. Jak już wspomniałem, poziom ten nie występuje dla globalnej biblioteki cyfrowej, która nie jest instytucją. Składa się ona jednak z wielu serwisów, w tym także instytucjonalnych (często właśnie dlatego szczególnie cennych), które można badać indywidualnie lub grupowo (na przykład grupę repozytoriów uczelnianych) pod względem poziomu realizacji misji i celów instytucji lub organizacji, poziomu integracji zasobów i usług różnych instytucji.

– Poziom indywidualny. Oceny realizacji potrzeb, zadań, działań informacyjnych na poziomie globalnej biblioteki cyfrowej lub jej usług dla indywidualnego użytkownika. Na tym poziomie realizowana jest większość badań, być może ze względu na łatwość uzyskania wyników.

– Poziom interfejsu. Podstawą interfejsu globalnej biblioteki cyfrowej jest ekran wy-

świetlany przez przeglądarki internetowe i wyszukiwarkę Google, ale prawdopodobnie nigdy (albo przynajmniej do czasu pełnego ujednolicenia serwisów globalnej biblioteki cyfrowej) nie dojdzie do scalenia różnych interfejsów serwisów GBC w jeden, uniwersalny. Poszczególne interfejsy mogą być więc oceniane pod względem poziomu wspomaganie dostępu, wyszukiwania, nawigacji, przeglądania i interakcji z zasobami globalnej biblioteki cyfrowej.

– Poziom rozwiązań technicznych. Badana jest poprawność funkcjonowania sprzętu, oprogramowania i ich konfiguracji. Na tym poziomie oraz na obu kolejnych szczególną rolę odgrywa zapewnienie współdziałania wymienionych elementów.

– Poziom przetwarzania. Dotyczy poziomu poprawności realizacji procedur, technik, algorytmów, operacji. Szczególną rolę odgrywają protokoły stosowane na różnych poziomach struktury globalnej biblioteki cyfrowej, począwszy od zapewniających poprawne funkcjonowanie całej sieci, po protokoły stosowane w lokalnych aplikacjach.

– Poziom treści. Dotyczy dostosowania zasobów lub źródeł informacji, zasad ich selekcji, reprezentacji, organizacji, tworzenia struktur i zarządzania do potrzeb użytkowników. Treści mogą być badane na różnych poziomach – Internetu jako całości (znane narzekania na „wielki śmietnik”), wielkich serwisów informacyjnych GBC, aż po prywatne zasoby danych i informacji udostępniane wybranym użytkownikom. Ich zasoby mogą być traktowane w kategoriach ekonomicznych; wówczas ich wartość, rozumianą na przykład jako wartość posiadanych praw własności intelektualnych, można wyrazić liczbowo [Zhang 2007, s. 268].

3. Metodologii oceny: metodologia pomiarów, wybór próby, procedury wyboru zbioru i sposobów analizy danych. Wymienia się takie metody oceny, jak analiza logów transakcji (technika specyficzna dla systemów zautomatyzowanych) oraz techniki od dawna stosowane w badaniach dotyczących funkcjonowania bibliotek tradycyjnych, takich jak ankieta, wywiad i obserwacja [Bertot i in. 2006, s. 25].

Walka o podnoszenie efektywności systemów informacyjno-wyszukiwawczych rozpoczęła się na długo przed rozpowszechnieniem Internetu i jego usług. Wraz z pojawieniem się wyszukiwarek internetowych sytuacja uległa wyraźnej zmianie; ich twórcy nie uzależniali sukcesu swojego produktu od poziomu tradycyjnie rozumianej efektywności wyszukiwania, czyli wzrostu jego trafności i/lub kompletności. Badania uwzględniające te wskaźniki zawodzą, prowadząc czasem do konstatacji, że 85% wyszukujących w Web jest zadowolonych z wyników bez względu na to, co otrzymają, więc system nie funkcjonuje zbyt dobrze, ale użytkownicy są zadowoleni, bo o tym nie wiedzą [Lesk 2005, s. 226]. Oczywiście nie jest to prawdą; potrzeby użytkowników GBC oceniane i zaspakajane są w inny niż wcześniej, co nie znaczy – gorszy, sposób.

Barczo często twórcy serwisów GBC nie dysponują wiedzą na temat rzeczywistych potrzeb użytkowników. Konstruują systemy według własnych wyobrażeń potrzeb wyszukiwawczych. Takie twierdzenie było między innymi podstawą tworzenia uproszczonych (w stosunku do MARC) schematów metadanych, gdyż jak wynika z badań nad serwisami informacyjnymi, oferującymi ustrukturyzowane metadane, użytkownicy zainteresowani są wyszukiwaniami według najprostszych kryteriów: autora i tytułu, najczęściej pomijając wszelkie inne kryteria formalne i rzeczowe⁸³.

Uzyskanie pełnej wiedzy o poziomie realizacji potrzeb użytkowników przez serwisy GBC może wymagać spełnienia postulatów Williama Armsa, który proponuje całkowitą zmianę sposobu oceny tych serwisów [Arms 2005]. Nawet ci, którzy uznają użytkownika za główny przedmiot badań, standardowo oceniają serwisy GBC badając sposób reali-

⁸³ Przykładowo por. dane dotyczące wyszukiwania w katalogu KaRo przedstawione przez Tomasza Wolniewicza: 98,5% wszystkich wyszukiwań prowadzonych jest według jednego lub dwóch kryteriów, którymi są najczęściej tytuł (62%), autor (18%), ISBN (16%), dla dwóch kryteriów: autor+tytuł (95% wyszukiwań według połączenia tych dwóch kryteriów) [Wolniewicz 2008].

zacji zadań przez użytkowników w obrębie ograniczonej usługi serwisu. Jest to ocena z organizacyjnego lub technicznego punktu widzenia. Tymczasem użytkownik nie widzi globalnej biblioteki cyfrowej jako pojedynczej instytucji czy współdziałających zasobów metadanych, lecz jako globalny zbiór przydatnych mu dokumentów, w związku z czym bardziej użyteczne byłoby podejście holistyczne. Badania holistyczne mogłyby dotyczyć użytkownika, na przykład nauczyciela akademickiego przygotowującego zajęcia, a polegać na obserwacji wszystkich narzędzi, które ta osoba wykorzystuje (a nie tylko tych, które oferuje jakaś biblioteka tradycyjna lub serwis GBC) i sprawdzaniu ich efektywności we wzajemnych kombinacjach.

Zwolennikiem holistycznego podejścia do oceny bibliotek cyfrowych jest Scott Nicholson, który przedstawił koncepcję takiej oceny z kilku punktów widzenia: użytkownika, bibliotekarza, decydenta [Nicholson 2004]. Ocenę prowadzić można za pomocą matrycy tematów i punktów widzenia na ocenę (elementy składowe systemu, efektywność systemu, użyteczność systemu, sposób posługiwania się systemem, interakcje z użytkownikiem), dzięki której wybierane są: przedmiot i sposób oceny. Zastosowanie takiego modelu pozwala na uzyskanie interesujących danych z każdej perspektywy i o wszystkich zagadnieniach. Nie daje on możliwości zastosowania jednej miary oceny, gdyż tylko jednoczesne użycie wielu miar może pozwolić na holistyczne zrozumienie całego systemu.

4.9. Semantyczny Web oraz Web 2.0 i Nauka 2.0

Duże znaczenie dla przyszłego rozwoju GBC, jako internetowej infrastruktury wiedzy, ma idea Semantycznego Webu. Jest on nową generacją Webu, którego twórcy starają się używać takiej reprezentacji informacji, aby była ona zrozumiała nie tylko dla ludzi, ale także dla maszyn i nie tylko do celów poprawnego jej wyświetlenia, ale również dla automatycznego stosowania, integracji i wielokrotnego wykorzystania przez wiele aplikacji. Jest to jeden z najbardziej nośnych tematów badawczych ostatnich lat, z pogranicza sztucznej inteligencji i sieci rozległych – i jedna z głównych sfer działalności World Wide Web Consortium (W3C)⁸⁴.

U podstaw Semantycznego Webu leży zmiana sposobu myślenia o danych [Daconta, Obst, Smith 2003, s. 2]. Początkowo były one zamknięte w odrębnych aplikacjach; uważano je za wtórne w stosunku do przetwarzania danych. Tego typu niepoprawne podejście doprowadziło do zjawisk, które były przyczyną powstania wyrażenia GIGO⁸⁵, wskazującego na zależność pomiędzy jakością danych i rezultatami ich przetwarzania. Oznacza ono, że dobre oprogramowanie jest całkowicie zależne od jakości danych. Informatycy zauważyli to, co dla bibliotekarzy zawsze było oczywiste: że dane są najważniejsze, muszą więc być odpowiednio weryfikowane i chronione. Semantyczny Web ma zapewnić obsługę danych w Sieci na poziomie odpowiadającym zakresowi zarządzania danymi w bibliotekach [Greenberg 2007, s. 207]. W tym celu powstały obiektowe języki programowania, stawiające dane na pierwszym miejscu. Wciąż jednak dane zamknięte były w poszczególnych aplikacjach, więc ich dostawcy mogli zamykać dostęp do danych, na przykład dla zwiększenia swojej konkurencyjności. Wraz z rozwojem Internetu i Webu, takich narzędzi jak XML oraz Semantycznego Webu, dane uwalniane są z poszczególnych aplikacji. Ukierunkowanie ku danym przetwarzanym automatycznie wymusza tworzenie bardziej inteligentnych systemów, w tym serwisów GBC.

⁸⁴ <http://www.w3.org/>.

⁸⁵ GIGO – skrót od „garbage in, garbage out”, czyli śmieci na wejściu, śmieci na wyjściu (chodzi o wejście i wyjście systemu przetwarzania danych).

Semantyczny Web ma być środowiskiem, w którym dane będą bardziej zrozumiałe dla komputerów dzięki strukturalnemu opisowi ich zawartości. Jego elementem jest również infrastruktura tworzona dla „inteligentnych” agentów (oprogramowanie), przemierzających Web w celu wykonywania złożonych działań na danych dla swoich użytkowników, co pociąga za sobą potrzebę pełnej integracji agentów z Webem i optymalne wykorzystanie jego infrastruktury (wysyłanie komunikatów, bezpieczeństwo, uwierzytelnienie, usługi katalogowe i zasady usług aplikacyjnych). W Semantycznym Webie następuje jawna deklaracja wiedzy, zawartej w aplikacjach Webowych, inteligentna integracja informacji, semantyczny dostęp do Internetu i automatyczne pobieranie informacji z tekstu. Usługi Web współdziałają ze sobą na wielką skalę, dzięki czemu powstaje Web usług interpretowalnych przez komputery, wyszukiwanych, wykonywanych i komponowanych przez inteligentnych agentów.

Można wyróżnić cztery etapy wzrostu „inteligencji” danych, od najprostszych do danych zawierających ilość informacji wystarczającą do prowadzenia maszynowych wnioskowań na ich temat:

- Teksty i bazy danych przed XML. Jest to etap początkowy, gdzie większość danych podporządkowuje się aplikacji. Inteligencja dotyczy więc raczej aplikacji, niż danych.
- Dokumenty XML dla określonych zastosowań (dziedzin). Na tym etapie dane uniezależniają się od aplikacji w obrębie określonej dziedziny. Mogą przemieszczać się między aplikacjami, ich semantyka jest automatycznie interpretowana przez narzędzia indeksujące.
- Przestrzenie nazw, takie jak taksonomie i inne słowniki kontrolowane, przypominające tradycyjne JIW, stosowane w SIW. Na tym etapie dane mogą pochodzić z różnych dziedzin i zastosowań. Są one klasyfikowane za pomocą hierarchicznych taksonomii, które mogą być także wykorzystywane do wyszukiwania danych. Taksonomia opisuje relacje pomiędzy poszczególnymi kategoriami danych.
- Metadane, pozwalające na identyfikację oraz wykorzystanie relacji pomiędzy obiektami. W coraz większym stopniu tworzone są automatycznie (zob. p. 4.2).
- Ontologie i reguły (zob. p. 2.7). Ostatni etap pozwala na wnioskowanie nowych danych z danych istniejących na podstawie reguł wnioskowania, stanowiących łącznie model dziedziny. Dane są opisywane przez relacje i skomplikowane wyrażenia formalne, na których wykonywane są operacje logiczne. Na tym etapie nie istnieją już one jako pojedyncze, odrębne jednostki, lecz jako skomplikowany mikrosystem.

Semantyczny Web związany jest z danymi „inteligentnymi”; jest to maszynowo przetwarzalna sieć inteligentnych danych, niezależnych od aplikacji, łączliwych, sklasyfikowanych, stanowiących część większego systemu (ontologii) [Daconta, Obst, Smith 2003, s. 4]. Dla umożliwienia realizacji idei Semantycznego Webu stworzona została bogata infrastruktura, którą można podzielić na cztery kategorie:

- Języki dla Semantycznego Webu: głównie XML i jego aplikacje: RDF, RDFS, OWL⁸⁶.
- Ontologie, umożliwiające dostęp do sieci wiedzy czytelnej dla maszyn i przez nie przetwarzalnej. Wiedza kodowana jest w formatach opartych na XML.
- Kodowanie semantyczne stron Semantycznego Webu: dla umożliwienia rozumienia treści udostępnianych w Web muszą one być kodowane, czyli opisane za pomocą terminologii zdefiniowanej w ontologii lub różnych ontologiach.
- Usługi Semantycznego Webu: są to aplikacje programowe pozwalające użytkownikom i agentom na bezpośrednie ich wywoływanie. Mogą być wyszukiwane, opisywane i udostępniane w intranetach, extranetach i w Internecie, w oparciu o XML i standardowe protokoły Webu.

Architektura Semantycznego Webu przedstawiona została w mojej poprzedniej publi-

⁸⁶ OWL – Web Ontology Language; podstawą tej technologii jest wyróżnianie klas i ich własności wraz z tworzeniem złożonych relacji pomiędzy nimi.

kacji [Nahotko 2006b, s. 30-36]. Jej realizacja może mieć zasadnicze znaczenie dla funkcjonowania GBC, gdyż umożliwia semantyczne współdziałanie pomiędzy jej serwisami, dzięki czemu powstać powinno zintegrowane środowisko informacyjne o globalnej skali. Semantyczny Web znacznie ułatwia funkcjonowanie GBC, stwarzając zupełnie nowe możliwości wykorzystania gromadzonych danych i informacji naukowych.

Wiele aplikacji Semantycznego Webu związanych jest ze stosowaniem założeń **Webu 2.0 i sieci społecznościowych** [Burke 2009, s. 317]. Pojęcie Web 2.0 pojawiło się według encyklopedii Wikipedia⁸⁷ w 2001 r. i oznacza serwisy internetowe, w funkcjonowaniu których główną rolę odgrywają treści tworzone przez samych użytkowników, skierowane do wszystkich zainteresowanych, wspólnie decydujących o ich ważności i sposobie prezentacji.

Twórcy Web 2.0 opisują jego ideę wskazując na zestaw cech, z których wiele związanych jest z koncepcją globalnej biblioteki cyfrowej. W Wikipedii wyodrębniono trzy grupy cech:

- Techniczne: wykorzystanie mechanizmu wiki i blogów; udostępnianie interfejsów XML, pozwalających innym stronom i programom na korzystanie z danych Web 2.0 (przede wszystkim przez RSS i Atom⁸⁸); używanie nowych technologii, jak na przykład AJAX⁸⁹.
- Społeczne: generowanie treści przez użytkowników; użycie folksonomii⁹⁰; tworzenie rozbudowanych społeczności wokół serwisów; wykorzystanie inteligencji zbiorowej; wykorzystanie otwartych licencji, jak Creative Commons czy GNU GFDL.
- Specyficzny wygląd: pastelowe barwy, gradienty, zaokrąglenia, duże czcionki.

Zdaniem jego propagatorów, globalna sieć przechodzi gruntowną metamorfozę. Serwisy budowane według „filozofii Web 2.0” w znacznie większym stopniu niż wcześniejsze otwierają użytkownikom nowe perspektywy dla własnej aktywności. Dla budowy takich serwisów wykorzystuje się nowe technologie – jak wiki czy weblog (blog) – udostępniające mechanizmy pozwalające internautom na łatwe tworzenie i dodawanie treści w Internecie oraz na pracę zespołową nad zawartością serwisu⁹¹.

Specyfikę Webu 2.0 często wyjaśnia się na przykładach nowych serwisów, typowych dla sieci społecznych. Web 2.0 to jednak nie tylko różnice w rodzajach powstających serwisów. Zmieniają się też obyczaje internautów, którzy coraz chętniej próbują wychodzić poza role biernych konsumentów newsów i multimedialnych. Szukają alternatywnych źródeł informacji lub sami starają się je dostarczać. Publikują, komentują i wchodzą w interakcje z innymi użytkownikami sieci⁹². Podejmując decyzje, skłonni są raczej oprzeć

⁸⁷ http://pl.wikipedia.org/wiki/Web_2.0.

⁸⁸ Protokół Atom jest standardem poziomu aplikacji, służącym do publikowania i edycji zasobów Web (typu blogów, Wiki, czasopism elektronicznych), wykorzystującym HTTP (operacje GET, POST, PUT i DELETE) i XML. Protokół ten stał się m.in. podstawą, na której utworzono protokół SWORD, służący przekazywaniu publikacji do repozytoriów cyfrowych [Allinson, François, Lewis 2008].

⁸⁹ AJAX (Asynchronous JavaScript and XML) – grupa technologii służących tworzeniu interaktywnych aplikacji Web.

⁹⁰ Folksonomia oznacza praktykę opisu treści (najczęściej internetowych) z wykorzystaniem dowolnie dobranych słów kluczowych czy innych wyrażen (etykiet – *tag*), dobieranych przez użytkowników opisywanych treści (nic-ekspertów). Opisy tworzone są wspólnie, przez społeczność użytkowników Webu.

⁹¹ Właśnie współdziałanie, szybkie sprzężenie zwrotne i zespołowość prowadzonych prac stanowią powód, dla którego Web 2.0 nazywany jest siecią społeczną.

⁹² W Webie 2.0, podobnie zresztą jak w każdym rodzaju publikowania, wyróżnia się uczestników głównych, którzy dodają nowe treści, modyfikują je i komentują, uczestników drugorzędnych, którzy wyłącznie komentują i korzystają z informacji oraz użytkowników pasywnych, wyłącznie korzystających z informacji. Co interesujące tylko kilka procent użytkowników Webu 2.0 należy do pierwszej grupy tworzącej własne publikacje (YouTube – 0,16%, Flickr – 0,2%, Wikipedia – 4,6%), z czego wynika że sieć społeczna nie wyzwala twórczości naprawdę masowej. Sytuacja wygląda lepiej, gdy uwzględnia się, że aktywność użytkowników może przyjmować

się na zbiorowej opinii innych internautów, wśród których nie brak również zawodowych ekspertów.

Zasady działania Webu 2.0 wpływają na sposób funkcjonowania nauki; dla tych zastosowań stosowany jest nawet termin Nauka 2.0. W porównaniu z tradycyjnymi narzędziami komunikacji naukowej umożliwiają znaczne rozszerzenie zasięgu i zakresu dyskursu. Artykuł w czasopiśmie przedstawia pozytywne wyniki badań i eksperymentów, ale zazwyczaj nie można tam znaleźć informacji o tym, co się nie udało i jaki miało to wpływ na realizację eksperymentu. Nauka 2.0 umożliwia swobodną dyskusję w trybie Open Access o dowolnych aspektach przeprowadzonych prac (także o popełnionych błędach), co znacznie przyspiesza ich realizację. Uczni przedstawiają często kolegom wyniki badań na bieżąco, każdego dnia, od pierwszych notatek w laboratorium do artykułu w czasopiśmie, co znamionuje tzw. otwartą naukę – *Open Science*. Mówi się nawet o tzw. *Open Notebook Science*, od notatek tworzonych w laboratorium w komputerowym notatniku powszechnie dostępnym, zarówno dla ludzi, jak również systemów komputerowych. Pomimo konieczności upublicznienia wyników jeszcze przed ich formalnym opublikowaniem, korzyści odnoszą wszystkie strony takiego dialogu, który zmienia współzawodnictwo pomiędzy uczonymi we współpracę [Waldrop 2008]. Dzięki temu powstająca Nauka 2.0 jest nie tylko bardziej kolektywna od tradycyjnej, ale także znacznie bardziej efektywna, wykorzystuje bowiem zdolność wirtualnych społeczności uczonych do wykorzystywania uzdolnień wszystkich ich członków [Jenkins 2007, s. 31].

Wdrażanie Nauki 2.0 w procesach badawczych następuje w zakresie:

- Szeroko rozumianej komunikacji naukowej dla upowszechniania wyników badań, rozwoju współpracy, likwidacji barier między dyscyplinami.
- Zmiany rozumienia „naukowości”, spowodowanej modyfikacją parametrów oceny (wspólne filtrowanie⁹³, tagowanie, *reblogging*⁹⁴), statusu publikacji, metodologii dziedzin.
- Powstawania wirtualnych wspólnot uczonych i w ich obrębie nowych autorytetów.
- Efektów organizacyjnych powodujących intermedialność, interdyscyplinarność, interaktywność, digitalizację i cyfryzację danych i zasobów informacji.
- Konkretnego stosowania narzędzi – blogosfera, panele dyskusyjne, serwisy typu Wiki, MySpace, Facebook.
- W ostateczności zmiany paradygmatu nauki i kryterium naukowości; metodyka badań naukowych polegać będzie na łączeniu badań nauk ścisłych z metodami nauk społecznych [Solska 2009, s. 136].

Co warte podkreślenia, Nauka 2.0 funkcjonuje nie dzięki wielkim inwestycjom w infrastrukturę nauki (jak w IIN), lecz na zasadzie inicjatyw oddolnych. Są nawet tacy, jak Alexander Bard i Jan Söderqvist, którzy twierdzą, że jest ona reakcją na przerosł administracji i snobizm intelektualny tradycyjnej nauki [Bard, Söderqvist 2006, s. 234]. Uczni z własnej inicjatywy stosują darmowe narzędzia typowe dla Webu 2.0 do komunikowania się w społeczności pracowników nauki zajmujących się podobnymi problemami, co zresztą powoduje szybkie poszerzenie kontaktów i rozwój współpracy. Dzieje się tak, ponieważ zasady Nauki 2.0 wydają się być naturalnym sposobem postępowania uczonych. Od cza-

różne formy; Jose Bernoff i Charlene Li dzielą użytkowników technologii społecznościowych na: twórców treści, krytyków – autorów komentarzy i uczestników forum, kolekcjonerów organizujących treści, członków sieci społecznych, widzów konsumujących cudze treści oraz innych, nie tworzących ani konsumujących treści. Liczbę tych ostatnich oblicza się na 44% [Li, Bernoff 2008, s. 43].

⁹³ Wspólne filtrowanie (*collaborative filtering*) jest procesem wydobycia informacji lub typowych wzorców badanych zjawisk realizowanym we współpracy wielu agentów w wielkich zasobach danych. Wynikiem mogą być na przykład listy bestsellerów lub rekomendacje.

⁹⁴ Reblogging polega na przekazywaniu informacji odnalezionych w Internecie, w szczególności na blogach, na kolejne blogi. Działanie takie umożliwiając niemal wszystkie programy do prowadzenia blogów. Liczba prześłań świadczy o popularności blogu.

sów rewolucji naukowej rozwijają oni swoją wiedzę przez uczestnictwo wielu badaczy w powszechnej debacie, umożliwiającej wybór teorii najlepiej opisujących rzeczywistość. Można więc powiedzieć, że Nauka 2.0 świetnie odpowiada sposobowi, w jakim funkcjonuje nauka, pomijając może szczególnie przypadki badań objętych tajemnicą lub prowadzących do zastrzeżeń patentowych, chociaż w tym drugim przypadku problem wydaje się natury organizacyjnej⁹⁵.

Warto wspomnieć o zastosowaniu narzędzi Web 2.0 w działalności bibliotek tradycyjnych i cyfrowych w zakresie interakcji między bibliotekarzami i użytkownikami, czyli o tak zwanym zjawisku Biblioteka 2.0 (*Library 2.0*) [Holmberg i in. 2009, s. 677; Gmirek 2007]. Wykorzystanie takich serwisów i usług, jak wspólne redagowanie stron Web (w tym naukowych), społecznościowe OPAC (SOPAC), blogi, podcasty, RSS, komunikatory, aplikacje typu Wiki i Flickr stanowią dalsze (poza tworzeniem bibliotek cyfrowych i repozytoriów) rozszerzenie działalności wydawniczej bibliotek w środowisku cyfrowym, a szerzej działalności w zakresie komunikacji naukowej, zarówno formalnej, jak i nieformalnej. Ułatwiają także wzmacnianie relacji biblioteki w środowisku. Stosowanie tego typu środków wzbogaca i zwiększa dostępność cyfrowych i tradycyjnych zasobów bibliotek naukowych, przyczyniając się do lepszego funkcjonowania GBC.

Nowe technologie informatyczne i telekomunikacyjne, zamiast oczekiwanego rozwoju „sztucznej inteligencji”, rozumianej jako technologia zastępująca ludzkie myślenie, raczej spowodowały olbrzymie zwiększenie możliwości wszelkiego rodzaju komunikowania między ludźmi – w tym także komunikacji naukowej [Bard, Söderqvist 2006, s. 34]. Przed powstaniem serwisów Webu 2.0 stosunki panujące w Internecie stanowiły przeniesienie zasad, funkcjonujących w świecie publikacji tradycyjnych (drukowanych). Role w łańcuchu informacji były znane: z jednej strony niewielka grupa nadawców informacji i jeszcze mniejsza grupa organizatorów procesu komunikacji, a z drugiej wielka zbiorowość odbiorców. Technologia Web 2.0 łamie te schematy; tutaj każdy odbiorca (konsument) jest jednocześnie nadawcą (twórcą i producentem), może bezpośrednio i aktywnie uczestniczyć w komunikacji. Wykorzystanie Webu 2.0 w publikacji czasopism naukowych nowego rodzaju zostało przedstawione w mojej wcześniejszej pracy [Nahotko 2007b, s. 159-160]. Model taki świetnie nadaje się do realizacji komunikacji naukowej, gdzie pracownicy nauki zawsze odgrywają na przemian wszystkie role w łańcuchu informacji: jednocześnie są autorami, redaktorami, recenzentami i czytelnikami. W komunikacji naukowej bardzo istotne znaczenie ma współpraca i szybkie sprzężenie zwrotne, uzyskiwane między innymi dzięki wykorzystaniu środków komunikacji nieformalnej, więc Web 2.0, łączący oba rodzaje komunikacji (formalną i nieformalną), może być szczególnie przydatny w działalności naukowej. Przy tym idea Web 2.0 opiera się w znacznym stopniu na pomysłach i mechanizmach, wykorzystujących zbiorową mądrość internautów („mądrość tłumów”), a więc jej zasady są łatwe do stosowania w komunikacji naukowej, gdyż rozwój nauki opiera się na pomysłach i funkcjach, wykorzystujących zbiorową mądrość uczonych. Bardzo interesujące efekty przynosi połączenie zbiorowej inteligencji z elementami Semantycznego Webu, co według Toma Grubera powoduje pojawienie się systemów zbiorowej wiedzy [Gruber 2008].

Według Mary Burke, Semantyczny Web i Web 2.0 są twórcami komplementarnymi; ten pierwszy pozwala na tworzenie inteligentnych odnośników i wykorzystanie zasobów, które są następnie udostępniane powszechnie poprzez interfejsy Webu 2.0. Użytkownik, podczas przydzielania tagów do obiektów Webu 2.0, może być przełączany do formalnych słowników w Semantycznym Webie. Najbardziej rozwojowym obszarem przyszłych badań i działań jest integracja Semantycznego Webu i Webu 2.0 w semantycznie powiązane ze sobą serwisy społecznościowe [Burke 2009, s. 318]. Połączenie to często nazywane

⁹⁵ Chodzi o to, aby urzędy patentowe uznawały pierwszeństwo do wynalazku na podstawie daty nadesłania informacji do serwisu społecznościowego, np. Wiki.

jest Webem 3.0 [Das i in. 2009]. Jest on traktowany jako praktyczne wdrożenie idei Semantycznego Webu, choć z niezbędnymi modyfikacjami pierwotnych zamierzeń.

Poglądy na kierunki rozwoju współczesnego Webu są bardzo zróżnicowane, od rozwiązań dotyczących nowych zastosowań sztucznej inteligencji, po przyspieszające działanie łącz internetowych czy też umożliwiające rozwój grafiki komputerowej. Są tacy autorzy, jak David Shotton, którzy twierdzą, iż użytkownicy Webu już zaczynają odczuwać korzyści z implementacji rozwiązań, przygotowywanych z myślą o Semantycznym Webie [Shotton 2009, s. 86]. Korzyści te wynikają m.in. z wykorzystania metadanych czytelnych dla komputera, dotyczących publikacji w Webie, tworzonych w oparciu o standardy Semantycznego Webu, pozwalających na komputerowe gromadzenie i integrację informacji. Procesy te realizowane są głównie na poziomie serwisów GBC.

5. SPOŁECZNOŚCI

Spoleczności globalnej biblioteki cyfrowej tworzą i wykorzystują serwisy tej biblioteki; te ostatnie powstają dla zaspokajania potrzeb grup użytkowników. W tym rozdziale podejmię próbę odpowiedzi na pytanie o powody, dla których zasoby GBC wykorzystywane są przez poszczególne społeczności oraz o relacje, funkcjonujące w ich obrębie oraz pomiędzy nimi. Liczne związki zachodzą także pomiędzy społecznościami i ich członkami a współczesnymi narzędziami stosowanymi w GBC.

Użytkownicy łatwo przyzwyczajają się do usług online. Bibliotekarze w okresie uruchamiania katalogów online (OPAC) szybko zorientowali się, że czytelnicy polegają na zasobach informacyjnych online nawet, gdy nie są one kompletne. W trakcie konwersji katalogów kartkowych do elektronicznych starsze materiały, dostępne wyłącznie poprzez katalogi kartkowe (często prowadzone równoległe z OPAC, a więc bardziej kompletne), szybko przestały być zamawiane, co skłaniało bibliotekarzy do przyspieszania konwersji. Podobnie jest z czasopismami elektronicznymi; wraz ze wzrostem oferty biblioteki w tym zakresie, rosną także oczekiwania użytkowników, którzy nie chcą już przesiadywać w czytelniach. W obliczu globalnej biblioteki cyfrowej, tradycyjna biblioteka naukowa w mniejszym stopniu staje się właścicielem i depozytariuszem zasobu, a bardziej organizatorem dostępu: płaci za usługi online w imieniu swoich użytkowników, którzy nie chcą już do niej przychodzić. Stąd wynika m.in. coraz częstsze przechodzenie z udostępniania zasobów online poprzez IP (a więc dostępnych z komputerów biblioteki lub kampusu) na udostępnianie poprzez hasło (z jakiegokolwiek komputera dostępnego w sieci).

Realizowane badania wskazują na następujące oczekiwania użytkowników w stosunku do serwisów GBC. Powinny one być:

- Uniwersalne – zawierać wszystko.
- Dostępne – wszystko udostępniać każdemu.
- Natychmiast zaspokajające – szybkość odpowiedzi.
- Bezgraniczne – łatwość przekraczania granic serwisów.
- Łatwe w użyciu – pojedynczy interfejs.
- Multimedialne – tekst, obraz, dźwięk [Bawden, Vilar 2006, s. 347].

Bibliotekarze, zamiast szukać możliwości zaspokajania takich potrzeb użytkowników poprzez wyjście poza własny serwis, nawet, jeśli tworzony na poziomie narodowym, zwykle narzekają na to, że żądania są nierealne. Gary Marchionini sugeruje, że bibliotekarze powinni możliwie szybko uświadomić użytkownikom:

- Których obiektów brak w ich serwisie.
- Że niewiele dokumentów prymarnych jest dostępnych online.
- Że istnieje wiele, niezbędnych do przejścia, poziomów wyszukiwania.
- Że istnieje wiele punktów dostępu, ale nie są one ujednolicone [Marchionini, Plaisant, Komlodi 2003, s. 154].

Nic dziwnego, że w takich warunkach użytkownicy odwracają się od bibliotek w kierunku wyszukiwarek, bez względu na ich efektywność wyszukiwawczą.

W poprzednim rozdziale mowa była o procesach związanych z obsługą danych cyfrowych (*digital curation*). W tym rozdziale zajmę się instytucjami i osobami związanymi z realizacją tych procesów, odbywającą się według Chrisa Rusbridge'a na wielu różnych

poziomach, od indywidualnych osób (w tym pracowników nauki), poprzez grupy osób, instytucje i ich związki (formalne i nieformalne), również związane z komunikacją naukową (np. wydawcy) i serwisy narodowe (biblioteki, archiwa narodowe i ich sieci) [Rusbridge 2006]. Osoby i instytucje najczęściej funkcjonują w obrębie poszczególnych serwisów GBC, produkując i konsumując usługi informacyjne. Niektórych z wymienionych uczestników procesów obsługi danych przedstawię szczegółowiej w dalszej części rozdziału.

5.1. Osoby i instytucje

Osoby skupione wokół globalnej biblioteki cyfrowej tworzą społeczności, współpracujące w trakcie realizacji procesów komunikacji naukowej. Współpracując realizują oni wspólne cele, chociaż zdarzają się także przypadki występowania sprzeczności interesów, szczególnie, gdy wcześniej ustalone reguły komunikacji naukowej ulegają zmianom. Wówczas niezbędne jest osiągnięcie nowego konsensusu, pozwalającego na dalsze uzyskiwanie korzyści przez wszystkie zainteresowane społeczności. Poszczególne osoby działają zazwyczaj w sposób dzinstytucjonalizowany: pracownicy nauki zatrudniani są przez instytucje naukowe (uczelnie, instytuty, stowarzyszenia), bez wsparcia których prowadzenie badań naukowych w większości dziedzin jest od dawna niemożliwe. Na te sformalizowane struktury nakładają się jednak inne sposoby organizacji współpracy, typu kolaboratoriów i typowe dla Nauki 2.0. Być może wkrótce do zespołów badawczych dołączą bibliotekarze, na razie jednak są oni niemal w całości zatrudniani przez biblioteki naukowe, stanowiące część instytucji nauki. Najbardziej mobilną i dynamiczną częścią systemu komunikacji naukowej są wydawcy i zatrudniające ich wydawnictwa. Na rynku wydawniczym wciąż pojawiają się i znikają nowi wydawcy; wielkie wydawnictwa masowo wykupują mniejsze, a wraz z nimi wiele innych instytucji z tego samego segmentu rynku (na przykład firmy fonograficzne, producentów gier komputerowych).

Ze względu na wielki wpływ technologii informacyjnych i informatyki na funkcjonowanie GBC, nie do przecenienia jest rola specjalistów w tym zakresie, czyli informatyków. Funkcjonują oni na wszystkich poziomach i procesach GBC i jej serwisów. Dla innych osób (bibliotekarze, uczeni, wydawcy), jako minimum kompetencji technologicznych traktowana jest umiejętność posługiwania się komputerem i korzystania z usług sieciowych, bez czego nie ma praktycznej możliwości uczestnictwa we współczesnym systemie komunikacji naukowej.

Naukowe zasoby Internetu, jako jednolity zasób globalnej biblioteki cyfrowej, najlepiej i najpełniej jawią się jej użytkownikom – **pracownikom nauki**. Jak twierdzi William Arms, dzisiejsze zasoby obiektów cyfrowych, dostępne poprzez Internet mogą być traktowane jako podstawa globalnej biblioteki naukowej. Już obecnie wielu uczonych nigdy nie trafia do biblioteki (tradycyjnej), ani nie bierze do ręki drukowanego egzemplarza czasopisma naukowego. Raporty, materiały robocze, materiały konferencyjne, zbiory danych, materiały multimedialne oraz artykuły tworzone są w formie cyfrowej i udostępniane w sieci. Również starsze materiały są coraz częściej digitalizowane. W efekcie uczeni posługują się wyłącznie zasobami w formie cyfrowej [Arms 2008].

Pracownicy nauki, będący twórcami danych i informacji, są zainteresowani rozwojem swoich dyscyplin, swoich karier i budowaniem reputacji własnej i zatrudniających ich instytucji. Pragną uznania i chcą jak najszerszego rozpowszechnienia wyników własnych badań, z czym może wiązać się gratyfikacja finansowa [Willinsky 2006, s. 49]. Umieszczenie (opublikowanie) pracy w środowisku globalnej biblioteki cyfrowej zwiększa dostępność tego materiału i zainteresowanie nim (a w efekcie IF uczonego), utrudnia kradzież pomysłów i plagiaty, gdyż środowisko cyfrowe (szczególnie, gdy istnieje pełny dostęp do wszystkich dokumentów) umożliwia z dużą dokładnością odnajdywanie identycznych lub

bardzo podobnych prac. Łatwiejsze jest także określenie pierwszeństwa badań, ze względu na szybkość publikowania i dzięki opatrywaniu datą każdego pliku.

Miejsce i rola pracowników naukowych w globalnej bibliotece cyfrowej jest wieloraka¹. Wielość ról wynika z tego, że tworzenie informacji naukowej jest procesem odbywającym się w obrębie zamkniętej społeczności, w stopniu znacznie większym niż innego rodzaju twórczość lub wytwórczość. Każdy twórca uwzględnia potrzeby i gusta odbiorcy, jednak uczony jest znacznie bardziej związany z odbiorcami właśnie przez wielość i przemienność pełnionych ról – staje się tzw. prosumentem (zob. rodz. 6). Role twórcy i odbiorcy są ze sobą tak silnie związane, że stanowią różne aspekty tej samej działalności naukowej. Osoby realizujące proces badawczy występują w obu rolach: odbiorcy i twórcy jednocześnie, niesekwencyjnie (czyta się m.in. po to, aby coś nowego napisać, pisze się dzięki temu, że coś się przeczytało), gdyż obie role w komunikacji naukowej zazwyczaj są nierozłączne, inaczej niż w innych rodzajach komunikacji. To, która z ról jest wykonywana, zależy od tego, w jakiej fazie procesu badawczego znajduje się pracownik nauki. Jak twierdzi George Landow, środowisko cyfrowe jest szczególnie podatne do takiej konwergencji ról. Tutaj zanika ciąg procesów czytania – pisania – kolejnego czytania. Hipertekst tworzy czytelnika aktywnego, wręcz atakującego, co powoduje, że przejmując on część uprawnień autora. Nawet, jeżeli czytelnik nie ingeruje bezpośrednio w tekst autora, nie dokonuje w nim zmian, to poprzez dodawanie komentarzy i tworzenie połączeń pomiędzy tekstami zmniejsza dystans fenomenologiczny istniejący wcześniej pomiędzy tekstami drukowanymi. Ograniczenie autonomii tekstu powoduje ograniczenie autonomii autora [Landow 2006, s. 125]. O ile w literaturze pięknej taki wzrost roli czytelnika może wydawać się przesadny, w nauce ułatwia prowadzenie żywego dyskursu, co jest podstawą jej rozwoju².

Najczęściej, gdy myśli się o współdziałaniu pracowników nauki i biblioteki cyfrowej, chodzi o etap poszukiwania informacji na określony temat z zastosowaniem określonych narzędzi. Rola pracownika nauki jako wyszukującego jest dla wszystkich oczywista, gdyż biblioteki naukowe od zawsze służyły pomocą na tym etapie badań naukowych, a uczony poszukujący danych i informacji jest stałym i zwyczajnym widokiem w (tradycyjnej) bibliotece naukowej. Powstanie GBC przynosi w tym zakresie możliwość zacieśnienia współpracy pomiędzy bibliotekarzem i uczonym, na przykład w postaci wyszukiwania wspólnego, realizowanego w ramach funkcjonowania kolaboratorium. Niewątpliwie jednak powszechne jest przejmowanie przez uczonych w tym zakresie ról wcześniej typowych dla bibliotekarzy: pracownicy nauki, tak jak pozostali użytkownicy Internetu, sami wytyczają drogi poszukiwań informacji [Bojar 2009, s. 19].

Konwergencja ról przybiera także inne formy. Obecnie, dzięki nowym technologiom, w globalnej bibliotece cyfrowej uczeni funkcjonują jednocześnie jako autorzy i wydawcy swoich dzieł. W tej roli uczeni rzadko i raczej w szczególnych przypadkach pojawiali się dotychczas w bibliotekach naukowych, nie było więc potrzeby tworzenia dla nich specjalnych serwisów. Obecnie biblioteki (tradycyjne) coraz częściej wspomagają procesy samopublikowania treści naukowych, poprzez udostępnianie odpowiednich usług w tworzonych przez siebie serwisach GBC. Nowe role wymagają nowych umiejętności – od pracowników nauki oczekuje się na przykład dostarczenia tekstu zredagowanego i sformatowanego w sposób umożliwiający jego natychmiastowe rozpowszechnianie.

¹ Oprócz roli autorów i użytkowników pełnią oni także role selekcjonerów materiałów do publikacji, redaktorów, recenzentów, wydawców, administratorów.

² Niektórzy mówią w tym kontekście wręcz o „śmierci autora”, w sensie odejścia od romantycznej koncepcji autora jako jedynego twórcy. Akt twórczy jest niekompletny dopóki dzieło nie zostanie przeczytane, więc czytelnik staje się współtwórcą, interpretującym dzieło w sposób właściwy tylko dla siebie. Efektem takiego podejścia jest opinia, że dzieło uzyskuje pełne znaczenie dopiero po jego przeczytaniu. Na autora dzieła naukowego w szczególności wpływ mają także wcześniejsze dzieła innych autorów [Feather 2003, s. 100].

Różnicowanie ról ma znaczenie nie tylko teoretyczne: uczeni w zależności od wykonywanej roli mają różne potrzeby, przejawiają także różne formy zachowań [Sale 2006, s. 87]. W efekcie zmianie ulegają formy funkcjonowania różnego rodzaju serwisów globalnej biblioteki cyfrowej, takich jak na przykład repozytoriów instytucjonalnych czy dziedzinowych. W dalszej części rozdziału przedstawiam główne role uczonych w komunikacji naukowej.

Uczeni jako poszukujący i użytkownicy

Jednym z głównych zagadnień, związanych z tworzeniem wszelkiego rodzaju serwisów globalnej biblioteki cyfrowej, jest pytanie o profil (model) użytkownika tych serwisów. Uzyskanie odpowiedzi na to pytanie w globalnym środowisku informacyjnym nie jest łatwe z kilku powodów:

- Wspomniana globalizacja oznacza, że poszczególne serwisy, gromadząc zasoby lokalnie, udostępniają je globalnie, wszystkim użytkownikom Internetu. Globalny dostęp łączony jest ze zjawiskiem tzw. globalizacji, czyli specyficznym zaspokajaniem potrzeb lokalnych, co w nauce oznacza zwykle grupy specjalistów z wąskich dyscyplin. Wyniki badań realizowanych lokalnie dostępne są globalnie.
- Kolejną grupą użytkowników są instytucje wspomagające naukę i zarządzające, takie jak jednostki zarządzające w instytucjach naukowych, finansujących naukę oraz agendy rządowe. Ze względu na władzę posiadaną przez tę grupę, ich potrzeby metainformacyjne mogą stanowić niemal równie wysoki priorytet, jak potrzeby samych pracowników nauki. Potrzeby te mogą wymuszać tworzenie nikomu innemu niepotrzebnych meta-danych i możliwości wyszukiwawczych. Specyficzną grupę tego typu użytkowników stanowią badacze samych serwisów GBC.
- Ostatnia grupa może być określona jako pozostali użytkownicy; składa się ona na przykład z nauczycieli szkolnych, uczniów oraz wszystkich innych zainteresowanych. Większość badań adresowana jest do innego kręgu odbiorców, więc w publikacjach naukowych stosowany jest wysocy specjalistyczny język, utrudniający odbiór tej grupie użytkowników. Istnieje jednak pewna część badań i publikacji, która wyłamuje się z tej reguły. Najczęściej są to prace z zakresu zdrowia, ekologii i ochrony środowiska, polityki, historii, kultury i sztuki. Niektóre prace z innych dziedzin, szczególnie dotyczące zagadnień ogólnych, także mogą spotykać się z zainteresowaniem grup amatorów. Te osoby mogą zajmować się także tzw. Citizen Science, polegającą na wykonywaniu pewnych czynności badawczych (obserwacji, pomiarów, przetwarzania) przez połączone siecią osoby bez specjalnego szkolenia.

Głównym przedmiotem mojego zainteresowania w tym rozdziale jest grupa pracowników nauki. Mają oni zwykle wyraźnie określone priorytety, które mogą być zaskakujące dla twórców serwisów globalnej biblioteki cyfrowej. Uczeni oczekują, że:

- Możliwe jest szybkie i łatwe odszukanie tego, co zaspokaja potrzeby.
- Wszystko, co dana instytucja ma do zaoferowania, powinno być dostępne online.

Uczeni nie są zainteresowani zasobami trudnymi do pozyskania, przez co jeżeli drugi wymieniony warunek nie zostanie spełniony, to zasoby będą mało wykorzystywane. Żądają oni natychmiastowego dostępu do wszystkich, kiedykolwiek opublikowanych tekstów; ważne są szybkość dostępu i bogactwo wyboru [Nicholas i in. 2008b, s. 133]. Zubożonym wariantem realizacji tych potrzeb jest dostarczenie wyłącznie opisu bibliograficznego, bez możliwości dotarcia do pełnego tekstu. Większość pracowników nauki po prostu pominięto tego typu zasoby, gdyż uzna je za niewarte wysiłku, związanego z ich odszukaniem. Pewnym wyjściem z sytuacji jest tworzenie w rekordzie metadanych odnośników do usługi pozwalającej zamówić egzemplarz publikacji, co minimalizuje niezbędny wysiłek do kilku kliknięć (oraz czasem, w przypadku zasobów o płatnym dostępie, podania źródła finansowania). Istotną rolę odgrywa koszt pozyskania publikacji. Uczony zwykle będzie

się starał uzyskać do niej dostęp po jak najniższych kosztach. Najwyższe koszty zwykle generuje prenumerata indywidualna; niezbędne jest ponoszenie opłat za subskrypcję oraz kosztów przechowywania kolejnych roczników. Znacznie niższe koszty zapewniają dostęp poprzez bibliotekę naukową do tradycyjnych (koszt przyjscia do biblioteki), a zwłaszcza elektronicznych wersji czasopism komercyjnych. Najlepszą sytuację zapewniłaby po-
wieszna dostępność do publikacji w trybie Open Access.

Obecnie nie ma problemów technicznych z dostępem do naukowych publikacji elektronicznych. Instytucje nauki są w wystarczającym stopniu nasycone odpowiednim sprzętem komputerowym. Wciąż jednym z najważniejszych urządzeń, służących do korzystania z zasobów elektronicznych, jest drukarka; uczeni drukują materiały, które wyselekcjonowali jako interesujące, aby bliżej się z nimi zapoznać i poczynić notatki. Jak bowiem pisze Jacek Wojciechowski, tekst drukowany sprzyja starannemu czytaniu – czasopisma drukowane są starannie czytane przez 75% odbiorców, podczas gdy online – przez 50%. Papier jest także nośnikiem zapewniającym najlepszy komfort czytania [Wojciechowski 2007, s. 162]. Być może jednak dane te są odzwierciedleniem zmian w sposobach percepcji tekstu naukowego, o czym wspominałem w p. 2.6.

Trudno powiedzieć, czy tego typu zachowania ulegną zmianie, gdy w życie wejdzie pokolenie mające od dziecka do czynienia z komputerami oraz obrazami i tekstami na ekranie. Już obecnie zmiana zachowań komunikacyjnych jest tak duża, że można mówić o nowym e-stylu życia ludzi korzystających z Internetu, który stał się medium co najmniej równorzędnym do dawniej istniejących (z punktu widzenia częstości wykorzystania), a przy tym znacznie bardziej uniwersalnym (zastąpić może telefon i pocztę, telewizję i kino, gazetę i książkę, sklep i niektóre usługi oraz wiele innych) [Niezgoda 2006, s. 112-113]. Według Diany i Jamesa Oblingerów w latach 1982-1991 (w Polsce zapewne nieco później, być może o 10 lat) rozpoczęło życie pokolenie Sieci (*Net generation*), które uczyło się czytać na tekstach w Webie [Oblinger Diana, Oblinger James, red. 2005, s. 2.1-2.5]. Mówi się także o pokoleniu ekranowym (*screen generation*), od urodzenia korzystającym z przekazu telewizyjnego, a od lat szkolnych z komputera. Traktuje komunikaty złożone głównie z obrazów jako przekaz równorzędny wobec książkowego³ [Goban-Klas 2005, s. 108]. Osoby urodzone nieco później, w pierwszej połowie lat dziewięćdziesiątych XX wieku określane są nawet jako pokolenie Google (*Google generation*). Dorastali oni w świecie zdominowanym przez Internet [Rowlands i in. 2008, s. 292]. Marc Prensky nazywa ich „rdzennie cyfrowymi” (*digital natives*), w odróżnieniu od „cyfrowych imigrantów” (*digital immigrants*)⁴ [Prensky 2001, s. 3]. Autor twierdzi, że urodzeni w okresie funkcjonowania Internetu i Webu myślą oraz przetwarzają informacje w zasadniczo inny sposób, niż poprzednie pokolenia. Powołuje się na teorię neuroplastyczności, stwierdzając, że nowe doświadczenia powodują powstawanie odmiennych struktur mózgu, który ulega przemianom i reorganizacji pod wpływem impulsów zewnętrznych (w tym przypadku dostępu do sieciowych zasobów cyfrowych).

To do tego pokolenia stosują się wyniki badań OCLC, przedstawione w p. 4.3. Osoby te są zupełnie odmiennie pod względem zachowań i oczekiwań informacyjnych, od poprzednich pokoleń; funkcjonują wielozadaniowo (wykonują wiele zadań jednocześnie) i oczekują, że ich źródła informacji będą elektroniczne i dynamiczne. Ich podstawowym zasobem informacyjnym jest Web, w którym realizują wyszukiwania badawcze (*exploratory search* – zob. p. 4.3). Badania wskazują także, że młodzi pracownicy nauki z zakresu

³ Jest to przyczyną zjawisk negatywnych, gdyż odbiór komunikatów złożonych synkretycznie (strukturalnie), takich jak przekaz telewizyjny, redukuje myślenie intelektualne, a jeśli jest nadmierny – może utrudniać późniejsze czytanie dłuższych tekstów [Wojciechowski 2006, s. 25].

⁴ Wspomnieć można o innych określeniach stosowanych do młodych ludzi, wychowanych w otoczeniu urządzeń cyfrowych, takich jak: „pokolenie X”, „pokolenie Y”, „pokolenie M” (media), „pokolenie V” (*virtual*), „pokolenie C” (*connected, creative, click*) [Schwyn 2009, s. 365].

medycyny dwa razy częściej korzystają z czasopism w formie elektronicznej, niż ich starsi koledzy [Sathe, Grady, Giuse 2002]. Projektanci serwisów GBC muszą uwzględniać te cechy nowych użytkowników, nie można jednak oczekiwać, że ich kompetencje informacyjne automatycznie rosną dzięki umiejętności obsługi sieciowego sprzętu elektronicznego. Tak jak umiejętność czytania i pisania nie powoduje bezpośrednio posiadania wiedzy o stosowaniu bibliografii, podobnie nawet stały kontakt z komputerem nie musi oznaczać kompetencji informacyjnych. Stąd duża rola pracowników informacji jako pośredników i przewodników [Selwyn 2009, s. 374].

Kolejne priorytety uczonych, nieco mniej istotne, ale także wpływające na odbiór to:

- Określona jakość (wiarygodność) treści naukowych.
- Stosowanie zasad poprawnego tworzenia serwisów Web.

Wydawałoby się, że uczeni pragną dostępu do treści autoryzowanych, o określonej, wysokiej jakości, takich jak recenzowane artykuły w uznawanych czasopismach o wysokim współczynniku IF, jednak w rzeczywistości ich potrzeby są inne. Pracownicy nauki są zainteresowani w publikowaniu w tego typu czasopismach ze względu na prestiż i zasady oceny ich pracy, stosowane przez gremia zarządcze i finansujące. Ich głównym celem podczas poszukiwań literaturowych jest jednak dotarcie do przydatnego tekstu, nawet, jeśli jego pochodzenie nie jest autoryzowane, po to, aby samodzielnie określić jego poziom i przydatność w dalszych badaniach. Czasem nawet w niezbyt interesującym tekście trafić można na interesującą ideę, którą warto się zająć. W związku z tym użyteczna może być także wersja przed opublikowaniem (preprint), a nawet wersja nieformatowana i pozabawiona ilustracji, byle łatwo dostępna. Jeżeli publikacja okaże się interesująca, warto zaangażować więcej czasu i środków dla dotarcia do bardziej jakościowego źródła.

Ostatnia z wymienionych cech ma najniższy priorytet, ale nie jest bez znaczenia. Nie można wymagać, aby uczonego musiał przebić się przez różnego rodzaju zbędne dodatki i ozdobniki, aby dotrzeć do interesujących go treści. O kryteriach oceny poprawnego serwisu internetowego pisałem w p. 3.3. Stanowią one zbiór zasad, którymi powinni kierować się twórcy systemów, szczególnie tych, które mają stanowić serwis globalnej biblioteki cyfrowej. Niepoprawna strona Web serwisu w najlepszym wypadku jest nieekonomiczna; gorzej, gdy na tyle irytuje użytkownika, że zniechęca do korzystania. Z jednej strony zalecana jest prostota, z drugiej wszystkie niezbędne elementy muszą znajdować się we właściwym miejscu. Zamiast więc „wzbogacać” strony Web należy raczej zadbać o pełną dostępność tekstów online, jak najszerzą wyszukiwalność przy pomocy jak największej liczby narzędzi (przede wszystkim globalne wyszukiwarki) oraz zapewnienie dostępu do różnych wersji tekstu (na przykład na stronie Web czasopisma – płatny i w repozytorium – bezpłatny).

Uczeni jako autorzy

Wśród wielu ról, jakie odgrywają uczeni, bycie autorem pozwala na pozostawienie najtrwalszego śladu w zasobach literatury naukowej. Jako autorzy podejmują oni wiele decyzji, związanych z wypełnianiem kolejnych ról:

- Twórców, czyli podejmujących decyzje o możliwości pisania, dziedzinie, temacie i sposobie argumentacji, a także o sposobie układania zdań i wyborze odpowiednich słów.
- Cytujących lub linkujących, decydujących o wyborze publikacji do cytowania, odesłania czy wskazania w jakiś inny sposób.
- Dostarczycieli tekstów, decydujących o wyborze czasopisma lub innego źródła, w którym umieszczony będzie dokument.
- Współpracowników, decydujących o wyborze współautorów i instytucji afiliujących ich prace.

Decyzje podejmowane w przedstawionym powyżej zakresie mają istotny wpływ na funkcjonowanie globalnej biblioteki cyfrowej. Twórcą decydujący się na publikowanie

w formie elektronicznej uzyskuje dodatkowe korzyści, niedostępne w druku, na przykład możliwość uzupełnienia tekstu danymi z badań. Cytowania publikacji cyfrowych mogą odsyłać bezpośrednio do pełnego tekstu. Dzięki nowym formom publikowania elektronicznego znacznie rozszerzyły się możliwości wyboru miejsc, do których można dostarczać publikacje. W środowisku cyfrowym także współpraca naukowa nabrała nowego znaczenia, czego wyrazem jest budowa IIN.

Właściwe funkcjonowanie globalnej biblioteki cyfrowej umożliwia optymalne rozpowszechnianie wyników pracy naukowej. Środowisko cyfrowe daje możliwość najszybszego docierania do starannie wyselekcjonowanego odbiorcy. Autorem publikacji naukowych zależy na jak najszerszym rozpowszechnieniu ich publikacji. Wynik badań naukowych, aby można go było uznać za wartościowy, musi być poznany, wykorzystany i oceniony przez innych. John Willinsky pisze w tym kontekście o „próżności” uczonych, dumnych z widoku swojego nazwiska w druku lub tytułu swojej pracy w cytacie kolegi [Willinsky 2006, s. 21]. Autorzy najlepszych publikacji mogą liczyć na cytowania, umieszczanie na listach publikacji w ważnych instytucjach, zaproszenia do wygłoszenia referatów. Na podstawie tych zdarzeń uczone może ocenić przydatność i poprawność ponoszonych wysiłków i funduszy wydatkowanych na własne badania. Właściwe rozpowszechnienie wyników prowadzi zwykle (choć niebezpośrednio) do korzyści finansowych w postaci prestiżu, stanowisk, promocji i wzrostu liczby grantów. Stevan Harnad podzielił autorów na komercyjnych i „ezoterycznych” (uczonych). Ci pierwsi tworzą publikacje w celu uzyskania bezpośrednich zysków, natomiast uczeni publikują dla osiągnięcia wymienionych korzyści pośrednich, a nawet po prostu dla przyjemności, jaką daje rozpowszechnianie wiedzy. Stąd interesy komercyjnych autorów i wydawców są zbieżne – obie grupy chcą ograniczyć dostęp do publikacji tylko do tych odbiorców, którzy za niego zapłacą. Potrzeby autorów ezoterycznych i ich wydawców są, przynajmniej w tym zakresie, sprzeczne, gdyż w interesie autorów nie leżą żadne ograniczenia w dostępie do ich publikacji [Harnad 1998, s. 186]. Właściwe rozpowszechnianie nie oznacza jednak konieczności dotarcia z publikacją tak szeroko, jak to tylko możliwe – dla uczonego wystarczające jest, aby jego dzieło znane było właściwym specjalistom i decydentom, często nie więcej niż stu osobom [Parks 2007, s. 326].

W warunkach publikowania tradycyjnego autorzy, ze współpracą z wydawcami, osiągają pewne korzyści. Proces technologiczny przygotowania publikacji, a także jej dystrybucja w druku są na tyle skomplikowane, że autorom oplaca się współpraca z kimś, kto wykona profesjonalnie te czynności za nich. Z jednej strony wydawcy utrudniają bowiem dostęp, żądając za swe usługi opłat od użytkowników, ale z drugiej w ogóle umożliwiają jakikolwiek dostęp. Dzięki zastosowaniu nośników cyfrowych w globalnym środowisku elektronicznym potrzeba pośrednictwa wydawców pomiędzy uczonymi znika.

Obecnie, dla osiągnięcia wymienionych wcześniej korzyści, pracownicy nauki muszą, poprzez swoje publikacje, zaistnieć w sieci. Mogą w tym celu oczekiwać na efekty aktywności wydawców i często się doczekują, gdyż obecnie wszystkie poważne czasopisma naukowe mają swoje wersje sieciowe. Nie musi to jednak oznaczać optymalnego rozpowszechniania, gdyż dla użytkownika końcowego dostęp w Sieci niekoniecznie jest bezpłatny. Dla zwiększenia wpływu (*impact*) swoich publikacji autorzy decydują się na autopublikowanie na własnych stronach Web lub w repozytoriach, gdzie oprócz pełnych tekstów publikacji podają swoje CV, osiągnięcia, przebieg pracy dydaktycznej. Podjęcie tego typu decyzji wiąże się z koniecznością dokonania selekcji (wyboru) materiału do opublikowania i wyboru miejsca publikowania. W przypadku umieszczania publikacji w repozytorium wystarczy wstawić odnośnik ze strony domowej do stałego adresu w repozytorium. Na ogół możliwe jest także udostępnienie jedynie odnośnika typu „Zobacz moje publikacje”, powodującego wyszukiwanie we właściwym repozytorium według określonego nazwiska. W ten sposób autorzy stają się wydawcami własnych prac,

korzystając z faktu, że publikowanie w Web jest nadzwyczaj proste i tanie, szczególnie, gdy istnieje gotowy eprint do udostępnienia.

Często do rozpoczęcia autopublikowania potrzebne jest przełamanie pewnej bariery psychologicznej. Po części wynika ona z obawy przed wchodzeniem w zakres działania innych specjalistów (redaktorów, wydawców), częściowo z nieznajomości zagadnień prawa autorskiego lub odpowiednich technologii. Przełamanie tej bariery powoduje często gwałtowny wzrost zainteresowania pracami danego autora, rozwój nowych kontaktów i dyskusji (m.in. w formie poczty elektronicznej, list dyskusyjnych), co powoduje, że większość uczonych staje się zagorzałymi zwolennikami umieszczania swych prac w Internecie.

Decyzje autorów o rodzaju materiałów i czasie ich rozpowszechnienia online, zależy od wielu czynników, z których część jest jeszcze nie do końca zrozumiała. Niektórzy skłaniają się do umieszczania w Sieci wszystkich swoich publikacji lub kompilacji (materiałów dydaktycznych, zestawów danych), natychmiast po ich przygotowaniu z nadzieją na jak najszerzy odbiór. Inni nie umieszczają w Sieci żadnych dokumentów, pozostawiając to fachowym wydawcom. Pomiędzy tymi dwiema skrajnościami znajduje się całe spektrum zachowań, polegających na umieszczaniu w Internecie dokumentów przed, po i jednocześnie (a także zamiast) z publikowaniem poprzez tradycyjne kanały publikowania.

Osobiste strony Web są co prawda wystarczającym narzędziem dla rozpowszechniania publikacji, jednak do realizacji funkcji archiwizacji nadają się znacznie gorzej. Strony te nie zapewniają długotrwałej dostępności ani skalowalności³, cechujące rozwiązania stosowane w bardziej zinstytucjonalizowanych serwisach GBC.

Autopublikowanie, oprócz korzyści, nakłada na autorów także pewne dodatkowe wy-mogi. Autorzy odpowiedzialni są między innymi za:

- Zgodność publikacji ze standardami stosowanymi w danej społeczności, dotyczący-mi zapisu danych i metadanych adekwatnie opisującymi kontekst i jakość danych oraz wspomagającymi wyszukiwanie i stosowanie danych.
- Udostępnianie danych, najlepiej w trybie Open Access, zgodne z obowiązującymi stan-dardami, dotyczącymi bezpieczeństwa, prywatności i praw autorskich.
- Zgodność z przyjętymi standardami dotyczącymi rodzaju, jakości i treści danych, wraz z dołączonymi metadanymi pozwalającymi na umieszczenie danych w odpowiednim repozytorium.
- Realizację zaleceń związanych z zarządzaniem danymi, określonymi w grantach, kon-traktach, umowach z fundatorami.
- Zapewnienie możliwie długotrwałej archiwizacji danych, na przykład przez umieszcze-nie w odpowiednim repozytorium, a przynajmniej określenie przewidywanego okresu ich przechowywania.

Wymienione działania stają się częścią warsztatu uczonego. Coraz powszechniej zda-rza się, że autorzy odpowiedzialność za te prace przenoszą na bibliotekarzy, którzy zajmu-ją się ich publikowaniem przejmując zadania komercyjnych wydawców.

Uczeni jako recenzenci

Recenzenci zazwyczaj wykonują swoje obowiązki nieodpłatnie, lub za symboliczną opłatą. Uznaje się, że ich praca traktowana jest jako usługa na rzecz środowiska, dziedziny i specjalności, jaką reprezentują, a korzyści osobiste polegają na wzmoczeniu reputacji i pozycji w środowisku. Stopień wzrostu osobistej reputacji zależy od reputacji czasopisma, z którym uczone współpracuje jako recenzent. Na ogół jednak korzyści z przyjęcia funkcji recenzenta są pośrednie.

³ Skalowalność jest to cecha systemów, sieci lub procesów komputerowych polegająca na zdolności do rozbudowy wraz ze wzrostem zapotrzebowania na usługi lub powiększaniem się zasobów (sprzętowych, danych). Rozbudowa taka powinna przebiegać w sposób jak najmniej uciążliwy dla użytkowników.

Ze względu na zazwyczaj wyższą reputację czasopism komercyjnych (choćby wyjątki od tej reguły), redakcyjnym tych czasopism łatwiej jest stworzyć grupę kompetentnych recenzentów. Decyzje podejmowane przez uczonych w zakresie współpracy z określoną redakcją, poza reputacją czasopisma, uzależnione są od wielu innych czynników, na przykład aktywności uczonego w innych rolach, w tym jako autora; jeżeli uczony sam publikuje w czasopismach Open Access, to zapewne łatwiej również przystanie na recenzowanie artykułów w takim czasopiśmie.

W efekcie cyfryzacji procesu publikowania naukowego zmianie ulega także rola recenzenta. Tradycyjne recenzowanie nie wytrzymuje próby czasu, uczeni jako recenzenci są krytykowanymi m.in. za tworzenie „klik”, utrudniających rozwój nowych idei [Poynder 2008b]. Prowadzone są prace i eksperymenty zmieniające rolę recenzenta. Tradycyjne recenzje uzupełniane są, a nawet zastępowane, tzw. recenzowaniem otwartym, a więc komentowaniem na bieżąco nowych publikacji, jednak już po ich opublikowaniu. W ten sposób w publikowaniu naukowym wykorzystywane są zasady funkcjonujące w Webie społecznościowym, w tym tzw. inteligencja zbiorowa, chociaż „zbiór” osób uczestniczących w takim recenzowaniu musi być oczywiście bardzo ograniczony, do grona osób zainteresowanych danym zagadnieniem naukowym.

Powstanie globalnej biblioteki cyfrowej wywołuje wiele dyskusji dotyczących przyszłości zawodu **bibliotekarza**⁶ oraz **pracownika informacji**, w tym potrzeby uzyskiwania przez nich nowych umiejętności [Isfandyari-Moghaddam, Bayat 2008, s. 852; Nahotko 2006a; Pindłowa 2000] i cech [Zhou 2005, s. 437]. Nie może to być przy tym wysiłek jednorazowy, niezbędny jest ciągły rozwój zawodowy, gdyż zmiany w komunikacji naukowej i jej technologiach stale postępują. Zmiany te, związane głównie z rozwojem technologii informacyjnej (ICT), traktowane bywają przez bibliotekarzy ambiwalentnie – z jednej strony obawiamy się, że użytkownicy mogą odwrócić się od bibliotek naukowych z ich tradycyjnymi zasobami, opisywanymi w OPAC, uważając, że to, czego nie ma w Internecie (czyli nie da się wyszukać w Google) nie jest warte zachodu, a z drugiej strony zauważamy niepowtarzalne szanse związane z Webem, takie jak możliwość szerokiej prezentacji swojej oferty usług, a także prowadzenie działalności wydawniczej w formie tworzenia bibliotek cyfrowych. Paradoksalne jest to, że wraz ze wzrostem liczby danych, udostępnianych przez biblioteki, a „widzianych” przez wyszukiwarki internetowe, wzrasta efektywność tych ostatnich jako narzędzia wyszukiwawczego, a w związku z tym stają się one coraz większym zagrożeniem dla tradycyjnych usług bibliotecznych, realizowanych głównie poprzez OPAC [Pearce, Gatenby 2005]. Nie musi to jednak oznaczać końca usług informacyjnych biblioteki. Jak piszą autorzy raportu dotyczącego pokolenia Google, umiejętności korzystania z technologii informacyjnych nie muszą iść w parze z kompetencjami informacyjnymi [Centre for Information Behaviour and the Evaluation of Research 2008, s. 20]. Coraz większa liczba dostępnych materiałów w Sieci powoduje więc coraz większe zapotrzebowanie na medycyjnne usługi specjalistów w zakresie działalności informacyjnej [Golka 2008, s. 156; Williams, Rowlands, Fieldhouse 2008, s. 166]. Na ogół wszyscy zgadzają się, że rola bibliotekarza ulega zmianie⁷. Wynika to z nowych potrzeb społeczeństwa sieciowego, przejawiających się takimi zjawiskami, jak szybki rozwój rynku informacji i usług w zakresie wiedzy i przetwarzania informacji [Barney 2008, s. 95]. Już obecnie absolwenci kierunków bibliotekoznawczych często znajdują zatrudnienie poza instytucjami tradycyjnie oferującym im pracę, zmianie ulegają także formy zatrudnienia. Świadczą o tym także stosowane i proponowane nowe nazwy tego fachu, których długą listę przedstawiła Katarzyna Materska [Materska 2007, s. 354-355].

⁶ Por. np. materiały z II Konferencji Bibl. Głównej Polit. Łódzkiej: Biblioteki XXI wieku: czy przetrwamy? <http://www.ebib.info/publikacje/matkonf/biblio21/>.

⁷ W dalszej części rozdziału zajmują się typowymi funkcjami zawodu bibliotekarskiego, pomijając fakt, że bibliotekarze są również pracownikami nauki i/lub administratorami.

Zaangażowanie się bibliotekarzy w proces tworzenia publikacji cyfrowych powoduje przejmowanie przez nich wielu ról, typowych wcześniej dla innych zawodów, związanych z publikowaniem tradycyjnym. Bibliotekarze tworzący biblioteki cyfrowe wykonują czynności typowe dla drukarzy, a więc techniczne czynności służące przygotowaniu publikacji, takie jak przygotowanie składu, montaż publikacji (impozycja), skanowanie i konwersja plików (w tym wybór formatów). W dalszej kolejności niezbędne jest wykonywanie czynności obciążających wcześniej wydawnictwa, takich jak przygotowanie umów licencyjnych, orientacja w sytuacji prawnej związanej z zasadami prawa autorskiego, czasem prace redakcyjne. Niejednokrotnie bibliotekarze muszą także dbać o zapewnienie technicznych możliwości udostępniania przygotowanych zasobów, poprzez zapewnienie dostępu sieciowego, ciągłej pracy serwerów i łączy komputerowych.

Transformacja bibliotek naukowych z wyizolowanej biblioteki tradycyjnej w kierunku serwisu GBC powoduje, że bibliotekarze muszą funkcjonować w większym stopniu w reżimie zorientowanym na zarządzanie projektami [Dahl, Banerjee, Spali 2006, s. 175]. Realizacja wielu funkcji biblioteki tradycyjnej, takich jak katalogowanie i działalność informacyjna, oparte są na działaniach rutynowych. Większość działań realizowanych w serwisach GBC, takich jak na przykład projektowanie portalu bibliotecznego lub internetowych narzędzi wyszukiwawczych, ma natomiast charakter realizacji projektów. W związku z tym bibliotekarze powinni zdobywać umiejętności zarządzania projektami i posługiwania się oprogramowaniem wspomagającym to zarządzanie.

Z tym związane jest wspomniane już zjawisko wykorzystywania niestandardowych form zatrudnienia, w tym głównie telepracy, na której zalety wskazuje Marian Golka [Golka 2008, s. 91]. Oprócz telepracy wymienić można „telecommuting”⁸, samozatrudnienie, consulting, pracę jako „wolny strzelec” [Barney 2008, s. 114]; wszystkie powodują rozluźnienie związków między pracownikiem a pracodawcą oraz zjawiska związane z czasową i przestrzenną dyslokacją pracy.

Bardzo ważną wydaje się rola bibliotekarzy jako osób łączących, za pomocą swojej wiedzy i umiejętności, dwa odrębne światy, które funkcjonują obecnie obok siebie: świat publikacji tradycyjnych i cyfrowych, wraz ze wszystkimi towarzyszącymi im narzędziami. W bibliotekach można uzyskać informacje zarówno o publikacjach drukowanych, jak i internetowych; bibliotekarze zajmują się organizacją tej metainformacji oraz organizują dostęp do każdego rodzaju materiałów, bez względu na medium. Biblioteki naukowe mogą być instytucjami wspomagającymi użytkowników zagubionych w cyberświecie oraz archiwizującymi wartościowe zasoby cyfrowe. Taka sytuacja może trwać jeszcze długo, przynajmniej do czasu zdigitalizowania znacznej większości publikacji tradycyjnych oraz pełnej cyfryzacji bieżącej produkcji wydawniczej. Pełna digitalizacja także jednak nie eliminuje potrzeby pośrednictwa bibliotekarza, gdyż wzrasta jego rola mediacyjna, służąca opracowaniu podaży zasobów informacyjnych przez ich selekcję: im więcej więc materiałów w Sieci, tym bardziej wzrasta zapotrzebowanie na usługi informacyjne.

Innym obszarem stanowiącym przyszłość zawodu bibliotekarza jest uczestnictwo w dwóch etapach procesu realizacji badań naukowych: przed publikowaniem wyników i po pierwszej ich publikacji [Gold 2007]. Bibliotekarze od dawna funkcjonują w części cyklu życia działań naukowych, których efektem są publikacje. Zanim jednak wyniki badań zostaną opublikowane, efektem prac badawczych są zestawy nieprzetworzonych danych. Bibliotekarze powinni włączać się także w funkcjonowanie tej części cyklu życia badań. Oprócz pracy z publikacjami, powinni oni zainteresować się także zestawami danych, wymienianymi pomiędzy pracownikami nauki przed fazą publikacji. Środowisko cyfrowe pozwala na tworzenie dynamicznych i płynnych związków w trakcie cyklu życia danych, w tym głównie na ich wielokrotne wykorzystywanie (*reuse*). Odbywa się ono zarówno

⁸ Telecommuting to praca wykonywana okresowo w domu, ale na stałe umiejscowiona w standardowym miejscu pracy.

przed, jak i po ich opublikowaniu. Wytyczenie ścisłej granicy pomiędzy etapami cyklu życia danych przed i po opublikowaniu jest nieco sztuczne, może jednak być użyteczne, szczególnie ze względu na to, że tradycyjne role i zadania bibliotek naukowych mają tendencję umieszczania się w zasięgu danych po opublikowaniu, a nowe role są sadowane bardziej w kierunku cyklu badań naukowych realizowanych przed publikowaniem wyników, co pozwala bibliotekarzom na uczestnictwo w procesach ochrony danych.

Role bibliotekarza i zarządzającego danymi, choć różne, mają wiele podobieństw. Dla zrozumienia i opisanie tych relacji przydatny może być model „provincji wiedzy” przedstawiony przez Karen Baker i Florence Millerand [Baker, Millerand 2007]. Model ten prezentuje role i relacje pomiędzy zarządzaniem danymi, zarządzaniem informacją oraz informatyczną infrastrukturą nauki (nazywane prowincjami wiedzy), w zakresie badań naukowych realizowanych w różnej skali, od małych do wielkich i od prostych do złożonych. Prowincja zarządzania danymi skupia się na prostych danych i zautomatyzowanych, zróżnicowanych (od małych do dużych) strumieniach danych; obszar zarządzania informacją zajmuje się wieloma odmiennymi rodzajami danych oraz tworzeniem relacji między nimi; natomiast infrastruktura nauki rozciąga się pomiędzy archiwizacją i udostępnianiem wielkich zasobów danych oraz złożonymi przepływami strumieni danych, typowymi dla narodowych ośrodków informacji.

Tworząc biblioteki cyfrowe, w których umieszczane są zdigitalizowane wersje tradycyjnych publikacji posiadanych przez biblioteki, bibliotekarze stają się wydawcami publikacji cyfrowych. Publikacje oraz ich metadane są najczęściej dostępne dla narzędzi indeksujących wyszukiwarek. Ze względu na prawa autorskie w bibliotekach cyfrowych zazwyczaj publikowane są dzieła powstałe wystarczająco dawno. Bibliotekarze dokonują selekcji publikacji do zdigitalizowania, ich digitalizacji, przygotowują metadane oraz publikują wersje cyfrowe poprzez ich umieszczenie w sieci, zazwyczaj udostępniając je w trybie Open Access. W ramach tej działalności nie dokonują oni selekcji z punktu widzenia jakości publikacji; selekcja taka nie była także udziałem bibliotek tradycyjnych, gdyż dokonywał jej wydawca publikacji. Bibliotekarze, kiedyś i obecnie, dokonują tylko selekcji z punktu widzenia potrzeb użytkowników; kiedyś kupowali, a teraz digitalizują to, co uważają za warte udostępnienia użytkownikom w pierwszej kolejności. W ten sposób porządkują zasoby według kryterium udostępniania i dostępności (fizycznej i intelektualnej), na pierwszym miejscu stawiając dokumenty, o które użytkownicy najczęściej pytają [Lancaster 2003, s. 138].

Inna forma działalności bibliotekarzy, związana z cyfrowym publikowaniem, może wiązać się z długotrwałą archiwizacją materiałów, powstających w wyniku działalności naukowej. Materiały te w większości są obiektami tekstowymi, gromadzonymi w serwisach GBC, które funkcjonują niezależnie od klasycznego procesu publikowania, przez co mogą zajmować różne miejsca w cyklu działalności naukowej, zarówno przed, jak i po publikacji, łącząc ze sobą obie części dokonanego tu rozdziału roli bibliotek naukowych. Istnieją na przykład plany tworzenia systemu repozytoriów lokalnych, gromadzących dane na wstępnych etapach ich istnienia, w połączeniu z repozytoriami centralnymi, przejmującymi te dane po przejściu do kolejnych etapów procesu publikowania, w celu ich archiwizacji. Ponieważ główne cele tych przedsięwzięć – zapewnienie długotrwałej ochrony i archiwizacji materiałów naukowych są od dawna realizowane w bibliotekach naukowych, mogą one uczestniczyć w tych przedsięwzięciach, wykonując prace polegające na wyrażaniu semantyki heterogenicznych zasobów danych, dostarczaniu funkcji obsługi użytkowników oraz stymulowaniu dystrybucji wyników badań naukowych.

Biblioteki naukowe mogą także współpracować na podobnych zasadach z wydawcami komercyjnymi, przejmując od nich dane przeznaczone do długotrwałej archiwizacji. Tego typu działalność może być realizowana po formalnym opublikowaniu materiałów, a także z zastosowaniem pewnego okresu, w którym zarządzanie nimi pozostawać będzie w gestii wydawcy (*moving wall*).

Role bibliotekarzy w zarządzaniu cyfrowymi danymi i publikacjami naukowymi wciąż nie są do końca określone, co znajduje odbicie w prowadzonych dyskusjach i eksperymentach służących próbom doskonalenia zasad postępowania. Eksperymenty te nie są jednak na ogół prowadzone w kierunku systemów globalnych i IIN. Przyczyna może leżeć w tym, że IIN wykorzystywana jest najczęściej do tworzenia skomplikowanych i wielkich systemów, służących działalności naukowej na dużą skalę. Nie wszystkie jednak badania naukowe realizowane są w takiej skali, a nawet, gdy skala jest duża, występują znaczne różnice w zakresie stosowanej infrastruktury, sposobów działania oraz warunków współpracy interdyscyplinarnej. Rola bibliotekarzy w przedsięwzięciach na wielką skalę jest szczególnie nieokreślona, na co składają się problemy z brakiem doświadczeń w obsłudze wielkich zasobów danych, stosowaniem wielkich mocy obliczeniowych i przetwarzania danych. Potrzeby w zakresie infrastruktury dla badań na średnią i małą skalę, często prowadzone lokalnie na podstawie miejscowych doświadczeń i zasad, stanowią odrębne zagadnienie.

Zarządzanie danymi, tworzoneymi przez zespół badawczy i jednocześnie obsługa informacyjna tego zespołu wymaga większego rozproszenia pracowników, wykonujących te zadania. Podobnie jak rozproszone są zasoby i użytkownicy globalnej biblioteki cyfrowej, tak rozproszeni są jej bibliotekarze. Stają się oni częścią zespołów badawczych, kolaboratorów, a nie pracownikami instytucji bibliotecznych, utrzymującymi tylko odległe relacje z pracownikami nauki. Powoduje to wnikanie bibliotekarzy do środowisk naukowych, co wywiera poważny wpływ na działalność biblioteczną, powodując rozwój bibliotekarstwa uczestniczącego. Taka tendencja zgodna jest z poglądami o potrzebie częstszego opuszczania bibliotek przez bibliotekarzy w poszukiwaniu relacji z innymi użytkownikami, którzy nie przychodzą już do bibliotek, bo nie mają takiej potrzeby. Relacje te są bibliotekarzom niezbędne dla rozumienia, a w rezultacie realizacji potrzeb użytkowników [Plutchak 2007, s. 88].

Warto wspomnieć o roli specyficznej grupy specjalistów – bibliotekoznawców. Tak, jak posiadanie wiedzy z obsługiwanego dziedziny nauki może być istotne dla bibliotekarzy zarządzających danymi i publikacjami, również bibliotekoznawcy mogą przedstawiać interesujące poglądy na relacje pomiędzy różnymi częściami zasobów publikacji naukowych, na wzorce postępowania podczas wyszukiwania i użytkowania informacji, a także na możliwości długotrwałego i interdyscyplinarnego wykorzystania wyników badań. Zagadnienia te mogą umknąć specjalistom z innych dziedzin wiedzy.

Środowiskiem, w którym można spodziewać się znacznych zmian, a które przed tymi zmianami broni się najmocniej, jest środowisko **wydawców komercyjnych**. Część autorów piszących o przyszłości publikacji naukowych przewiduje całkowitą likwidację tej profesji, jako zbędnego pośrednika między autorem a użytkownikiem [Tenopir, King 2000]. Często łączy się to z przewidywaniami zaniku czasopism elektronicznych na rzecz repozytoriów lub innych sposobów publikowania elektronicznego.

Trudno jednak oczekiwać tak radykalnych zmian w przewidywalnym czasie. Wydawcy utrzymują swoją pozycję na rynku, z jednej strony dzięki opisanym już działaniom marketingowym, a z drugiej dzięki świadczeniu pewnej wartości dodanej, do której przyzwyczajeni są wszyscy korzystający z publikacji tradycyjnych, a która to wartość, według wymienionych autorów, nie jest niezbędna w środowisku cyfrowym. Oprócz samego publikowania, wydawcy świadczą usługi polegające na organizowaniu procesów prowadzących do selekcji tekstów odpowiadających ustalonym kryteriom poprawności (głównie poprzez recenzowanie) oraz redakcji technicznej i merytorycznej tekstu. Według Petera Subera, wydawcy oferują jednak wyraźnie mniejszą wartość dodaną niż autorzy i fundatorzy badań, jednocześnie żądając największych lub wyłącznych zysków, będąc powodem tworzenia sytuacji sztucznego niedoboru poprzez kontrolę dostępu [Suber 2008]. Być może wydawcy komercyjni w przyszłości zajmą się rozpowszechnianiem komercyjnych treści naukowych, przeznaczonych na sprzedaż, takich, jak specyfikacje

nowych technologii. Rozwój biblioteki globalnej może silnie zmienić rynek komercyjnej komunikacji naukowej, ze względu na coraz bardziej powszechny dostęp Open Access do treści naukowych. Wydawcy starają się utrzymać pewną równowagę pomiędzy tymi sektorami, „uwalniając” część treści o mniejszym znaczeniu komercyjnym (na przykład publikacje starsze). Poszukują także aktywnie nowych modeli publikowania w ramach OA, w tym nowych źródeł finansowania działalności.

Rynek wydawniczy publikacji naukowych zdominowany został przez wydawców komercyjnych, którzy swoje przedsiębiorstwa tworzyli często przez przejęcia tytułów (czasopism, serii) publikowanych wcześniej przez wydawców uczelnianych. Rynek ten przeżywa obecnie gwałtowną komasację; połowa publikacji naukowych na świecie wydawana jest obecnie przez kilka koncernów [Nahotko 2007b, s. 23]. Koncerny te opanowały także produkcję pokrewną, na przykład multimediiów i gier komputerowych.

Wydawcy komercyjni, podobnie jak inne przedsiębiorstwa komercyjne, funkcjonują w celu osiągnięcia zysków. Z ich punktu widzenia publikacje elektroniczne stwarzają kłopoty, spowodowane koniecznością zmiany modelu ekonomicznego i sposobu pracy. Nowe możliwości, wynikające z technologii cyfrowej, nie są dla nich najważniejsze, a nawet traktowane są jako zagrożenie, o ile nie zwiększają zysku. Wydawcom wszystkie publikacje, również publikacje elektroniczne, kojarzą się z dochodem wynikającym z pracy nad tekstem, natomiast wybór formatu (tradycyjnego czy cyfrowego) jest mniej istotny [Fether 2003, s. 168]. W systemie komunikacji tradycyjnej osiągnęli oni zyski, jednocześnie dostarczając korzyści środowisku naukowemu, w postaci usług na tyle specjalistycznych, że nie opłacała się ich realizacja bezpośrednio przez uczonych. W środowisku elektronicznym potrzeba takiej specjalizacji i podziału prac już nie jest taka oczywista. Demokryzacja publikowania, zawdzięczana nowemu medium, silnie związanemu z ICT, może dla wydawców oznaczać także utratę władzy; źródłem autorytetu staje się maszyna [Wenz 2008, s. 115]. Podobnie jak bibliotekarze, również wydawcy mogą obawiać się nowych technologii umożliwiających globalny dostęp do danych i informacji. Paradoksalnie obie te grupy bardzo szybko włączyły się do prac nad ich wykorzystaniem. Wydawcy przy tym wykonują zadania, w których niełatwo ich zastąpić. Zwraca się uwagę na potrzebę realizacji takich usług, jak redakcja językowa, ujednoczenie formy publikacji, marketing, generacja przychodów z reklam, współpraca z serwisami indeksowymi. Nowe zadania wydawców wiążą się także z profesjonalnym prowadzeniem serwisów internetowych oraz długotrwałą archiwizacją.

Role wydawców była zawsze realizacja funkcji rozpowszechniania publikacji naukowych. W środowisku cyfrowym rozpowszechnianie może być trudne do odróżnienia od udostępniania. Wydawcy zaprzestają dostarczania obiektów fizycznych do bibliotek, które wcześniej realizowały funkcję udostępniania, zamiast tego umieszczają obiekty cyfrowe we własnych bazach danych, przejmując tę funkcję. Bazy te oprócz opisów bibliograficznych i abstraktów coraz częściej zawierają pełne teksty dokumentów, a nawet indeksy cytowań. W ten sposób wydawcy zachowują kolejną istotną przypisywaną im rolę: selekcję tekstów do publikacji, prowadzoną najczęściej na podstawie oceny jakościowej.

Środowisko elektroniczne niesie dla wydawców zmiany, zarówno negatywne, jak i pozytywne, zmuszając ich do modyfikacji sposobu funkcjonowania. Wśród efektów negatywnych cyfryzacji (z punktu widzenia wydawców) można wymienić:

- Łatwość publikowania. Przed rozwojem technik komputerowych: edytorów tekstów, drukarek laserowych, techniczne przygotowanie publikacji było zbyt skomplikowane, czasochłonne i kosztowne. Aby opłacało się wykonywać te prace indywidualnie. Obecnie większość tekstów przygotowywanych do publikacji powstaje od razu w formie elektronicznej, która w bardzo prosty sposób może być udostępniona w Internecie, przy czym wygląd tak przygotowanej publikacji nie odbiega od wydawnictwa profesjonalnego.
- Przejęcie zadań wydawców przez inne grupy. Te zmiany, o których już pisałem, dotyczą

zarówno autorów, jak i bibliotekarzy. Ze względu na repertuar, biblioteki jako wydawnictwa nie stanowią zasadniczego zagrożenia dla wydawnictw; obie te instytucje nie konkurują ze sobą, gdyż biblioteki upubliczniają na ogół dzieła, których wydania raczej nie podjęłoby się wydawnictwo komercyjne. Dla wydawnictw większym zagrożeniem mogą być bardziej radykalne odłamy ruchu Open Access.

- Open Access. Bezpłatne i powszechne udostępnianie publikacji naukowych w Internecie, jak najczęściej definiowane jest zjawisko Open Access, z pewnością może stanowić zagrożenie dla interesów wydawców (i tak często jest przez nich postrzegane). Ruch Open Access nie jest jednak jednorodny; oprócz dość radykalnych poglądów istnieją również inne zapatrywania. Przewiduje się funkcjonowanie publikacji Open Access nie zamiast, ale oprócz publikacji tradycyjnych, których wersje elektroniczne byłyby uzupełnieniem, przyspieszającym rozwój nauki dzięki poprawie funkcjonowania komunikacji naukowej. Pozostawiana jest wydawcom część ich zadań, w tym szczególnie w zakresie oceny jakości; wydawcy powinni zadbać o jakość publikacji organizując proces recenzowania, natomiast zrecenzowane publikacje, po ukazaniu się drukiem, powinny zostać umieszczone w Sieci i tam bezpłatnie udostępniane. W odpowiedzi wydawcy sami uczynią eksperymentować z OA, poszukując nowych modeli efektywnych finansowo.
- Specyfika Internetu jako środowiska wydawniczego sprawia, że paradoksalnie lepiej nadaje się ono do publikowania krótszych form o charakterze naukowym (jak artykuły, rozdziały z większych całości, notatki w blogach) oraz bardzo długich (obszernych) treści zawierających wyniki obliczeń towarzyszących badaniom naukowym (zestawy danych). Internet oraz środowisko elektroniczne znacznie gorzej nadaje się do publikowania i odbioru długich tekstów beletrystycznych, esejów. Wynika z tego, że do opublikowania w formie elektronicznej najlepiej nadają się materiały stanowiące podstawę komunikacji naukowej: pełne dane, stanowiące wynik badań, wraz ze stosunkowo krótkim opisem w postaci artykułu czy raportu. Dlatego funkcjonowanie Internetu szczególnie naraża na szwank pozycję wydawców, specjalizujących się w publikacjach naukowych, niebeletrystycznych. Podobnie część publikacji popularyzacyjnych o charakterze informacyjnym (słowniki, encyklopedie) znacznie lepiej funkcjonuje jako publikacje elektroniczne w formie baz danych.
- Nowe formy recenzji. Kontrola jakości publikacji traktowana jest jako jedna z najważniejszych wartości dodanych, tworzonych przez wydawców. Recenzowanie jest, co prawda, realizowane zazwyczaj przez samych pracowników nauki, jednak wydawcy wskazują na ten proces, jako jeden z najbardziej czasochłonnnych i kosztownych. Środowisko elektroniczne globalnej biblioteki cyfrowej także w tym zakresie oferuje nowe rozwiązania, niosące ze sobą zazwyczaj także nowe możliwości (nową wartość dodaną). Wydawcy dostrzegają także nowe szanse, których oczywiście nie omieszkają wykorzystywać:
- Łatwość dystrybucji publikacji w sieci: możliwość bardzo szybkiej dystrybucji dowolnych treści do ściśle określonego odbiorcy oraz łatwość i wygoda ich odbioru.
- Konserwatywnym środowiska naukowego powoduje, że publikacje tradycyjne wciąż uznawane są za lepsze niż elektroniczne: łatwiej uzyskać na nie fundusze, wyżej oceniane są w dorobku autora, wyżej cenione są także ich recenzje. Zapewniają autorom i recenzentom większy prestiż. Według części zwolenników OA zmianę tej sytuacji spowodują przepisy, obligatoryjnie nakazujące umieszczanie w Open Access publikacji wynikających z realizacji badań, finansowanych z środków publicznych. Lobby wydawców sprzeciwia się także tym rozwiązaniom.
- Biblioteki naukowe prenumerują czasopisma w wersji elektronicznej, co znacznie ułatwia realizację prenumeraty i udostępnianie. Zakup wersji elektronicznej bardziej wiąże bibliotekę z wydawcą, gdyż rezygnacja z licencji przez bibliotekę powoduje utratę

dostępu do całości czasopisma, nie tylko do przyszłych roczników, ale także do tych części, do których dostęp wcześniej był zapewniony.

- Wydawcy wykorzystują możliwości nowej technologii cyfrowej, w tym możliwości łatwego tworzenia odnośników między obiektami cyfrowymi. W tym kontekście nie dziwi, że wszystkie poważne serwisy wydawców tworzone są obecnie w formie indeksów cytowań.

Na rynku wydawniczym oprócz wydawców komercyjnych nadal funkcjonują wydawnictwa uczelni wyższych oraz towarzystw naukowych, traktujące publikowanie jako działalność niedochodową, a często nawet dotowaną. Jak wynika z moich wcześniejszych badań, w Europie Wschodniej wciąż obejmują one dość duży obszar rynku wydawniczego [Nahotko Małgorzata, Nahotko Marek 2009, s. 109-110], na Zachodzie, oprócz największych uniwersytetów, wydawnictwa uczelniane i towarzystw wykupione zostały przez wydawców komercyjnych. Wydawnictwa niekomercyjne zapewniają często hybrydowy dostęp do publikacji naukowych, traktując ich wersje elektroniczne jako uzupełnienie swojej tradycyjnej oferty. Wydawnictwa te nigdy nie traktowały swojej działalności dochodowo, starano się jedynie, aby wydatki nie przekroczyły przychodów, często uzupełnianych dotacjami. Niestety, aby utrzymać się na rynku, również mniejsi wydawcy muszą przejmować zachowania komercyjnych gigantów [Taylor, Perakakis, Trachana 2008, s. 18].

Obecnie, dzięki funkcjonowaniu GBC, możliwy staje się powrót do bardziej rozproszonego modelu publikowania naukowego, w którym znaczną rolę mogą odegrać małe, lokalne wydawnictwa, a nawet nieformalne grupy uczonych, dzięki technologiom cyfrowym uzyskujące globalny zasięg. Model ten, nazywany publikowaniem gildii uczonych⁹, oparty jest na sponsorowaniu publikacji przez instytucje naukowe [Kling, Spector, McKim 2002]. Publikowanie przez gildię jest powodowane przez chęć jak najszybszego udostępnienia publikowanych materiałów. Przynależność do gildii wysokiej klasy uczonych stanowi gwarancję jakości publikacji.

Wydawcy niekomercyjni, będący najczęściej gildiami pracowników nauki, działają w kierunku rozwoju dyscyplin, z którymi są związani, poprzez organizowanie kanałów komunikacyjnych i innych usług minimalizując koszty. Działając na granicy bezpieczeństwa ekonomicznego, bardzo niechętnie przyjmują jakiegokolwiek inicjatywy mogące je obniżyć. Do tej grupy wydawców zaliczyć można także tradycyjne biblioteki naukowe, digitalizujące swoje zbiory i udostępniające je w Internecie, na przykład w formie bibliotek cyfrowych. Globalna biblioteka cyfrowa może stanowić dla nich szansę na obniżenie kosztów dystrybucji lub znalezienia dla siebie nowego miejsca w elektronicznej komunikacji naukowej. Wydaje się, że współpraca pomiędzy tymi dwoma rodzajami instytucji (wydawcy niekomercyjni i biblioteki) może dać ciekawe rezultaty. Biblioteki i lokalne wydawnictwa uczelniane mogą współpracować w zakresie publikowania naukowego, każde realizując odpowiednie etapy tego procesu.

Oprócz wydawnictw, na rynku funkcjonują także pośrednicy, tak zwani agregatorzy, którzy oferują bibliotekom dostęp do publikacji wielu wydawców, a także do bibliograficznych baz danych. Obecnie publikacje elektroniczne bywają dostępne ze strony wydawcy oraz jednego lub kilku agregatorów. Ci ostatni oferują dostęp do zróżnicowanych zasobów przy pomocy jednolitego interfejsu, a także dostarczają zróżnicowanych usług, stanowiących ich wartość dodaną. W wyniku wspomnianej już konsolidacji rynku coraz częściej zdarza się, że największe wydawnictwa wykupują także firmy agregatorów.

⁹ Gildia uczonych jest tu rozumiana jako grupa osób o podobnych zainteresowaniach naukowych i wspólnie pracujących w ich zakresie. Grupy takie funkcjonują w oparciu o instytucje nauki. Gildia tego rodzaju publikuje własne materiały na użytek specjalistów. Model ten więc jest powrotem do koncepcji organizacji badań i publikowania przez uczonych dla uczonych, bez potrzeby pośrednictwa wydawców komercyjnych [Kling 2003].

5.2. Zagadnienia ekonomiczne

Ekonomiczne modele korzystania z globalnej biblioteki cyfrowej wciąż są na etapie rozwoju. Nie ma jednego sposobu na ponoszenie kosztów za dostarczanie informacji w sieci globalnej. Jak już pisałem, zagorzały zwolennik Open Access Stevan Harnad uważa, że dalszy rozwój nauki zależy od tego, czy potrafimy udostępnić recenzowane publikacje z czasopism naukowych w trybie online i bezpłatnie dla wszystkich. Wydawcy komercyjni natomiast przekonują, że bezpłatny dostęp do informacji naukowej nie jest korzystny, również dla samych pracowników nauki. Wiele informacji może być udostępnianych w Internecie bezpłatnie, co nie oznacza, że informacja wartościowa, o wysokiej jakości powstaje bez zwiększonych wysiłków i nakładów. Bezpłatna globalna biblioteka cyfrowa oznacza bezpłatne (dla użytkowników) serwisy na różnych poziomach. Czy możliwe jest zapewnienie bezpłatnego dostępu do wszystkich serwisów? Czy należy dążyć do takiego modelu, a jeżeli tak, to kto będzie za to płacił, gdyż jak wiadomo nie ma towarów ani usług darmowych. Jeżeli nie płaci bezpośredni klient (konsument, użytkownik), to znaczy, że zapłacił ktoś inny.

Ekonomiczne uzasadnienie działalności informacyjnej nie jest łatwe, bo nie wynika z prostego rachunku zysków i strat. Ekonomia skali w zakresie publikowania i działalności bibliotek utrudnia ukształtowanie odpowiednich cen; na przykład ograniczenia w prenumeracie prowadzą do wzrostu cen czasopism, związanych z tym kolejnych ograniczeń prenumeraty i w efekcie powstania destabilizującej spirali cięć wydatków i podwyżek cen. Wiele serwisów GBC, w tym biblioteki naukowe, rzadko może stosować ceny odzwierciedlające koszty, częściowo ze względu na to, że trudno jest ustalić właściwe ceny w sytuacji znacznego zróżnicowania możliwości ekonomicznych klientów (studenci, dysponujący skromnymi środkami na utrzymanie – uczeni dysponujący okazałymi grantami badawczymi). W ostatnich latach przemysł wydawniczy bardzo znacznie zwiększył skalę zysków (zob. p. 1.2), a publikowanie elektroniczne, po dokonaniu niezbędnych inwestycji, czyni tę działalność jeszcze bardziej zyskową, gdyż niezbędne jest ponoszenie jedynie kosztów pierwszego egzemplarza. Istniały nadzieje, że po rozpowszechnieniu się dystrybucji informacji w formie cyfrowej stanie się ona tańsza, jednak nic takiego jak dotąd nie nastąpiło [Shieber 2009].

Twórcy wielu serwisów GBC byli początkowo zaskakiwani globalnym zasięgiem ich usług, co wpływało także na kwestie finansowe. Tak było z niemieckim serwisem Subito, pozwalającym na zamawianie elektronicznych kopii artykułów z czasopism prenumerowanych przez biblioteki niemieckie. Początkowo można było zamówić tę usługę z dowolnego miejsca w Internecie; później ograniczono dostęp do użytkowników w Niemczech; w końcu zredukowano go do pracowników uczelni. Podobnie po stwierdzeniu, że tylko 15% użytkowników brytyjskiego serwisu BUBL pochodzi z Wlk. Brytanii, zaczęto zastanawiać się nad słusnością jego finansowania przez brytyjskie uniwersytety. Okazało się, że za globalnym zasięgiem informacji nie nadąża lokalna mentalność i dawne modele finansowania, sprawdzające się w przypadku biblioteki obsługującej społeczność lokalną.

Serwisy składające się na globalną bibliotekę cyfrową wymagają nowych modeli finansowania. Michael Lesk wymienia takie możliwości, jak wsparcie instytucjonalne, obciążanie użytkowników, zarabianie na reklamach, a nawet prośby o wsparcie [Lesk 2005, s. 262-263]. Wszystkie one nie wydają się jednak stwarzać solidnej podstawy ekonomicznej do prowadzenia profesjonalnej działalności informacyjnej, nie przewidują także nowych sposobów finansowania, poza wcześniej już stosowanymi w bibliotekach.

Nie tylko biblioteki naukowe dostarczają i zamierzają dostarczać bezpłatną informację cyfrową. Pod uwagę należy także brać, przynajmniej częściowo, usługi wydawców, księgarzy, ośrodki informacji i prywatne serwisy internetowe. Dotychczas biblioteki dostarczały informacje bezpłatnie, rozkładając koszty na całą społeczność potencjalnych użyt-

kowników. Powstaje pytanie, czy taki model może nadal funkcjonować w środowisku cyfrowym.

W świecie cyfrowego biznesu istnieje wiele sposobów na pobieranie opłat za udostępnianie materiałów cyfrowych. Wymienić można opłaty za czas dostępu, obciążenie CPU¹⁰, opłata za wyszukiwanie, za wyszukaną jednostkę, za kopiowanie (*download*). Stosowane są także roczne opłaty licencyjne, czy też opłaty kalkulowane z uwzględnieniem wielkości populacji potencjalnych użytkowników. Wszystkie stwarzają jakiegoś rodzaju problemy: użytkownicy mogą mieć kłopoty z ich zrozumieniem; wysokość opłat może być nieprzewidywalna; mogą one prowadzić do bezsensownych zachowań. Dostawcy starają się obciążać opłatami za coraz mniejsze jednostki usług (mikropłatności), co minimalizuje ryzyko utraty korzyści w przypadku wykrycia słabego punktu w metodzie pobierania opłat, powodującego dostęp do szerokiego zakresu usług za opłatę mniejszą, niż oczekiwana przez dostawcę. Z drugiej strony pobieranie wielu małych kwot za drobne transakcje zwiększa koszty administracyjne.

Mikropłatności związane są z nowymi możliwościami technologicznymi. Theodor Nelson, współtwórca projektu Xanadu zaproponował termin transkluzja¹¹ (*transclusion*) na określenie procesu polegającego na umieszczaniu cytowań obcych dokumentów. W przyszłości, zamiast cytować urywki tekstów, czy też wklejać grafikę odnalezioną w Sieci na własną stronę, autorzy mogą umieszczać do nich odnośniki. Po wykorzystaniu tego odnośnika (kliknięciu) przez użytkownika strony Web naliczana będzie opłata. Oczywiście opłaty powinny być bardzo niskie [Nelson, Smith 2007]. Najczęściej jednak wszelkie formy płatności za wykorzystanie (*pay-per-use*) powodują spadek poziomu korzystania ze źródeł. Wielu uczonych, postawionych w obliczu jakichkolwiek płatności, zaprzestanie korzystania z płatnych usług. Korzyści z mikropłatności są także trudne do oszacowania z góry.

Wydawcy komercyjni stosują następujące modele pobierania opłat:

- Miesięczne, kwartalne lub roczne opłaty za prenumeratę. Często zdarza się, że koszt prenumeraty dwóch wersji: tradycyjnej i elektronicznej jest znacznie niższy, niż suma kosztów obu wersji. Ten model, wywodzący się z publikacji tradycyjnych jest wciąż bardzo popularny, szczególnie wykorzystywany przez biblioteki tradycyjne pragnące zapewnić swoim użytkownikom dostęp do wersji elektronicznych czasopism.
- Opłaty za czas korzystania (na przykład za każdą minutę). Tego typu metoda naliczania opłat jest atrakcyjna dla użytkownika, jednak pod warunkiem, że płatności są minimalne. Zamiast więc nastawiać się na obsługę kilku użytkowników instytucjonalnych, mogących płacić 100 USD za godzinę wyszukiwania w serwisie Dialog, być może należy wziąć pod uwagę potrzeby milionów użytkowników skłonnych zapłacić 0,1 USD za godzinę korzystania.
- Nagrody za przyprowadzenie nowego użytkownika. Przez pewien czas serwisy online płaciły wydawcom, dzięki którym zdobywały nowych użytkowników. Podobnie Amazon płacił za skierowanie do księgarni nowego klienta.
- Opłaty za kopiowanie (*download*). Za każde ściągnięcie tekstu, na przykład artykułu z czasopisma na ekran komputera, na dysk lub drukarkę można pobierać opłaty. Tego typu usługi oferuje większość wydawców komercyjnych.
- Reklamy. Wiele stron Web usiłuje funkcjonować dzięki reklamom, co było łatwiejsze przed załamaniem się rynku firm internetowych. Obecnie rynek reklam jest skoncentrowany w niewielkiej liczbie portali i wyszukiwarek internetowych.

¹⁰ CPU – central processing unit, główny procesor komputera.

¹¹ Transkluzja definiowana jest przez Theodora Nelsona jako ta sama treść znana z więcej niż jednego miejsca. W zastosowaniu do stron Web każda zmiana na stronie powoduje utworzenie połączenia między tymi samymi (niezmienionymi) częściami strony w obu wersjach; przejście między częściami w obu wersjach dostępne jest poprzez kliknięcie. Dzięki temu możliwy jest powrót do pierwotnego kontekstu (otoczenia) wybranej części strony Web.

- Oplaty od autorów. Część wydawnictw pobiera opłaty od autorów za umieszczenie ich tekstów dostępnych online. Zwykle teksty te dla użytkowników dostępne są bezpłatnie.
- Strona Web jako infolinia. Stosowana w przypadkach, gdy potrzebny jest kontakt z klientem, ale nie musi on być w pełni symultaniczny. Dzięki uruchomieniu specjalnej strony Web można uniknąć poważnych kosztów utrzymywania stanowiska pomocy telefonicznej. Model ten rzadziej będzie miał zastosowanie w zakresie publikowania, ale stosowany bywa na przykład przez biblioteki.

Różne sposoby finansowania publikacji OA przedstawione zostały w jednej z moich poprzednich książek [Nahotko 2007b, s. 98-104]. Tu również funkcjonuje wiele modeli ekonomicznych publikowania, gdyż „Open Access” nie oznacza „bez żadnych kosztów”. Bardziej więc chodzi o to, kto i w jaki sposób ma uczestniczyć w kosztach; jasne jest tylko, że użytkownik końcowy zwolniony jest z płatności¹². Najogólniej można powiedzieć, że istnieją dwie grupy modeli. Pierwszy z nich, zwany opłatą za przygotowanie artykułu, bywa także nazywany płatnościami autorskimi, gdyż koszty przygotowania tekstu do publikacji i samej publikacji pokrywają autorzy, a dokładniej zatrudniające ich instytucje naukowe lub organizacje przyznające granty na badania naukowe. Obecnie już setki czasopism (również całkowicie komercyjnych) realizuje tego typu model, czasem dla całości publikacji, czasem jako opcję do wyboru przez autora. Zaletami tego modelu jest jego prostota oraz to, że zwraca on autorowi uwagę na koszty publikowania, co zostało całkowicie zagubione w systemie opartym na prenumeracie.

Innym, często stosowanym modelem, jest bezpośrednie finansowanie publikacji przez towarzystwa i instytucje naukowe, fundacje, inne organizacje naukowe oraz jednostki administracji centralnej w ramach działalności statutowej (misji) tych organizacji. W takim przypadku publikacja jest bezpłatna zarówno z punktu widzenia czytelnika, jak i autora. Czasami takie czasopisma uznaje się za subsydiowane, w odróżnieniu od czasopism komercyjnych, co jest nie do końca usprawiedliwione. Dochody wydawców komercyjnych pochodzą bowiem z prenumeraty, czyli z budżetu bibliotek naukowych, więc również są w jakimś sensie „subsidiowane”, tylko nie tak bezpośrednio. Odmianą tego modelu jest coraz częściej dyskutowane centralne finansowanie publikacji czasopism; fundusze byłyby przekazywane bezpośrednio wydawcom komercyjnym przez centralną instytucję finansującą. Tego typu model realizowany jest w ramach konsorcjum SCOAP (CERN) dla czasopism z dziedziny fizyki [Bianco i in. 2007].

Istnieją także inne możliwości publikowania OA, jak chociażby poprzez wspomniane już umieszczanie materiałów na stronach domowych uczonych, czy też w repozytoriach. Takie publikowanie może być jednak traktowane jedynie uzupełniająco, gdyż udostępniane teksty zazwyczaj są jednocześnie publikowane w sposób tradycyjny.

Publikowanie naukowe jest usługą, w większości przypadków finansowaną przez społeczność naukową w szerokim rozumieniu [Cockerill 2006, s. 113]. Finansowanie zysków wydawców, w połączeniu ze świadomością o możliwościach, oferowanych przez nowe technologie, oraz zwyczaj pozyskiwania usług autorów, recenzentów i redaktorów bez wynagrodzenia spowodowały krytykę postępowania wydawców komercyjnych przez środowiska naukowe. Przedstawicielem tego nurtu jest Jeremy Rifkin, który krytykuje komercjalizację kultury i udostępniania informacji. Według niego możliwości sieci rozległych, zamiast do maksymalizacji poziomu dostępu do informacji i treści, są wykorzystywane przez wydawców komercyjnych do pełnej kontroli dostępu i jego ograniczania tylko dla tych, którzy mogą zapłacić. Sam dostęp również został ograniczony przez zastąpienie prawa własności prawem do dostępu [Rifkin 2003]. Taka sytuacja powoduje zwrot ku OA

¹² Z punktu widzenia użytkownika, usługi tradycyjnej biblioteki także zawsze były „Open Access”, gdyż użytkownik nie płacił za nie bezpośrednio; do współczesnego OA brakowało tylko udostępniania treści w Internecie.

[Święćkowska 2008, s. 51]. Istotnym elementem walki jego zwolenników jest żądanie wprowadzenia obligatoryjności udostępniania w trybie OA publikacji, które są wynikiem badań naukowych, finansowanych z środków publicznych.

Richard Poynder przedstawił dyskusję dotyczącą kosztów publikowania elektronicznego, szczególnie w kontekście oczekiwań redukcji tych kosztów dzięki publikowaniu OA. Wynika z niej, że zastąpienie jednego kosztownego modelu publikowania innym, równie kosztownym, nie rozwiąże problemów, a tylko pogłębi kryzys komunikacji naukowej [Poynder 2008a]. Tego typu dyskusje świadczą o tym, że rozwiązania społeczne, głównie ekonomiczne, wciąż pozostają w tyle za możliwościami technicznymi. Globalna biblioteka cyfrowa jest bardzo odpowiednim miejscem na dystrybucję bezpłatnej (z punktu widzenia użytkownika końcowego) informacji naukowej w formie produktów, oferowanych przez jej serwisy. Wciąż jednak nie do końca rozwiązany pozostaje sposób docierania z właściwymi płatnościami do właściwych miejsc w cyklu publikowania. Zmiany w tym zakresie mogą doprowadzić do redystrybucji kosztów i dochodów, powodującej głębokie zmiany na rynku elektronicznych publikacji naukowych, na przykład wypadanie z tego rynku pewnych grup instytucji (jak wydawnictwa) i gruntowne zmiany zakresu działalności pozostałych (w tym bibliotek naukowych) [Feather 2003, s. 204].

5.3. Zagadnienia prawne

Jak pisze Andrzej Matlak, udostępnianie dzieł naukowych w postaci cyfrowej, oprócz korzyści, niesie także obawy wystąpienia zjawisk negatywnych, do których autor zaliczył przede wszystkim łatwość naruszenia praw autorskich na niespotykaną wcześniej skalę [Matlak 2001, s. 162]. Zagrożenia mogą wynikać z korzystania z obiektów cyfrowych bez zgody właściciela praw lub manipulowania treścią utworów rozpowszechnianych w formie cyfrowej. Sytuację komplikuje brak typowych regulacji prawnych dla Internetu jako całości, co powoduje, że w wielu przypadkach stosowane są normy zwyczajowe, których przykładem może być prawo do swobodnego tworzenia sieci odnośników [Polański 2006, s. 181]; prawa zwyczajowe ograniczane są często przez unormowania prawne – na przykład twórca odnośnika może jednak zostać pociągnięty do odpowiedzialności, gdy poprowadzi hiperlink do strony, zawierającej treści zakazane prawem [Barta, Markiewicz 2008, s. 246]. Podobnie niejasna jest sytuacja prawna tzw. „głębokiego linkowania”¹³ [Oppenheim 2008, s. 946].

Revolucja technologiczna związana z cyfryzacją spowodowała zburzenie wcześniejszego ładu prawnego, przyczyniając się do powstania dwóch, przeciwstawnych postaw: pierwsza związana jest z wykorzystaniem nowych możliwości do zaostrzenia kontroli użycia i wykluczenia, druga postuluje zastosowanie tych samych technologii do umożliwienia jak najszerszego udostępnienia wiedzy [Hofmokr 2009, s. 152; Reichman, Uhlir 2003, s. 318-319]. Oba stanowiska mają swoje konsekwencje prawne.

Wszystkie instytucje, tworzące serwisy globalnej biblioteki cyfrowej, muszą przestrzegać regulacji prawnych, wśród których na pierwszym miejscu wymienić należy prawo własności intelektualnej. W toczonej dyskusjach padają różne propozycje co do tego, kto powinien posiadać prawa do jakich treści, przez jaki czas i za jakie kwoty (zob. np. [Howorka 2009]). Termin „własność intelektualna” jest hierarchicznie nadrzędny w stosunku do trzech węższych terminów: prawo autorskie (*copyright*), prawo patentowe i znaki własnościowe. W zakresie publikowania naukowego, szczególnie elektronicznego, największe zastosowanie ma prawo autorskie.

Janusz Barta i Ryszard Markiewicz jako jeden z istotnych problemów prawnych, powodowanych przez technikę informacyjną, wymienili zmianę dotychczasowych para-

¹³ „Głębokie linki” to odnośniki pozwalające na bezpośrednie dotarcie do informacji lub obiektu znajdującego się na podstronie, z pominięciem strony głównej.

dygmatów działalności wydawniczej, związaną z udostępnianiem utworów w sieciach komputerowych [Barta, Markiewicz 2008, s. 231]. Wynikają one z dwóch przyczyn: po pierwsze, z zacieraania różnic pomiędzy publikowaniem a rozpowszechnianiem, po drugie, z przejmowania kompetencji wydawców przez twórców i instytucje informacyjne, w tym biblioteki. Ze względu na kwestionowanie cech charakterystycznych publikacji elektronicznej, takich jak: charakter publiczny, niemożność definitywnego usunięcia ze zbiorów oraz istnienie utrwalonego egzemplarza, niezbędne może być utworzenie nowej, prawnej definicji utworu opublikowanego. To nie jedyny przykład trudności natury prawnej, wynikających z nieprzystawania wcześniejszych rozwiązań do nowej sytuacji. Dla materiałów tradycyjnych łatwo można było na przykład odróżnić wykorzystanie (przeczytanie) od kopiowania. Każda kolejna kopia zazwyczaj była coraz gorszej jakości. Dla zasobów cyfrowych znacznie trudniej o takie rozróżnienie; zanim użytkownik zobaczy żądany dokument na ekranie, w celu jego użycia, jest on nawet kilkakrotnie kopiowany, a jakość każdej kopii jest dokładnie identyczna z oryginałem [Delany, Landow 2008, s. 82].

Zagadnienia prawne dotyczą wszystkich etapów publikowania cyfrowego: tworzenia, przechowywania i rozpowszechniania zasobów informacji cyfrowej. Stąd znaczne uwrażliwienie procesów publikowania cyfrowego na rozwiązania prawne. Nieograniczony rozwój nauki może być zagrożony przez stałe rozszerzanie zasięgu praw autorskich i patentowych. Prawo to stało się zbyt restrykcyjne. Dodatkową presję wywierają coraz bliższe powiązania pomiędzy środowiskiem nauki a przemysłem i biznesem, który w imię ochrony własnych interesów utrudnia czy wręcz uniemożliwia powszechny dostęp do wyników prac badawczych. W efekcie tych poczynań ochroną prawną obejmowane są takie obiekty, jak na przykład surowe dane z badań, w tym poszczególne sekwencje genu ludzkiego.

Prawo autorskie ma dwa zasadnicze cele:

- Zapewnienie autorowi odpowiednich zysków z jego twórczej działalności intelektualnej; zyski mogą być pojmowane nie tylko w sensie bezpośrednio finansowym, ale także dotyczą korzyści z udziału w rozwoju wiedzy oraz ze wzrostu reputacji.
- Zapewnienie właścicielowi praw autorskich kontroli nad sposobem wykorzystania dzieła, objętego tym prawem.

Zgodnie z prawem międzynarodowym i postanowieniami większości uregulowań krajowych (w tym polskich) wyróżnić można dwie grupy praw autorskich:

- Autorskie prawo osobiste, niezbywalne (nie można go odsprzedać) i będące prawem do autorstwa utworu.
- Majątkowe prawo autorskie, do rozporządzania swoim utworem i otrzymywania wynagrodzenia za jego wykorzystywanie; jest ono zbywalne i zgodnie z wolą autora może być przeniesione na inne osoby. W dalszej części rozdziału zajmę się tylko tą grupą praw, gdyż publikowanie elektroniczne nie powoduje zazwyczaj zmiany autorstwa¹⁴.
- Podstawowe rozwiązania prawa autorskiego są podobne we wszystkich krajach.
- Brak potrzeby rejestracji – dzieło po utworzeniu automatycznie staje się przedmiotem prawa autorskiego.
- Większość krajów ma własne przepisy prawa autorskiego i często urzędy stojące na straży jego przestrzegania. Prawa autorskie w poszczególnych krajach dostosowane zostały do potrzeb ochrony praw do publikacji elektronicznych i dostępu do nich.
- Prawa autorskie wygasają po pewnym okresie od daty publikacji lub śmierci autora (ostatniego z wielu autorów); zazwyczaj jest to 70 lat.

¹⁴ Zdarzają się publikacje, dla których ustalenie autorstwa jest trudne (np. teksty tworzone przez wiele, czasem przypadkowych osób korzystających z sieci i dopisujących kolejne części fabuły). Podobne przypadki zdarzają się także w zakresie tekstów naukowych, na przykład w sytuacji, gdy tekst powstał w wyniku pracy zespołu dziesiątek osób, a sygnowany jest nazwiskiem osoby pełniącej funkcję kierownika projektu, która mogła mieć niewielki wpływ na zawartość tego właśnie tekstu.

- Dzieło może posiadać więcej niż jednego właściciela praw autorskich (współautorzy¹⁵).

Twórcy serwisów globalnej biblioteki cyfrowej muszą uwzględniać wymogi prawa autorskiego odnośnie zasobów informacji o różnej proveniencji, zarówno tworzonych lokalnie (digitalizowanych oraz *born digital*), jak i otrzymywanych z innych źródeł. Serwisy te dbają również o ochronę prawną udostępnianych materiałów, w zakresie ich właściwego użytkowania w sensie prawnym.

Twórcy serwisów GBC mogą spotkać się z kilkoma sytuacjami prawnymi:

- Materiały nie są objęte prawem autorskim.
- Materiały są objęte prawem autorskim i znany jest podmiot tych praw (właściciel).
- Materiały są objęte prawem autorskim, ale właściciel jest trudny do odszukania lub jest nieznany (dzieła osierocone).

Postępowanie twórców serwisu GBC przebiega w zależności od stanu prawnego dzieła; czasem wymusza on decyzję o zaniechaniu digitalizacji i nieumieszczeniu w zasobach serwisu. Wówczas regulacje prawa autorskiego mogą działać na niekorzyść autora; dzieje się tak w przypadku dzieł osieroconych, oraz tych, dla których właściciel praw (wydawnictwo) nie zamierza z nich korzystać, tworząc reedycję dzieła. Wówczas dzieło nie może być publikowane ani drukiem ani w postaci cyfrowej, pomimo że nie działa to na niczyją szkodę (właściciel praw i tak z nich nie korzysta), a nie przynosi korzyści w postaci zwiększonej dostępności.

Bardzo często właścicielem majątkowych praw autorskich stają się wydawnictwa komercyjne, które zawierają standardowe umowy z autorami, przewidujące zrzekanie się praw (majątkowych). Taki sposób postępowania krytykowany jest głównie z tego powodu, że autor przestaje dysponować swoim utworem; nie może go na przykład zdeponować w archiwum cyfrowym (repozytorium), co także jest formą upublicznienia dzieła. Rozwiązaniem może być model tzw. licencji na publikowanie (*licence to publish*). Licencja taka pozwala wydawcy na jednorazowe opublikowanie artykułu w czasopiśmie bez przejmowania na stałe praw autorskich do tekstu. Jest ona wystarczająca dla zapewnienia wydawcy odpowiedniej pozycji na rynku, dzięki temu, że uzyskuje on zapewnienie pierwszeństwa publikacji.

Polityka wydawnictw często wykracza poza ochronę praw autora do osiągnięcia korzyści z twórczego i intelektualnego aktu opublikowania oryginalnego dzieła [Willinsky 2006, s. 47]. Prawo autorskie jest zbyt często wykorzystywane dla ochrony praw wydawcy do pobierania opłat za publikowanie, czyli za blokowanie dostępu do dobra wspólnego w rozumieniu Roberta Mertona (zob. p. 1.2). Prawa autorskie przez część wydawców wykorzystywane są w celu zapewnienia utrzymania zysków w trudnym okresie transferu od druku do mediów cyfrowych.

Wyróżnić można cztery sposoby postępowania, stosowane przez właścicieli praw autorskich w odpowiedzi na rozwój publikacji cyfrowych. Po pierwsze, lobbują oni za zwiększeniem restrykcyjności prawa autorskiego, czasem znajdując popleczników, na przykład w przemyśle muzycznym. Rezultatem jest m.in. ograniczanie dozwolonego użytku. Po drugie rozwijane są techniczne sposoby ograniczania dostępu, zwane Digital Rights Management Systems (DRM), ograniczające prawo do czytania i kopiowania¹⁶. Po trzecie tworzone są umowy licencyjne, szczegółowo opisujące sposoby korzystania

¹⁵ Jak piszą Janusz Barta i Ryszard Markiewicz, sytuacja w zakresie współautorstwa dzieł naukowych jest specyficzna, ze względu na to, że prawo autorskie nie dotyczy wartości odkrywczej dzieła, co jest cechą zasadniczą z punktu widzenia naukowości, a także oznaczania współautorstwa dzieł naukowych. Zazwyczaj tylko twórca tekstu naukowego jest uważany za jego autora; nie ma jednak przeszkód prawnych do oznaczania jako autorów również osób współuczestniczących w uzyskiwaniu wyniku naukowego [Barta, Markiewicz 2008, s. 57].

¹⁶ Ograniczenia te powodują, że kontrola kopiowania, zawsze funkcjonująca w prawie autorskim, obecnie przybiera formę ścisłej kontroli użycia utworu za pośrednictwem komputera [Benkler 2008, s. 456].

z licencjonowanych treści [Oppenheim 2008, s. 940]. Na koniec stosowana jest perswazja, że jakiekolwiek odstępstwa od ochrony własności intelektualnej są niemoralne i muszą być traktowane jako przestępstwo [Bard, Söderqvist 2006, s. 69]. Rezultatem tych poczynań mogą być zaskakujące sytuacje, wynikające z wykorzystywania braku przystawania rozwiązań prawnych do nowych okoliczności, na przykład wydawcy dostarczają bibliotekom jako egzemplarz obowiązkowy dokument w formie elektronicznej (plik), którego biblioteki nie mogą legalnie udostępnić ani w tej formie, ani w formie wydruków. Rozwiązania prawne, stosowane w zakresie publikacji tradycyjnych, nie zdają już egzaminu, więc dla zachowania swoich dotychczasowych zysków właściciele praw autorskich do publikacji cyfrowych żądają znacznego wzrostu restrykcyjności rozwiązań¹⁷. Zjawiska te potwierdzają tezę Ithiel de Sola Pool, który już w latach osiemdziesiątych XX wieku twierdził, że ponieważ formy ochrony dostosowane do dokumentów drukowanych są nieużyteczne w nowych warunkach, powstają nowe rozwiązania, które są efektem walki pomiędzy zwolennikami wolności wypowiedzi a zwolennikami restrykcji nie spotykanych w publikowaniu tradycyjnym [Pool 1983].

Na szkodliwość stosowania zbyt restrykcyjnych ograniczeń prawnych w zakresie ochrony własności intelektualnej wskazuje Yochai Benkler, tłumacząc ją specyficznymi cechami informacji jako dobra: niekonkurencyjnością informacji naukowej i efektem „stawania na ramionach olbrzymów”; dzisiejsze odkrycie naukowe jest oparte na wielu poprzednio dokonanych odkryciach i daje podstawy do następnych [Benkler 2008, s. 52-54]. Jeżeli dostęp do informacji o postępie nauki będzie utrudniany przez wprowadzanie opłat przekraczających koszt krańcowy (wynoszący zero), to nastąpi spadek twórczości i innowacyjności w najbliższej przyszłości. Wysoka cena nie tylko ogranicza bieżącą konsumpcję informacji, ale również szybko zmniejsza jej dalszą produkcję.

John Willinsky wskazuje dwa wyjątki w ogólnej sytuacji wzrostu kontroli, której narzędziem jest prawo autorskie. Pierwszy dotyczy uczelni w krajach rozwijających się, którym wydawnictwa komercyjne udostępniają swoje czasopisma w trybie Open Access. Innym ważnym osiągnięciem jest coraz powszechniejsza zgoda wydawców, zawierana w umowach wydawniczych, na autoarchiwizację dzieła przez autora. Już około 95% wydawnictw udziela takiej zgody; może ona dotyczyć wersji przed lub po opublikowaniu (preprint lub postprint) [Willinsky 2006, s. 48].

Po wygaśnięciu praw autorskich, dzieło przechodzi do tak zwanej domeny publicznej, którą stanowi abstrakcyjny ogół utworów, dla których wygasły prawa autorskie lub autorzy zrzekli się ich. Do domeny publicznej część materiałów może trafiać także z mocy prawa; dotyczy to publikacji, będących efektem badań finansowanych przez rząd federalny w USA [Reichmann, Uhlir 2003, s. 326]. Domena publiczna stale się kurczy z powodu powiększania zakresu materiałów, dla których stosowane jest prawo autorskie oraz wydłużania czasu trwania ograniczeń¹⁸. Stosowane są także prawne sztuczki, jak na przykład ochrona równoległa¹⁹. Ograniczanie domeny publicznej jest bardzo ważne dla

¹⁷ Takim rozwiązaniem jest amerykański Digital Millennium Copyright Act, zakazujący właścicielom urządzeń i oprogramowania ich modyfikacji, a nawet analizy ich działania, o ile mogłoby to prowadzić do obejścia wbudowanych systemów DRM.

¹⁸ Jak podaje Christine Borgman, w USA czas trwania copyright wynosił początkowo 14 lat; później został przedłużony do 28 lat, a obecnie wynosi 70 lat od śmierci autora i wciąż pojawiają się propozycje jego przedłużania. Europejskie prawo jest jeszcze bardziej restrykcyjne, m.in. z powodu większej grupy dokumentów objętych przez copyright [Borgman 2007, s. 107]. W Polsce od 1952 r. okres ochrony praw wynosił 20 lat od śmierci autora. Został on przedłużony w 1994 r. do 50 lat, a w 2000 r. do 70 lat. Mówi się obecnie o potrzebie jego przedłużenia do 100 lat [Szczepańska 2007a, s. 53]. Żywa jest obecnie dyskusja w UE o potrzebie i możliwości przedłużenia ochrony własności intelektualnej do nagrań dźwiękowych z 50 do 95 lat (propozycja zmiany dyrektywy UE 2006/116/EC).

¹⁹ Ochrona równoległa polega na równoczesnej ochronie innowacji na podstawie wielu różnych przepisów, np. prawa autorskiego, patentowego, znaków towarowych i nieuczciwej konkurencji dla maksymalnego wydłużenia okresu ochrony.

twórców zasobów cyfrowych i serwisów globalnej biblioteki cyfrowej. Znacznie ponad 90% produkcji wydawniczej przestaje funkcjonować na rynku komercyjnym już po kilku latach od opublikowania, nie przynosząc żadnych dochodów ani autorowi, ani wydawcy. Mogłyby one zostać włączone do domeny publicznej i zdigitalizowane, gdyby wydawcy w obawie przed utratą dochodów ze sprzedaży kilku procent publikacji nie blokowali dostępu do wszystkich pozostałych na długie dziesiątki lat [Lessig 2005, s. 249]. Problemy te stały się dobrze widoczne w związku z wielkimi projektami digitalizacji, powodującymi potrzebę negocjacji warunków realizacji copyrightu pomiędzy Google i organizacjami wydawców²⁰.

Czas niezbędny do określenia właściciela praw do digitalizowanych obiektów i ich zabezpieczenia może być długi. Badanie to może znacznie wydłużyć proces digitalizacji i zwiększyć jego koszty, czyniąc go wręcz nieopłacalnym. Wiele instytucji skupia się więc na digitalizacji materiałów, do których posiada prawa lub takich, do których prawa wygasły. Uniwersytety i ośrodki badawcze rozpoczęły napełnianie repozytoriów od materiałów, do których prawa autorskie są znane, bo należą albo do uczonych albo do instytucji, w których oni pracują (wyniki badań, raporty, dysertacje, sprawozdania i materiały konferencyjne) [Szczepańska 2007b, s. 201]. Jednocześnie twórcy archiwów cyfrowych zazwyczaj zapewniają, że prawa autorskie do utworów przekazywanych do repozytoriów przez ich autorów pozostają przy autorach. Na korzystanie ze swoich dzieł udzielają oni licencji niewyłącznej (utwór może być przekazany w tej samej postaci do innej bazy lub innemu podmiotowi bez pytania o zgodę właścicieli repozytorium). Udzielenie licencji na zwykle miejsce w trakcie tworzenia rekordów w bazie, jako jeden z elementów procesu autoarchiwizacji lub poprzez podpisanie i odesłanie umowy licencyjnej. Część organizatorów repozytoriów w ogóle nie wymaga żadnych deklaracji ze strony autorów. Na stronie projektu pojawia się jedynie zapis, że prawa autorskie pozostają przy autorach. W takim przypadku, przysyłając materiały do bazy, autor akceptuje warunki korzystania z repozytorium opisane na stronach projektu (udostępnianie materiałów w modelu Open Access bez żadnych ograniczeń i za darmo) i tym samym udziela dorozumianej zgody na ich przechowywanie i przetwarzanie. Twórcy niektórych repozytoriów szczegółowo opisują procesy, którym poddawane będą archiwizowane obiekty (przetwarzanie do różnych formatów, udostępnianie, drukowanie na żądanie, przechowywanie) uważając, że autor wyraża dorozumianą zgodę na takie działania.

Odrębnego omówienia wymagają tzw. dzieła osierocone (*orphan works*), czyli takie, dla których właściciel praw autorskich nie jest znany lub nie ma możliwości jego lokalizacji. Problemy z tego typu materiałami występują wszędzie tam, gdzie wojny dokonywały spustoszeń, zmieniały się granice państw, znikwały instytucje, w tym wydawnicze; takim krajem jest niewątpliwie Polska. Trudności związane z odnalezieniem właściciela praw autorskich powodują niejednokrotnie, że dzieło nie zostaje zdigitalizowane i opublikowane w nowej formie [Szczepańska 2007a, s. 60]. Możliwość przeciwdziałania temu zjawisku przedstawiła Agnieszka Vetulani; prace w tym zakresie trwają m.in. w ramach Unii Europejskiej, która skłania się do rozwiązywania zagadnień prawnych dzieł osieroconych na poziomie poszczególnych krajów członkowskich [Vetulani 2009, s. 30].

W globalnej bibliotece cyfrowej podejmowane są próby zachowania, z prawnego punktu widzenia, dzieł naukowych (i innych także) jako wspólnej własności, wspólnego dobra²¹, na wzór istniejących niegdyś wspólnie użytkowanych lasów czy pastwisk (zwanym

²⁰ Zob. <http://www.googlebooksettlement.com/help/bin/answer.py?answer=128055&hl=pl>.

²¹ Dobra wspólne były krytykowane przez Garretta Hardina w jego artykule [Hardin 1968], w którym dowodził, że wolność wspólnot prowadzi do ruiny, od której ratunkiem jest jedynie prywatyzacja dobra wspólnego lub jego nacjonalizacja (upaństwowienie). Dywagacje te dotyczyły w szczególności dóbr otwartych i jednocześnie rywalizacyjnych [Hofmokl 2009, s. 25], a więc nie zasobów wiedzy w Sieci, które uważa się za co prawda otwarte, ale też łatwo podzielne. Późniejsze badania, prowadzone m.in. przez Elinor Ostrom nie potwierdziły tych tez.

w Anglii *commons*). Chodzi o to, aby z jednej strony pewne prawa autora dzieła mogły być przez niego określone i wskazane, a z drugiej strony, aby prawa te nie uniemożliwiały swobodnego korzystania z dzieł w trybie Open Access, czyli działały na wzór wcześniejszej istniejących licencji *free source* – wolnego oprogramowania (głównie GNU GPL). W ten sposób powstał pomysł organizacji Creative Commons (CC), która stworzona została w 2002 r. przez amerykańskich prawników i uczonych (w tym głównie Lawrence'a Lessiga), zaniepokojonych procesami komercjalizacji cyfrowych dóbr nauki i kultury [Hofmokr, Tarkowski 2005]. Organizacja ta działa na rzecz rozszerzenia zasobów wspólnych, dostępnych w globalnej bibliotece cyfrowej, poprzez tworzenie narzędzi ułatwiających twórcom otwieranie dostępu do własnych dzieł. Narzędziami tymi są głównie nowego rodzaju licencje autorskie, stanowiące rozwiązanie pośrednie między bardzo restrykcyjnymi licencjami stosowanymi przez wydawnictwa komercyjne, odbierającymi autorom prawa majątkowe do utworu, a pełną domeną publiczną.

Częścią Creative Commons jest Science Commons, którego zadaniem jest pobudzanie rozwoju nauki przez ułatwienie dostępu do publikacji, danych i innych materiałów naukowych, chronionych prawami autorskimi, wszystkim potrzebującym (nauce, przemysłowi), co ma wpływać na lepsze rozprzestrzenianie wiedzy. Twórcy Science Commons starają się zaradzić nieprzystawalności współczesnego prawa do nowej rzeczywistości komunikacji naukowej, w której nowe technologie wykorzystywane są do szybkiej archiwizacji i udostępniania danych, a Internet funkcjonuje jako globalna biblioteka cyfrowa. Open Data Commons – Public Domain Dedication & License (PDDL) jest natomiast licencją pozwalającą na otwarty dostęp do danych, w szczególności gromadzonych w bazach danych, w tym bibliograficznych²².

Zagrożeniem dla swobodnego rozwoju nauki i komunikacji naukowej może być cenzura w Internecie. Oprócz walki ze zjawiskami o podłożu kryminalnym (terroryzm, pornografia, przemoc) cenzura może być także wykorzystywana do walki z poglądami politycznymi lub naukowymi, o czym świadczy przykład porozumienia Google z rządem Chin. Odróżnianie treści etycznie nieodpowiednich od cenzurowanych z powodów ograniczających wolność wypowiedzi zawsze było zadaniem bibliotekarzy, więc mogą oni pełnić podobną rolę dla zasobów GBC [Oppenheim 2008, s. 948].

W wyniku tworzenia coraz nowszych restrykcji prawnych pogłębia się rozdźwięk pomiędzy możliwościami, jakie daje zastosowanie technologii cyfrowych ułatwiających rozpowszechnianie i dostęp do treści naukowych a ograniczeniami narzucanymi przez prawa własności i politykę ekonomiczną w zakresie dostępu do tych samych zasobów. Wydaje się, że rozwiązaniem tego paradoksu może być bardziej powszechne stosowanie polityki Open Access wraz z licencjami typu CC. Pozwoli to na traktowanie informacji naukowej mniej jako dobra wspólnego, a bardziej – publicznego, co jest jedną z podstawowych zasad funkcjonowania IIN i globalnej biblioteki cyfrowej.

Niewątpliwie zagadnienia prawne stanowią wciąż nierozwiązany problem, utrudniający rozwój GBC. W szczególności prawa autorskie muszą być dostosowane do nowych możliwości, jakie daje środowisko cyfrowe. Wzrost kontroli i restrykcyjności prawa, osiągnięty dzięki większym możliwościom technologii cyfrowej, z pewnością możnaby uznać za porażkę idei globalnej biblioteki cyfrowej, pomimo tego, że dopuszcza ona komercyjne udostępnianie treści.

użytkownicy dóbr wspólnych zachowują się inaczej, gdy umożliwi się im łatwą komunikację, co pozwala na stworzenie im własnych zasad postępowania w przyszłości [Hess, Ostrom 2006, s. 11].

²² Zob. <https://biblios.net/pddl>.

6. MODEL – METAFORA KOMUNIKACJI NAUKOWEJ W GBC

Obecnie społeczność globalna stanęła w obliczu rozdźwięku pomiędzy niemal nieograniczonym zasięgiem sieci komputerowych a rzeczywistymi ograniczeniami możliwości ludzkiej percepcji. Powstaje pytanie o przyszłość komunikacji naukowej w środowisku, w którym niemal każdy może publikować swoje dzieła, a wielu realnie korzysta z tej możliwości. Wraz z oczywistą łatwością komunikowania się w Internecie ograniczeniu ulega znaczenie poprzednich, instytucjonalnych i hierarchicznych struktur w nauce. Sukces poglądów filozoficznych, głoszonych przez zwolenników „mądrości tłumów” [Surowiecki 2005], sugeruje zwrot ku egalitaryzmowi w nauce, w której uznane autorytety naukowe tracą znaczenie na rzecz uśrednionej opinii tłumu.

Bez względu na przyszłą realizację tego typu idei, z lektury poprzednich rozdziałów mojej książki wynika fakt odbywania się na naszych oczach bardzo głębokich zmian we wszystkich dziedzinach życia, których dotyka jakikolwiek rodzaj komunikacji (a więc właściwie we wszystkich dziedzinach życia), w tym w znacznym stopniu w nauce. Zmiany te mogą być porównywane jedynie do procesów mających miejsce w okresach wprowadzania do użytku rewolucyjnych, nowych technologii informacyjnych, takich jak pojawienie się języka naturalnego, wynalazek pisma i druku [Bard, Söderqvist 2006, s. 30-33]. Co paradoksalne, technologia elektroniczna umożliwia w jakimś sensie powrót do poprzednio zarzuconych, ze względu na wymogi technologiczne, sposobów komunikacji międzyludzkiej, w tym naukowej. Jej cechą charakterystyczną jest oralność – zarówno narzędzia, takie jak blogi i poczta elektroniczna, ale także przede wszystkim hipertekst, pomimo posługiwania się pismem, umożliwiają oderwanie komunikujących się stron od tradycyjnej liniowości analogowo powielanych tekstów, przez co więcej mają wspólnego z mową, niż z pismem. Rodzi to napięcia w świecie, w którym oficjalna nauka (sama będąc rezultatem rozwoju piśmienności) przez wieki wynosiła przekazy uwiecznione w piśmie ponad prymitywne kultury przekazów ustnych [Wright 2007, s. 232].

Na poziomie dynamiki organizacji komunikacji naukowej opisane zmiany manifestują się zróżnicowaniem na stałe i zmienne, „płynne” elementy jej środowiska. Symbolami elementów stałych są książki, systemy biblioteczne i tradycyjne SOW, traktowane jako statyczne instrumenty piśmienności. Elementami zmiennymi są poczta elektroniczna, komunikatory, blogi, czaty, nowe SOW (zob. [Sosińska-Kalata 2008, s. 112]) i inne instrumenty kojarzone z Webem 2.0. Elementy statyczne uporządkowane są hierarchicznie, natomiast zmienne – sieciowo. Hierarchie i sieci nie muszą jednak pozostawać w opozycji, mogą nie tylko koegzystować, ale nawet uzupełniać się wzajemnie. Efekty tego współistnienia w krótkiej perspektywie sprawiają wrażenie chaosu, jednak w dłuższym okresie mogą doprowadzić do innego, sprawniejszego funkcjonowania komunikacji naukowej, w której w pełni wykorzystane będą nowe możliwości technologii informacyjnych.

Dla zobrazowania zachodzących zmian w dalszej części rozdziału przedstawiłem model obecnego, przejściowego stanu rzeczy w komunikacji naukowej. Dzięki takiemu opisowi stanu aktualnego możliwe jest wyznaczenie prawdopodobnych kierunków rozwoju, chociaż o jakiegokolwiek prognozy we współczesnym, płynnym środowisku komunikacyjnym bardzo trudno.

Metody modelowania często są uzależnione od rodzaju modelowanego systemu. Najłatwiej jest oczywiście modelować dobrze określone systemy o znanej strukturze i funkcjach. Ich wektory wejścia i wyjścia, i/lub interfejsy są w pełni zdefiniowane i mierzalne. Więcej kłopotów sprawiają systemy o atrybutach (wielkościach charakterystycznych, opisujących system) zdefiniowanych, lecz nie zawsze mierzalnych. Najtrudniejszym przypadkiem są systemy, które nie są w pełni obiektywnie zdefiniowane; brak opisu ich podsystemów i atrybutów. Do tej grupy należą tzw. modele mentalne, powstające w oparciu o wcześniejsze doświadczenia i będące niejednokrotnie metaforami¹ pomocnymi w myśleniu indywidualnym lub zbiorowym oraz reprezentacji informacji i wiedzy [Próchnicka 2006, s. 41].

Większość znanych modeli komunikacji naukowej (a nawet komunikacji w ogóle) oparta jest na koncepcji przewodu. Modele te opierają się na metaforze „przepływu” informacji przez przewód lub rurę. Najbardziej znany jest model transmisji danych Shannona i Weavera z 1948 r., który opisany został po raz pierwszy przez Michaela Reddy [Chandler 2000]. Autor ten zauważył, że istnieje spójny, metaforyczny substrat stanowiący podstawę wyrażania idei i komunikowania się [Reddy 1979]. Metafora ta oparta była na założeniu, że idee są podobne do obiektów fizycznych, a zadaniem języka jest przekazywanie pakietów idei pomiędzy umysłami. W tym i podobnych modelach zawsze wyróżniany jest początek procesu komunikacji, czyli „nadawca”, następnie „kanał”, w którym transmitowana jest informacja (tzn. zakodowany komunikat) oraz adresat, czyli „odbiorca”, gdzie informacja jest dekodowana i komunikat jest przetwarzany².

Modele przewodu opierają się zazwyczaj na deterministycznym poglądzie na komunikację naukową. Ich założeniem jest, że wszystko, co autor wprowadza do systemu, po określonej liczbie kroków, w końcu dociera do użytkownika finalnego. W rzeczywistości nie jest to prawdą; łańcuch informacji nie jest przewodem, w którym to, co na wejściu, musi zawsze pojawić się na wyjściu. Część publikacji (twierdzi się nawet, że większość) pozostaje niezauważona, inne zostaną pominięte, a efekt jest taki, że z pewnością nie każda publikacja trafi do każdego użytkownika końcowego. Modele te są zorientowane na rozpowszechnienie, przez co nazywane są modelami pchającymi (*push-models*). Nie uwzględniają one gromadzenia informacji (czyli modelu ciągnącego – *pull-model*), w którym komunikacja jest inicjowana przez użytkownika³. Bardziej realistyczną reprezentacją

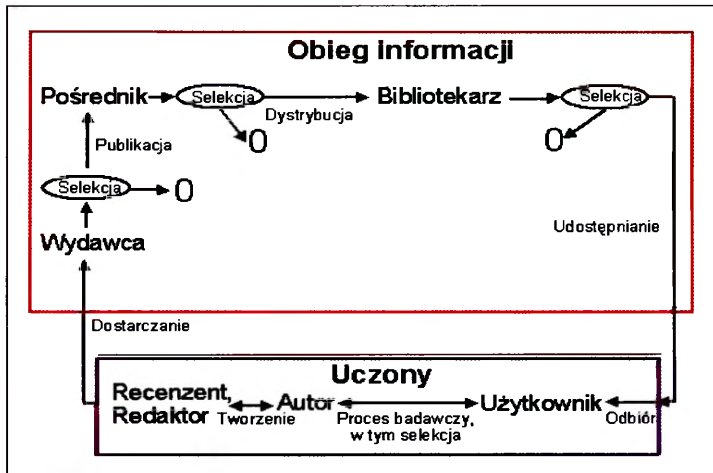
¹ Metafora jest tu rozumiana nie tyle jako struktura gramatyczna czy wyrażenie alegoryczne, lecz jako metoda percepcji. W tym sensie metafory używa się po to, aby osoby funkcjonujące w różnych kontekstach i posiadające różne doświadczenia rozumieli coś intuicyjnie. Za pomocą metafor syntezujemy w nowy sposób posiadane informacje i wyrażamy tę część wiedzy, której nie potrafimy jeszcze nazwać. Metafora pomaga osiągnąć ten cel przez połączenia różnych i często odległych doświadczeń w jeden obraz lub symbol – dwóch pojęć w jedno wyrażenie. Nowe wyrażenie często zawiera sprzeczności, które pozwalają uruchomić proces kreowania nowej wiedzy.

² Model Shannona i Weavera uwzględniał pięć elementów: źródło informacji, transmitter kodujący informację na sygnały, komunikat, kanał, odbiornik dekodujący komunikat z sygnału i odbiorcę. Podczas kodowania i dekodowania mogą pojawiać się szumy, czyli zakłócenia. Na przepływ informacji wpływ ma przepustowość i pojemność kanału.

³ Obecnie funkcjonujący model publikowania oparty jest na koncepcji „ciągnięcia” (*pull*): użytkownicy zainteresowani publikacją idą do biblioteki lub repozytorium i „wyciągają” dokument z zasobów. Od użytkownika wymaga się więc podjęcia konkretnych działań. Z tego powodu niezbędne jest zorganizowanie magazynowania; należy przechowywać informację, dopóki użytkownik o nią nie zapyta. W niektórych obszarach publikowania i być może znacznie szerzej w przyszłości model ten jest zastępowany przez koncepcję „pchania” (*push*): użytkownik określa rodzaj dokumentów, którymi jest zainteresowany, a relewantne materiały są mu natychmiast przesyłane po opublikowaniu lub utworzeniu, bez podejmowania przez niego bezpośrednich działań. W takim przypadku nie ma potrzeby realizacji pamięci w kanale dystrybucji [Owen 1998]. W praktyce częste są przypadki mieszania technologii *push* i *pull*; na przykład tworzenie i wysyłanie poczty elektronicznej można uznać za użycie technologii *push*, natomiast odebranie takiej wiadomości – za użycie technologii *pull*. Im bardziej używana technologia jest interakcyjna, tym mniej poddaje się takim dychotomicznym podziałom.

tego, co rzeczywiście dzieje się w łańcuchu informacji, jest sytuacja przedstawiona przeze mnie na rys. 11. Autorzy i użytkownicy przedstawieni są tu jako główne węzły, co pozwala na zróżnicowanie ich ról.

Ciągłe i swobodne przemieszczanie się aktora – uczonego w obrębie tych dwóch ról stanowi sedno procesu naukowego z punktu widzenia komunikacji naukowej. W modelu wskazałem ważną rolę selekcji, dokonywanej praktycznie na każdym etapie funkcjonowania łańcucha informacji. Jest ona podstawową zasadą działania obiegu informacji, który charakteryzuje się raczej nadmiarem, niż niedoborem. Selekcja odbywa się już na początku łańcucha informacji, gdy autor decyduje, z kim, o czym pisać i gdzie dostarczyć przygotowany tekst, a wydawca dokonuje selekcji dostarczonych dokumentów na podstawie określenia ich jakości (na przykład za pomocą recenzowania) i innych kryteriów (takich, jak spodziewany zysk). W centrum informacji (na przykład bibliotece), w którym zinstytucjonalizowane są funkcje pamięci, przewidziano dwa miejsca selekcji: podczas gromadzenia zbiorów oraz podczas wyboru przez użytkownika relewantnej informacji z zasobu informacji. Przepływ informacji od autora do użytkownika jest negocjowany w tych trzech punktach selekcji. W jej wyniku ilość informacji, dostarczanej użytkownikowi, maleje (dane odrzucone w wyniku selekcji oznaczone są jako 0). Tak dzieje się aż do samego końca tego łańcucha, gdzie użytkownik decyduje o tym, jakie informacje zechce odebrać, poprzez jakie kanały i jak je wykorzystać. Tradycyjnie taka selekcja była uznawana za domenę aktorów – pośredników, a nie autora/użytkownika: wydawcy i pośrednicy decydują, co jest potrzebne i możliwe na rynku, a bibliotekarze (poprzez tworzenie i realizację polityki gromadzenia zbiorów) określają, co i w jaki sposób zechcą przekazać swoim użytkownikom. W rzeczywistości jednakże w łańcuchu informacji istnieją dwa obiegi. Pierwszy jest „tradycyjnym” obiegiem informacji pomiędzy wydawcami, pośrednikami i bibliotekarzami. Drugi jest bardziej konceptualny; autorzy składają ofertę, a użytkownicy selekcjonują i pobierają informacje o wynikach badań i nowych ideach, transmitowanych poprzez obieg informacji za pomocą mediów drukowanych.



Rys. 11. Obieg informacji naukowej w systemie tradycyjnym (publikacji drukowanych)

Model ukazuje wielość ról pracowników nauki, którzy w procesie badawczym występują zawsze w dwóch (co najmniej) rolach jednocześnie. Można więc stwierdzić, że pracownik nauki zawsze odgrywa rolę prosumenta, którym według terminologii Alvina Tofflera jest konsument aktywny, uczestniczący w tworzeniu produktu lub usługi [Toffler 2001, s. 305]. Obecnie, dzięki dysponowaniu środkami komunikacji cyfrowej, uczony, jak każdy prosument, według Tofflera „przesuwa proces produkcji ze sfery gospodarki oficjalnie uznanej przez ekonomistów do sfery gospodarki lekceważonej”, czyli przenosi proces publikowania z oficjalnych przedsiębiorstw wydawniczych do Sieci.

Organizacja łańcucha informacji w oparciu o funkcje aktorów ulega zmianom, spowodowanym cyfryzacją formatów stosowanych w komunikacji naukowej oraz rozwojem komunikacji sieciowej. W celu opisu możliwych sposobów przejścia od publikowania drukiem do elektronicznego zastosowałem model dystrybucji Mackenzie Owena, skupiającego się na dwóch podstawowych funkcjach łańcucha informacji: selekcji i przechowywania (nazwanej funkcją „pamięci”) [Owen 1998].

Na rys. 12 prezentuję własny model, oparty na obiegu informacji w łańcuchu informacji cyfrowej. Cyfryzacja obiegu informacji znacznie skomplikowała ten proces. W tym modelu „selekcja” używana jest w nieco innym znaczeniu, niż w przedstawionym wcześniej modelu tradycyjnego obiegu informacji: tam selekcja polegała na kolejnym odrzucaniu materiałów nieodpowiednich (selekcja negatywna), tutaj na wyborze materiałów odpowiednich (selekcja pozytywna). Można wskazać także inne różnice: według Briana Whitwortha i Roba Friedmana w modelu tradycyjnym, nazywanym przez nich „feudalnym” niewielu decyduje o tym, co ma czytać wielu. Obieg informacji cyfrowej pozwala na „demokratyzację” systemu komunikacji naukowej, dzięki temu, że autorzy, oprócz wcześniejszych możliwości przekazania tekstu wydawcy, uzyskują zdolność do autopublikowania w wielu serwisach GBC [Whitworth, Friedman 2009]. Powoduje to zmianę ról wydawców i recenzentów, którzy zamiast ferowania wyroków zajmują się raczej udzielaniem wskazówek użytkownikom mającym swobodę wyboru.

Model umożliwia przedstawienie kilku (teoretycznych) opcji pozwalających na selekcję i przechowywanie informacji:

1. Najbardziej radykalną opcją jest likwidacja aktorów pośredniczących: użytkownicy selekcionują materiały w sieci, umieszczone tam bezpośrednio przez autora, który umożliwia także opiniowanie (na przykład w blogach uczonych).

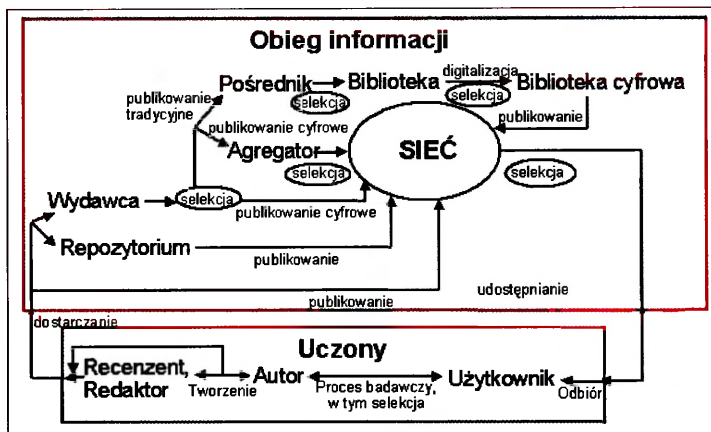
2. Innym rozwiązaniem jest eliminacja jedynie funkcji realizowanych przez ośrodki informacji (biblioteki) i publikowanie elektroniczne realizowane przez wydawców (czasem za pośrednictwem agregatorów): w tym przypadku użytkownicy otrzymują dostęp poprzez sieć do materiałów przechowywanych w serwisach wydawców⁴.

3. Kolejnym rozwiązaniem, mogącym powodować likwidację tradycyjnego ośrodka informacji jest publikowanie elektroniczne w różnego rodzaju repozytoriach (instytucjonalnych, dziedzinowych) – nowych ośrodkach informacji (serwisach GBC), w których dokumenty cyfrowe umieszczane są przez użytkownika. Repozytoria mogą zapewniać jakiś rodzaj selekcji (na przykład po opublikowaniu).

4. Ostatnim sposobem dystrybucji obiektów sieciowych jest ich wtórne publikowanie w serwisach GBC (bibliotekach cyfrowych), po zdigitalizowaniu ich wersji drukowanych.

Zauważmy, że w każdym z opisanych przypadków możemy mieć do czynienia z modelem pchania, realizowanym przy pomocy nowych technologii, na przykład RSS, w oparciu o spersonalizowane kryteria selekcji (profile użytkowników).

⁴ W przypadku dostępu płatnego na stronach wydawców komercyjnych, rolę biblioteki tradycyjnej staje się zapewnienie dostępu do publikacji wydawcy dla określonej grupy użytkowników przez wykupienie w ich imieniu licencji na dostęp; takie rozwiązanie powoduje jedynie modyfikację sposobu realizacji funkcji centrum informacji.



Rys. 12. Dystrybucja obiektów elektronicznych w Sieci

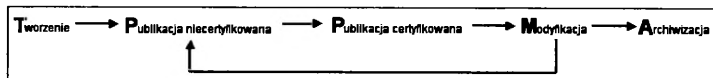
Z przedstawionego modelu wynika wtórna rola druku w komunikacji naukowej. Publikacje drukowane wcześniej czy później czeka digitalizacja w serwisach GBC. Proces digitalizacji spowalniany jest przez prawo autorskie; 70-letni okres ochrony tych praw ma być może uzasadnienie w przypadku literatury pięknej, ale nie tekstów naukowych (zob. p. 5.3). Zauważyć można także wielką rolę i potrzebę nowych sposobów zapewnienia jakości publikacji naukowych, gdyż dwa z wymienionych sposobów nie zapewniają kontroli jakości w tradycyjnym sensie (poprzez recenzowanie; na rys. 12 brak „selekcji” podczas samodzielnego publikowania w sieci oraz podczas publikowania w repozytoriach). W tych przypadkach niezbędne jest wypracowanie nowych sposobów zapewnienia jakości zasobów sieciowych. Model ten zwraca także uwagę na fakt, że obecnie niemal nie ma już publikacji czysto tradycyjnych, wyłącznie drukowanych; drukowi towarzyszy wersja cyfrowa, która na ogół także jest publikowana.

Analiza obiegu informacji w środowisku cyfrowym pozwala na wyciągnięcie wniosku, że model przewodu nie jest odpowiednią reprezentacją funkcjonowania komunikacji naukowej. Przeprowadzona analiza wskazuje, że łańcuch informacji funkcjonuje raczej jako przestrzeń, zawierająca wiele punktów wyboru, w których aktorzy negocjują włączenie własnego i uzyskanie dostępu do obcego, certyfikowanego zasobu wiedzy naukowej. Koncepcja takiej przestrzeni zawiera ideę użytkownika zaangażowanego, aktywnie włączającego się w procesy przepływu informacji [Owen 2007, s. 91]. W efekcie komunikacja naukowa musi być postrzegana poprzez paradygmat konstruktywistyczny, gdzie nie ma pojedynczego, obiektywnego „naukowego stanu rzeczy”, ale funkcjonuje mnóstwo „poglądów na świat” wynikających z konkretnych działań uczonych w obrębie systemu komunikacji naukowej.

Inną wadą modelu przewodu jest to, że upraszcza on rolę autora, nie biorąc pod uwagę złożoności relacji pomiędzy działalnością badawczą a informacją. Rolę autora przedstawić można w ramach cyklicznego modelu procesu badawczego, który wyróżnia trzy kolejne fazy badań: przygotowawczą, realizacji i upublicznienia. Na każdym z tych etapów pracownicy nauki muszą korzystać z danych i informacji zewnętrznych, w trakcie własnych badań gromadząc nowe informacje, dodawane do istniejących wcześniej i poszerzające

istniejącą wiedzę. Rodzaj i rola informacji zewnętrznej oraz sposób jej wykorzystywania jest różny na poszczególnych etapach badań. Formalna komunikacja naukowa pozostaje procesem dyskretnym⁵, w którym pracownicy nauki uzyskują wyniki badań raczej periodycznie, niż w sposób ciągły. Ponieważ komputeryzacja w kontekście IIN może powodować bardziej ciągłe formy komunikacji, problemy stwarzać może ocena efektów badań, tradycyjnie oparta raczej na dyskretnych wynikach niż na procesie. Dodatkowo uczeni, poza uczestnictwem w formalnym łańcuchu informacji, są obciążeni bieżącą, nieformalną działalnością komunikacyjną z kolegami.

W systemie komunikacji naukowej ważnym zagadnieniem jest funkcjonalność w odniesieniu do różnych aktorów. W tradycyjnym systemie komunikacji doszło do znacznego usztywnienia funkcjonalności systemu, zadań aktorów i podziału ról. Wydaje się, że komputeryzacja spowoduje ponowne otwarcie i uelastycznienie systemu. Wydawcy i bibliotekarze przejmują wzajemnie swoje funkcje, pewne zadania są włączane w system komunikacji sieciowej, a jeszcze inne przekazywane (przynajmniej w pewnym zakresie) autorom i użytkownikom lub ich instytucjom (na przykład w formie samopublikowania lub autoarchiwizacji). Dodatkowo powstają nowe funkcje (jak prepublikowanie lub długotrwała archiwizacja), wymagające jakiejś formy kontroli i zaangażowania aktorów. Powstają pomysły bardziej radykalnej cyfryzacji, prowadzącej do pełnej kontroli systemów komputerowych nad komunikacją, bez udziału aktorów instytucjonalnych. Jednak wiele ważnych funkcji, realizowanych w łańcuchu informacji, w szczególności ocena i zapewnienie jakości, wymaga wsparcia instytucjonalnego; takie rozwiązania uwzględniane są w ramach tworzonej IIN.



Rys. 13. Kontinuum publikacji elektronicznych

W okresie przejściowym komunikacja naukowa staje się coraz bardziej złożona. Obok starych form, wciąż istniejących, powstają nowe. Wiele z nich funkcjonuje równocześnie, powodując istnienie rozwiązań „hybrydowych”. Przepływ informacji naukowej od autora do użytkownika nie jest już pojedynczym, dobrze zdefiniowanym procesem, składającym się z odrębnych etapów, zarządzanym przez aktorów o jasno wyznaczonych rolach. Cyfryzacja spowodowała zwiększenie różnorodności dokumentów, aktorów i sposobów komunikacji, mnożąc ścieżki, po których informacja naukowa może się przemieszczać, zwiększając także liczbę punktów dostępu użytkowników do informacji, w zależności od miejsca w cyklu życia publikacji.

Komunikację naukową można przedstawiać jako proces, w którym treść i sens wyników badań naukowych ewoluują w czasie, podczas przechodzenia przez łańcuch informacji. Proces ten składa się co najmniej z następujących etapów (rys. 13):

- Faza tworzenia: w tej fazie wyniki badań (dane) transformowane są w formę informacji, która może być zapisana i upubliczniona.
- Faza publikacji niecertyfikowanej: w tej fazie publikacja przekazywana jest w sposób nieformalny, na przykład jako szara literatura lub preprint. Nie została ona poddana ocenie jakości, jednak dokument rozpoczął aktywne uczestnictwo w dyskursie naukowym.
- Faza publikacji certyfikowanej: dokument został oceniony albo przed (ocena recenzentów), albo po (ocena użytkowników) opublikowaniu.

⁵ Proces dyskretny to proces, składający się z wyodrębnionych faz, którego efektem są odrębne produkty, w odróżnieniu od procesów ciągłych.

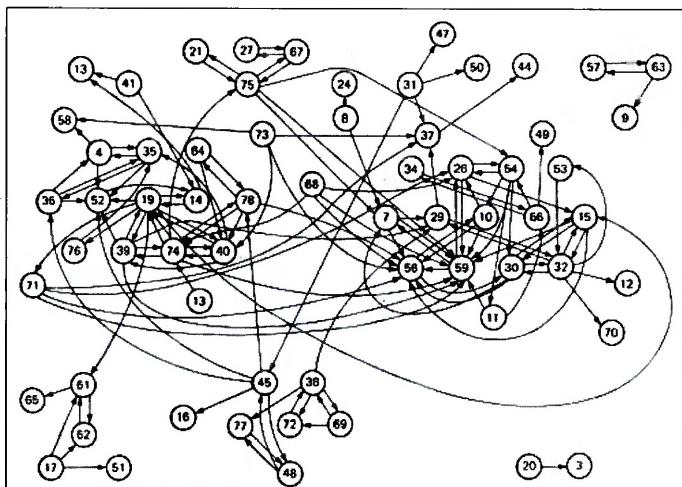
- Faza modyfikacji: w tej fazie dokument może być zmieniony (dodane mogą być uwagi lub nowe części), w celu usunięcia pomyłek lub uaktualnienia: w tym momencie następuje sprzężenie zwrotne do fazy niecertyfikowanej, czyli etapu komunikacji nieformalnej.
- Faza archiwizacji: na tym etapie publikacja przestaje odgrywać aktywną rolę w procesie naukowym (oraz prawdopodobnie traci swą wartość naukową), jednak pozostaje jako zapis w dorobku nauki, więc nadal może uczestniczyć w procesach komunikacji naukowej.

Tworzenie i rozpowszechnianie wiedzy jest procesem, w którym zgodnie z podejściem konstruktywistycznym, dochodzi do interakcji pomiędzy technologią a środowiskiem relacji społecznych [Barney 2008, s. 53]. Każda praca naukowa, służąca pomnażaniu wiedzy, opiera się na wcześniejszych osiągnięciach innych uczonych. Pomiędzy pracownikami nauki stale, dynamicznie odbywa się wymiana wartości, norm, metod i systemów nagradzania. Nauka akademicka funkcjonuje w sposób, który antropologzy nazywają „kulturą wymiany darów” [Lyman 1996]. Uczelnie opłacają swoich pracowników naukowych. Rząd i jego agendy pokrywają koszty badań i kształcenia studentów. Często mówi się, że pracownicy nauki są członkami „niewidzialnych uniwersytetów” (zob. p. 1.4), składających się z uczynnych o podobnych zainteresowaniach i wymieniających między sobą zasoby, dane i informacje. Stanowią grupę konstytuowaną wspólną wiedzą, celami i normami zachowań, a także realizującą intensywne kontakty społeczne i komunikację interpersonalną. Istnienie takiej grupy, którą nazwać można społecznością wiedzy w określonym obszarze nauki lub dyscypliny badań, nie poddaje się pełnemu opisowi zależności społecznych i sposobów komunikacji pomiędzy jej członkami ze względu na ich złożoność [Baptista, Rodrigues, Machado 1999]. Stanowi ona system wielu, wielopoziomowych i częściowo pokrywających się „podspołeczności”.

Uczni wykorzystują mechanizmy komunikacji formalnej, pisząc i czytając artykuły, referaty na konferencje, książki i inne publikacje. Z drugiej strony stosują mechanizmy komunikacji nieformalnej, z których część została przetransponowana z komunikacji formalnej. System komunikacji formalnej jest ściślej dopasowany do struktury danej społeczności wiedzy. Przed rozwojem Internetu komunikacja nieformalna ograniczona była głównie do bezpośrednich spotkań, rozmów telefonicznych, przesyłek pocztowych, adnotacji na dokumentach. Pracownicy nauki stosując nowe technologie telekomunikacyjne tworzą publikacje a następnie rozmawiają o nich, informują się nawzajem i pozyskują nową wiedzę. W takich warunkach granica między komunikacją formalną a nieformalną zaciera się coraz bardziej, co może doprowadzić do nieadekwatności takiego podziału.

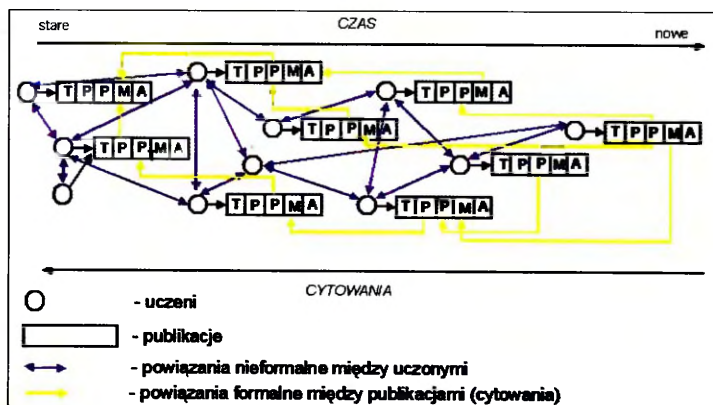
Usługi internetowe umożliwiają członkom społeczności wiedzy, pracownikom instytucji naukowych, regularne stosowanie mechanizmów komunikacji nieformalnej, takich jak poczta elektroniczna, fora dyskusyjne, blogi. Jednocześnie formy, które dotychczas uchodziły za element komunikacji formalnej stają się „mniej formalne”. W jaki sposób nazwać na przykład preprint umieszczony na stronie autora lub w repozytorium? Z jednej strony jest to artykuł wydrukowany później (lub jednocześnie) w recenzowanym czasopiśmie, z drugiej jest to tekst, który nie podlegał certyfikacji. Podobnie zatarciu ulegają różnice między rodzajami publikacji, takimi, jak obszerny artykuł w czasopiśmie, raport z badań i średniej objętości książka.

Komunikacja nieformalna prowadzi do powstawania tak zwanych społecznych sieci uczonych, opisanych przez Briana Vickery’ego [Vickery 2005]. Te sieci nieformalnych więzi międzyosobowych są sybystytutem wspólnot, działających zarówno offline, jak i online; GBC dostarcza w tym zakresie nowych możliwości [Castells 2008, s. 364]. W wyniku badania wzajemnych powiązań uczonych, tzn. określenia „kto konsultuje się z kim i jak często” otrzymuje się sieć podobną do przedstawionej na rys. 14.



Rys. 14. Nieformalne powiązania pomiędzy uczonymi

Ze względu na powstające więzi można wyróżnić kilka sposobów komunikowania się uczonych: wielu uczonych utrzymuje kontakty z wieloma innymi (np. nr 59 z rys. 14), inni przyjmują kontakty z zewnątrz oraz sami je inicjują (np. nr 32), niektórzy utrzymują kontakt tylko z jednym uczonym, zainicjowany przez niego (np. nr 47, 9, 12), z niektórymi nikt nie inicjuje kontaktu (np. nr 20, 31). Widoczne są także dwie lub trzy podgrupy (w rogach rysunku), które słabo lub w ogóle nie współpracują z resztą uczonych.



Rys. 15. Hybrydowy model komunikacji naukowej

Podobne związki i grupy uczonych można odnaleźć dzięki analizie cytowań i współcytowań. Ben Johnson i Charles Oppenheim stwierdzili ścisły związek między sieciami społecznymi uczonych i sieciami powstającymi w wyniku cytowań [Johnson, Oppenheim 2007, s. 624-625]. Możliwe jest graficzne obrazowanie relacji pomiędzy twórcami i użytkownikami obiektów informacyjnych w postaci grafu [Mayer, Rauber 2009; Strotmann, Zhao 2009, s. 507]. W zależności od tego, do jakiej grupy zaliczyć można uczonego, istnieją zwykle wyraźne różnice w rodzaju współpracy, jakiej oczekuje on od pracowników informacji naukowej.

Podczas realizacji komunikacji naukowej, sieć społeczna uczonych i realizowana w jej obrębie komunikacja nieformalna (w Internecie i poza nim), łączona jest z sieciami powstającymi w wyniku rozpowszechniania i cytowania treści kanałami komunikacji formalnej (rys. 15.). Komunikacja nieformalna jest uzupełniona powiązaniami komunikacyjnymi, wynikającymi z analizy cytowań. Przedstawiony wcześniej model kontinuum publikacji elektronicznych należy uzupełnić przez dodanie fazy cytowania, przez co dokument nie jest traktowany jako wyizolowana jednostka, ale włączony zostaje w strukturę sieci cytowań w nauce. Posłużyłem się w tym celu modelem zaprezentowanym przez Carla Lagoze, opisującym połączenie sieci publikacji naukowych z siecią społeczną uczonych w jedną sieć hybrydową⁶ [Lagoze 2004, s. 92]. Łączy on podejście stosujące łańcuch informacji (dla publikowania naukowego) z sieciami społecznymi uczonych. Oficjalny system komunikacji naukowej tworzy sieci dokumentów (żółte strzałki zwrócone w jedną stronę, od nowych publikacji do wcześniejszych). Dokumenty na rysunku są symbolizowane przez prostokąty podzielone na kolejne fazy, zgodnie z rys. 13. Jednym z efektów utworzenia repozytoriów elektronicznych jest zmniejszenie roli wymiaru czasu cytowań: sto lat temu można było jedynie cytować dokumenty opublikowane około roku wcześniej (i oczywiście wcześniejsze); obecnie cytować można artykuły, które jeszcze nie ukazały się formalnie drukiem. Graf cytowań jest podstawą oceny jakości publikacji, przy pomocy indeksów cytowań lub analizy hiperlinków. Uwzględnienie podczas analizy cytowań przedmiotów publikacji cytowanych i cytujących pozwala otrzymać graficzny obraz powiązań i relacji pomiędzy poszczególnymi dziedzinami i ich poddziedzinami.

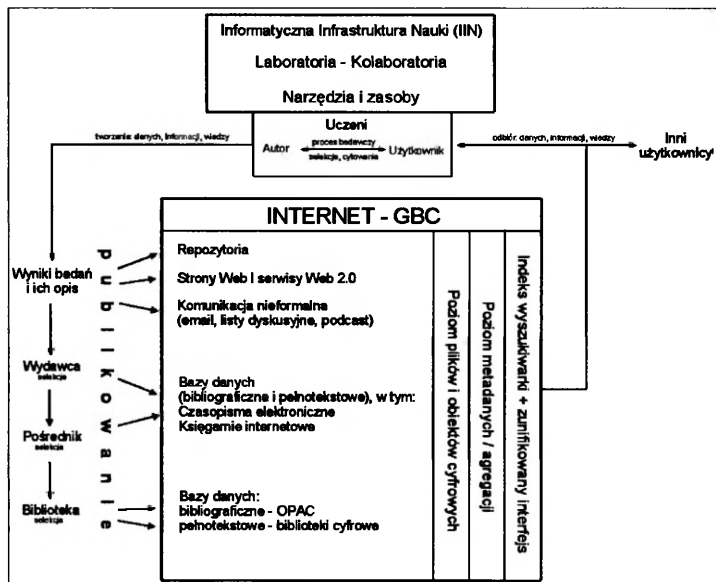
Na sieć dokumentów nakłada się społeczna sieć autorów i użytkowników, czyli uczonych (czarne kółka, połączone niebieskimi strzałkami). Sieci społeczne pomiędzy uczonymi są bardziej dynamiczne, niż sieci dokumentów. Uczni komunikują się nie tylko poprzez publikowanie formalne, ale także za pomocą nieformalnych dyskusji. Kanały komunikacji nieformalnej składają się z bogatych przepływów informacji, które ułatwiają tworzenie nowych idei i dojrzewanie teorii naukowych.

Jak z tego wynika, coraz trudniej jest uznać publikację naukową za jednostkę niezmienną⁷ i dobrze zdefiniowaną. System komunikacji naukowej stał się niezwykle dynamiczny. Duże możliwości modyfikacji dokumentów, szybko powstająca sieć cytowań i bogaty repertuar środków komunikowania nieformalnego stanowią główne elementy obecnej sytuacji. Formalna, certyfikowana publikacja naukowa wydaje się być jedną z wielu materializacji wyników badań naukowych, która może zostać poprzedzona i zastąpiona przez inne materializacje.

⁶ Podobną koncepcję reprezentował wcześniej Loet Leydesdorff pisząc o dwupoziomowym systemie sieciowym składającym się z (1) społecznych relacji między uczonymi i (2) relacji pomiędzy komunikatami. Cytowania wynikają z interakcji między tymi dwiema sieciami [Leydesdorff 1998].

⁷ O niezmienności publikacji drukowanej można mówić z punktu widzenia konkretnej materializacji dzieła: pomiędzy materializacjami już mogą zdarzać się różnice, chociaż zazwyczaj nie są one zasadnicze. Poważne zmiany zachodzą, gdy przyjmujemy punkt widzenia ciągłości dzieł, wynikającej z rozwoju nauki i dyskursu naukowego w czasie. Wówczas różnica pomiędzy poziomem zmienności publikacji drukowanych i elektronicznych nie będzie wyrażana dychotomicznie, jako niezmienna-zmienna, lecz jako bardziej dyskretny ciąg stanów: trudno zmienna-latwo zmienna i wolno zmienna-szybko zmienna.

Związki powstające między uczonymi w wyniku realizowanej przez nich komunikacji naukowej, przypominają wspomniane już sieci „małych światów”, charakteryzujących się tym, że powiązane siecią jednostki (uczni), oprócz wielu połączeń z najbliższymi sąsiadami w sieci, posiadają również mniejszą liczbę połączeń dalekiego zasięgu. Dzięki istnieniu tych połączeń dwa dowolne obiekty w sieci powiązane są krótkimi łańcuchami relacji – „sześcioma stopniami oddzielenia”. Sieci małych światów mogą być modelem zdecentralizowanego wyszukiwania informacji w GBC, pozwalającym wyjaśnić sposób poprawnego działania wielkich struktur.



Rys. 16. Model komunikacji naukowej w GBC

Na rys. 16 przedstawiłem model, zawierający wszystkie opisane przeze mnie elementy komunikacji naukowej, stanowiące część IIN, we wzajemnych relacjach. Stanowi on rozwinięcie zaprezentowanego wcześniej modelu. U góry rysunku prezentuję elementy IIN, takie jak zbiory danych, narzędzia do ich analizy, zasoby Gridu oraz wszystkie pozostałe zasoby umożliwiające prowadzenie badań naukowych w środowisku cyfrowym. Częścią IIN są także uczni, pełniący podwójną rolę twórców i użytkowników informacji naukowej. Poniżej znajduje się poziom elektronicznej komunikacji naukowej, gdzie umieszczonych zostało wiele środków komunikowania formalnego i nieformalnego. Uczni, jako prosumenci, sami rozwijali w Internecie efektywne formy komunikacji, pozwalające na szersze rozpowszechnianie wyników badań; część form istniała wcześniej, inne są typowe dla komunikowania w środowisku cyfrowym. W wyniku publikowania, obiekty cyfrowe, proste i złożone, umieszczane są w Internecie. Obiektom tym towarzyszą metadane, tworzone przez człowieka, systemy automatycznego indeksowania i/lub powstające w wyniku pełnotekstowego indeksowania wyszukiwarki. Interfejs wyszukiwarki stanowi

zunifikowane środowisko, udostępniające informację i/lub przekierowujące do odpowiedniego serwisu GBC. Wzrost widzialności wyników badań naukowych zależy więc z jednej strony od skali publikowania w Sieci, a z drugiej od zdolności wyszukiwarek do ich wyszukania.

We wszystkich miejscach przedstawionych na rysunku odbywa się tworzenie, przetwarzanie i udostępnianie – odbiór (poprzez umieszczanie w serwisach GBC) danych, informacji i wiedzy (zob. p. 1.4). Oprócz interpretacji danych uzyskanych w trakcie badań (informacji w postaci publikacji) uczony może udostępnić także same dane, gromadzone w specjalnych serwisach GBC.

Możliwość korzystania z wyników badań naukowych otrzymują także użytkownicy spoza instytucji nauki. Chodzi zarówno o działalność popularyzatorską samych uczonych, jak również o bezpośredni dostęp do publikacji poprzez wyszukiwarki typu Google, przybierający rozmiary nie spotykane wcześniej, w okresie funkcjonowania komunikacji naukowej, opartej na publikacjach drukowanych. Amatorskie użytkowanie informacji naukowej jest nawet powodem stawiania tez o szkodliwości łatwego uzyskiwania zbyt specjalistycznej informacji przez użytkownika nieprzygotowanego [Weitzman 2004].

Model komunikacji naukowej, opisany w tej części książki, przedstawia sposób postrzegania komunikacji naukowej przez naukę o informacji naukowej. Zmiany w modelach komunikacji naukowej odzwierciedlają ewolucję poglądów na komunikację, od linearnego, technokratycznego i systemowego punktu widzenia, do poglądów opartych na postrzeganiu komunikacji naukowej jako systemu społecznego. System komunikacji naukowej zawiera skomplikowaną sieć relacji, funkcjonujących w ramach sieci socjotechnicznej. Taki punkt widzenia pozwala nie tylko na znacznie bogatszą analizę łańcucha informacji, ale także wyjaśnia, że innowacje nie są prostym procesem technologicznym.

ZAKOŃCZENIE:

Kierunki rozwoju komunikacji naukowej

Początki bibliotek cyfrowych sięgają pierwszych lat dziewięćdziesiątych. XX wieku; projektowane one były w odpowiedzi na nowe możliwości sieciowego rozpowszechniania wiedzy. Już wówczas przyszłość komunikacji naukowej wiązano ze środowiskiem cyfrowym. Od tego czasu nastąpił znaczny postęp na drodze ku realizacji tej wizji, co nie znaczy, że nie napotymano na trudności. Wraz z nabywaniem nowych doświadczeń zmiany ulegały, często szybciej niż pierwotnie zakładano, wcześniej wytyczone cele. Cyfryzacja komunikacji naukowej spowodowała istotne, a przy tym w dużej części nieprzewidywane początkowo zmiany w jej organizacji. Jak już wcześniej stwierdziłem, takie zjawisko jest typowe dla powstawania wszelkich dużych infrastruktur: wraz z ich realizacją cel oddala się i często zmienia kształt. Niektóre pierwotne zamierzenia twórców bibliotek cyfrowych, czyli stworzenie sprawnie zarządzanych, sieciowych repozytoriów obiektów cyfrowych czy też potrzeba uwzględnienia w nich różnych rodzajów dokumentów, formatów i nośników, pozostały główną potrzebą infrastruktury, służącej realizacji badań naukowych i komunikacji naukowej. Jednak dopiero w XXI wieku rozpoczęły się poważne prace nad tworzeniem wielkich zasobów, w szczególności obiektów nietekstowych oraz różnicowanie ich zakresu. Wtedy też pojawiła się możliwość integracji wszystkich lokalnych działań w jedną, globalną infrastrukturę.

Obecne rozwiązania stwarzają możliwości transformacji sposobu prowadzenia badań w wielu dziedzinach nauki, a nawet umożliwiają tworzenie nowych obszarów badawczych, często interdyscyplinarnych. Głównym celem prac, opisanych w tej książce, stał się system, obsługujący badania i komunikację naukową, który wykorzystywałby nowe możliwości na tyle efektywnie, aby mogły być stosowane intuicyjnie, a jego funkcjonowanie pozostawało transparentne dla użytkowników. Rezultatem jest informatyczna infrastruktura nauki (IIN), obsługująca wielkie i zróżnicowane (ze względu na potrzeby informacyjne) grupy użytkowników, uzupełniona jednak narzędziami Nauki 2.0, pozwalającymi na konstytuowanie się w Sieci niewielkich grup badaczy, powiązanych wspólnymi celami. Powstający system zbliża się do wizji Memexu Vannevara Busha [Bush 1945]: urządzenia pozwalającego na elektroniczny zapis wszystkiego, co się słyszy, widzi i czyta, a nawet doświadcza w sposób cyfrowy czy rzeczywisty; indeksujący, wyszukujący i przetwarzający te dane na żądanie; a w końcu przedstawiający wyniki w zindywidualizowany sposób, w zależności od potrzeb i możliwości (na przykład dostępnych urządzeń) i zapewniający archiwizację danych i informacji.

GBC ma coraz większy wpływ na życie naukowe, a także ekonomiczne i społeczne, gdyż dzięki niej każdy ma możliwość tworzenia, wyszukania, selekcji, organizacji, archiwizacji i wielokrotnego wykorzystania treści cyfrowych na różne, nowe i efektywne sposoby. Stąd prosta droga do wzrostu produktywności, innowacyjności i kreatywności. Sądzi się nawet, że produktywność w zakresie działalności informacyjnej może wzrosnąć, od początku wieku do 2012 r., dwukrotnie [Larsen, Wactlar 2003, s. 25]. Serwisy GBC stanowią narzędzia umożliwiające integrację informacji na wyższym poziomie (globalnej biblioteki cyfrowej), co ułatwia lepsze zarządzanie zasobami po niższych kosztach. Dla

pełnego wykorzystania powstających zasobów informacyjnych konieczne jest jednak zapewnienie współdziałania wielkich zasobów informacji z małymi, osobistymi bibliotekami cyfrowymi, tak, aby każdy mógł ściągnąć potrzebne mu dane w celu ich przetworzenia (*mashup*) oraz rozpoznać swoje zasoby (surowe dane z badań, dane przetworzone – publikacje, jak i metadane).

W GBC, w której udostępniana jest cała ludzka wiedza, wartość informacji wynika nie z jej istnienia, lecz z wykorzystywania. Tylko ograniczony zasób informacji będzie podlegał bezpośrednio ludzkiemu przetwarzaniu. Przemiany GBC podążają w kierunku minimalizacji udziału ludzi w poszukiwaniu potrzebnej informacji, optymalizacji jej przetwarzania i transformacji, a także opracowania wyników tych procesów, ułatwiających jej zrozumienie i wykorzystanie. Wszystko to w celu wsparcia uczonego w realizacji kreatywnej, intelektualnej transformacji informacji, prowadzącej do odkrycia naukowego – nowej wiedzy.

Wiele technologii związanych z infrastrukturą taką, jak Internet i Web, rozwijało się od wykorzystania w instytucjach naukowych, czasem militarnych, poprzez zastosowania komercyjne, aż do etapu powszechnego stosowania przez wszystkich. Globalna biblioteka cyfrowa znajduje się na tej samej drodze, która prowadzi do powstania ogólnie dostępnego środowiska informacyjnego, zwielokrotniającego zmiany, spowodowane już przez sam Internet. Trudność może stanowić różnorodność serwisów GBC, a więc brak jednolitego modelu realizacji usług. Być może obecne zróżnicowanie w tym zakresie jest zjawiskiem okresu przejściowego i niedaleka przyszłość pozwoli na wyłonienie kilku najbardziej efektywnych modeli serwisów GBC, zróżnicowanych na przykład pod względem poziomu odbiorcy, na wzór Google i Google Scholar.

Inną, może jeszcze poważniejszą, przyczyną dysfunkcji staje się dążenie do zapanowania nad swobodnym, jak dotąd, przepływem treści w sieci globalnej. Paradoksalnie technologia cyfrowa, pozwalająca na niemal nieskrępowaną transmisję treści w przestrzeni i czasie, może jednocześnie służyć kontroli tej transmisji w skali wcześniej nieznannej. Propozycje dystrybutorów treści w zakresie nowych regulacji mogą doprowadzić do podziału użytkowników na niewielką grupę, posiadającą dostęp do szybkich łączy i pełnych zasobów, oraz na większość, korzystającą ze standardowych rozwiązań, zmuszoną do akceptacji ograniczonych możliwości komunikacyjnych i blokowania dostępu do konkurencyjnych zasobów Webu. Podobne, negatywne znaczenie mogą mieć niektóre proponowane rozwiązania prawne, na przykład postulujące ograniczenie dotychczasowych granic dozwolonego użytku osobistego w sieciach komputerowych [Barta, Markiewicz 2008, s. 232].

Globalna biblioteka cyfrowa jest skomplikowaną strukturą wielu elementów, które w obrębie Sieci globalnej współpracują ze sobą tworząc zintegrowane środowisko informacyjne. Można zadać pytanie, czy GBC jest realnie funkcjonującą infrastrukturą, czy tylko jedną z wielu metafor, jakimi opisywana jest globalna sieć – Internet¹. Szukając odpowiedzi na to pytanie, charakteryzowałem poszczególne elementy współczesnego środowiska cyfrowej komunikacji naukowej, wyszczególnione przez twórców modelu 5S jako niezbędne części składowe biblioteki cyfrowej [Gonçalves i in. 2007, s. 1420]. Opis tych elementów pozwolił mi na przedstawienie różnic pomiędzy tradycyjnym rozumieniem biblioteki naukowej, a globalną biblioteką cyfrową jako metaforą systemu cyfrowej i sieciowej komunikacji naukowej, w której funkcjonują serwisy GBC, m.in. biblioteki cyfrowe. Wyniki można przedstawić następująco:

¹ Pomijam opinie skrajne, których wyrazicielem jest Clifford Stoll, twierdzący, że wszystko, co się dzieje w Sieci jest metaforą: rozmowami bez mówienia, uśmiechami bez poruszania ustami i obejmowaniem bez dotykania. Clifford Stoll w ten sposób podkreśla, że internauci żyją metaforami, zamiast prawdziwym życiem [Stoll 2000, s. 56]. To samo jednak można chyba powiedzieć np. o czYTElnictwie powieści, czyli innych, od dawna istniejących formach komunikowania.

Zasoby danych i informacji

Repozytoria obiektów cyfrowych znajdują się na najniższym poziomie wielopiętrowej infrastruktury, na którą składają się: Internet – informatyczna infrastruktura nauki – globalna biblioteka cyfrowa – serwisy GBC. Globalna biblioteka cyfrowa jest, tak jak wszystkie infrastruktury, powiązana poprzez relacje społeczne i technologie z innymi strukturami. Jednocześnie w jej obrębie także funkcjonują bogate struktury serwisów GBC. Poziom treści sam może być uważany za infrastrukturę, która powinna służyć interpretacji, wykorzystaniu i swobodnemu udostępnianiu dokumentów, a więc działać jako infrastruktura informacyjna, uwzględniająca treść informacji. Do realizacji tego poziomu infrastruktury wykonano wiele prac w kierunku przygotowania odpowiednich standardów i narzędzi. Jednym z ich podstawowych zadań jest integracja elementów globalnej, wielopoziomowej infrastruktury informacyjnej w kierunku jednolitego, z punktu widzenia użytkownika, zasobu danych i informacji transformowanych w wiedzę. W przyszłości należy oczekiwać dalszej integracji tych struktur.

Dane i informacje, ich zasoby, a także organizacja tych zasobów, zawsze stanowiły jedną z najważniejszych części i funkcji biblioteki. W GBC dane występują w postaci elektronicznej, co narzuca stosowane rozwiązania, ale też otwiera nowe możliwości, które wciąż nie są jeszcze w pełni odkryte, a tym bardziej wykorzystywane. Należy oczekiwać powstawania nowych form publikowania, czemu towarzyszyć muszą nowe sposoby organizowania zasobów tych publikacji, w tym ich opisywania. W GBC publikowane są nie tylko przetworzone dane (informacje), jak w druku, ale także surowe, nieprzetworzone dane, pochodzące wprost z urządzeń laboratoryjnych, gdzie powstają w trakcie realizowanych eksperymentów badawczych. Współistnienie różnych zasobów informacyjnych, o różnym stopniu przetworzenia powoduje, że ulegają zatarciu różnice pomiędzy komunikacją formalną i nieformalną, na rzecz komunikacji utylitarnej, jak najlepiej wspomagającej procesy tworzenia wiedzy.

Te przemiany wyznaczają także kierunki zmian publikacji elektronicznych. Publikacje te nadal zachowują tradycyjną formę w dużej części ze względu na potrzeby administracji nauki, a mniej samych uczonych. Powszechną praktyką jest umieszczanie reprintu artykułu w repozytorium i powiadamianie o nim zainteresowanych odbiorców, którymi są koledzy zajmujący się badaniami w tej samej dyscyplinie. Wówczas publikowanie artykułu w czasopiśmie elektronicznym, wydawanym na wzór tradycyjnego, służy udostępnieniu tego tekstu badaczom z innych, pokrewnych dyscyplin, a także przedstawieniu tego faktu w sprawozdaniu dla celów administracyjnych. Wcześniej nieformalnie opublikowany tekst jest podstawą dyskusji, służącej m.in. weryfikacji jakości i może być uzupełniany dodatkowymi materiałami.

Zauważalny jest gwałtowny wzrost liczby dostępnych w Webie treści o wysokiej jakości, płatnych lub bezpłatnych z punktu widzenia użytkownika końcowego. Rozwojowi cyfrowej komunikacji naukowej towarzyszą oczekiwania otrzymywania cyfrowych treści bezpłatnie, co jest próbą kontynuowania wcześniejszej sytuacji: w tradycyjnych bibliotekach naukowych uczeni także otrzymują potrzebne im publikacje bez wnoszenia opłat. Rosnący popyt na informacje powoduje coraz większy nacisk na twórców rozwiązań prawnych i funduszoodawców w kierunku regulacji zapewniających bezpłatny dostęp do jak największej ilości informacji naukowych. Przejawem tych tendencji jest ruch Open Access, stanowiący ideologię bezpłatnego korzystania w Sieci z tego, za co już wcześniej zostało zapłacone w różny sposób, także poprzez redystrybucję środków, skierowanych do instytucji naukowych, a pochodzących z podatków osobistych obywateli, zysków przedsiębiorstw i innych źródeł. W ruchu tym zaczynają uczestniczyć nawet wydawcy komercyjni, tworząc modele ekonomiczne oparte na mikropłatnościach lub przerzucaniu opłat z użytkownika końcowego na kogoś innego. Z punktu widzenia użytkownika, którego biblioteka wykupiła dostęp, usługi serwisu takiego wydawcy to również typowy Open

Access; użytkownik ma bezpłatny dostęp online, często z własnego komputera domowego, do wykupionych (przez kogoś) zasobów². Chodzi w takim razie o znalezienie sposobu na bezpłatne udostępnienie w Sieci (w wersji elektronicznej) tego, co dotychczas, również bezpłatnie, dostępne było w bibliotekach naukowych (w wersji analogowej)³.

Zasadniczą rolę do odegrania mają w GBC dane o danych, czyli metadane. Metadane, powstające automatycznie w wyniku pracy narzędzi wyszukiwawczych, współistnieją z metadanymi (ustrukturyzowanymi), tworzonymi w serwisach GBC przez ludzi lub automatycznie. Ważne, aby w każdym przypadku zapewnić możliwość indeksowania zasobu przez wyszukiwarki, czyli maszynową interpretację rekordów bibliograficznych, dzięki stosowaniu właściwych standardów i formatów (takich, jak XML). Metadane ustrukturyzowane mogą być przydatne dla uczonych stale korzystających z usług określonego serwisu, zapewniającego dostęp do literatury specjalistycznej. Pracownicy naukowcy w standardowych sytuacjach badań literaturowych stosują bowiem zarówno wyszukiwarki, jak i kilka wybranych, znanych sobie baz danych z reprezentowanej przez siebie dziedziny [Young, Seggern 2001, s. 161]. Poza metadanymi opisowymi, w serwisach GBC powstają i wykorzystywane są także inne rodzaje metadanych, które mają podstawowe znaczenie dla funkcjonowania tych serwisów, w tym głównie metadane administracyjne i strukturalne.

Narzędzia wyszukiwawcze

Efektywne wyszukiwanie danych i informacji jest procesem o zasadniczym znaczeniu, gdyż warunkuje możliwość i sens realizacji większości innych procesów GBC. Jednym z najważniejszych zadań do rozwiązania, dotyczących stosowania informacji w formie cyfrowej, jest obecnie nie tyle wyszukiwanie informacji relewantnej, lecz sprawne posługiwanie się wielkimi zbiorami wyników, pochodzących z przeszukiwania wielu serwisów GBC jednocześnie, szczególnie, gdy wyniki zawierają olbrzymią liczbę nie powtarzających się obiektów relewantnych. Rozwiązaniem może być stosowanie filtrowania i dodatkowych kryteriów wyszukiwawczych, na przykład rodzaju nośnika lub języka dokumentu. Obecnie tworzy się wiele niezbędnych narzędzi, chociaż powstałe rozwiązania oczekują na bardziej powszechne zastosowanie i podbudowę teoretyczną. Wynika z tego, że aby stworzyć zadowolający system informacyjno-wyszukiwawczy – serwis GBC, nie wystarczy dostarczenie relewantnych wyników wyszukiwania; obecnie podczas projektowania takiego serwisu niezbędne staje się zastosowanie osiągnięć wielu dyscyplin, takich jak kogniwytyka, lingwistyka, psychologia oraz oczywiście informatyka.

Jedną z podstawowych cech GBC powinna być możliwość wyszukiwania, opartego na pojedynczym, prostym interfejsie narzędzia, które pozwala na dostęp do wszystkich globalnych zasobów lub informacji o nich. Użytkownik otrzymuje wyniki wyszukiwania, uszeregowane według specyficznie rozumianej relewancji. Takie podejście opiera się na sposobie funkcjonowania współczesnych wyszukiwarek. Powinny one kierować użytkownika bezpośrednio do relewantnych obiektów cyfrowych lub przynajmniej do właściwych serwisów, oferujących dostęp do obiektów istniejących w formie cyfrowej lub do metadanych obiektów tradycyjnych. Serwisy GBC stosują zazwyczaj własne, zróżnicowane

² Niektóre badania dowodzą, że dzięki cyfryzacji publikacji naukowych (głównie czasopism), w krajach najbardziej rozwiniętych (np. w USA i Wlk. Brytanii) liczba dostępnych w bibliotekach tytułów od 2001 r. niemal się podwoiła. Nic dziwnego, że ponad 60% uczonych jest w tych krajach zadowolonych z dostępu do literatury naukowej. Niestety w innych krajach dostęp do publikacji naukowych jest gorszy lub znacznie gorszy [Ware, Mabe 2009, s. 42-43].

³ Podobną myśl przedstawił w swojej książce Łukasz Gołębski: „Internet oferuje nam masę usług za darmo. Wyszukiwarka Google działa za darmo. Konto pocztowe zakładamy za darmo. Przez Skype rozmawiamy za darmo. Przyzwyczailiśmy się, że wiele rzeczy, które kiedyś były mniej lub bardziej kosztowne, dziś są za darmo. Oczywiście pozornie za darmo, bo przecież cena jest gdzieś ukryta” [Gołębski 2009, s. 34].

interfejsy, pozwalające na bardziej skomplikowane wyszukiwanie; ich przykładem mogą być biblioteczne OPAC. Katalogi te, szczególnie centralne (w tym na poziomie globalnym), pozostaną narzędziami łączącymi zasoby dokumentów tradycyjnych z cyfrowymi, zarówno w sensie ich wyszukania, jak i umożliwienia dostępu. Innym rozwiązaniem jest stosowanie wyszukiwania sfederowanego, które jednak nie ma możliwości zapewnienia dostępu do całości, ani nawet poważnej części zasobów światowych. Serwisy GBC będą także interpretować i dokonywać transformacji informacji, zawartych w dokumentach niesamoistnych; wykresach, tabelach czy mapach, jak i w dokumentach graficznych i dźwiękowych (wyszukiwanie uniwersalne).

Systemy informacyjne tworzą w GBC skomplikowaną, wielopoziomową strukturę, przechodzącą obecnie wiele transformacji, których efektem będzie zapewne ujednoczenie narzędzi i technik wyszukiwania informacji. Wyszukiwanie poprzez jeden wspólny indeks i interfejs wyszukiwarki jest uzupełniane innymi technikami, często pochodzącymi z systemów społecznościowych, takimi jak rekomendacje, powiązania pomiędzy publikacjami cytowanymi i cytującymi, opinie i komentarze użytkowników. Serwisy GBC mogą wykorzystywać dane, dotyczące poziomu udostępniania materiałów do doskonalenia procesu wyszukiwania.

Istotną cechą GBC jest tzw. globalizacja. Przykładem łączenia potrzeb lokalnych z globalnym zasobem jest możliwość tworzenia kastomizowanych wyszukiwarek na bazie wyszukiwarek globalnych typu Google. Ułatwiają one zaspokajanie indywidualnych potrzeb wyszukiwawczych. W pewnym sensie można do nich zaliczyć wyszukiwarkę Google Scholar, która ma realizować potrzeby wyszukiwawcze określonej grupy użytkowników – uczonych. Systemy specjalistyczne, pozwalające na współdziałanie z informacjami dostępnymi w sieci rozległej, mogą być cechą charakterystyczną serwisów GBC.

Serwisy i usługi

W zakresie usług, oferowanych przez serwisy GBC, niezbędne jest zachowanie właściwych proporcji pomiędzy tym, co jest lokalne a tym, co powinno być udostępniane globalnie. W tym zakresie możliwe są dwa skrajne podejścia. Można próbować tworzyć ogólny zestaw narzędzi i usług dla wszystkich dyscyplin, języków, krajów i kultur, co może być atrakcyjnym projektem z technicznego i ekonomicznego punktu widzenia. Podczas takiego postępowania należy jednak uwzględnić fakt, że każdy system, którego twórca stara się zaspokajać wszystkie potrzeby każdego użytkownika, zazwyczaj nie zaspokaja dobrze żadnych potrzeb. Na poziomie lokalnym, podczas projektowania narzędzi i usług, ściśle dopasowanych do indywidualnych potrzeb odrębnych grup użytkowników, społeczności i zespołów badawczych, może dochodzić natomiast do dublowania wysiłków. To podejście jest typowe dla bibliotekarzy i specjalistów z zakresu informacji naukowej, dotąd z powodzeniem stosowane w tradycyjnych systemach informacyjnych. Właściwym rozwiązaniem jest zapewne lokalna kastomizacja usług GBC dostępnych globalnie, czyli wspomniana globalizacja.

Stały rozwój technologii i spadek ich cen umożliwią tworzenie lokalnych banków danych i bibliotek osobistych, zawierających terabajty danych. Połączenie tych osiągnięć z przestrzenią Internetu pozwoli na powstawanie globalnych serwisów GBC, takich jak wideoteki, dające dostęp do wszystkich filmów, jakie kiedykolwiek wyprodukowano. W tym kierunku zmierzają niedoskonałe jeszcze próby tworzenia serwisów stanowiących „światowy katalog książek”⁴ poprzez scalanie opisów z wszelkich dostępnych źródeł (głównie OPAC bibliotecznych i księgarni internetowych), co pozwala użytkownikom tych serwisów na wymianę opinii i tworzenie własnych podkatalogów dzieł ulubionych.

⁴ Przykładami takich serwisów mogą być wspomniane wcześniej biblios.net oraz [Librarything](http://Librarything.com/) (<http://pl.librarything.com/>) a także doskonalszy [Open Library](http://openlibrary.org/) (<http://openlibrary.org/>), tworzony przez Internet Archive. W Polsce przykładem jest repozytorium [YADDA](http://yadda.icm.edu.pl/) (<http://yadda.icm.edu.pl/>).

Dokumenty udostępniane przez serwisy GBC z zasobów pasywnych, staną się podstawą aktywnych usług, dostosowujących interfejsy użytkowników do aktualnych wymagań, a gromadzących i filtrujących informacje na podstawie przewidywanych, potencjalnych potrzeb. Organizacja serwisów GBC ulega modyfikacjom od sztywnych struktur hierarchicznych w kierunku samoorganizujących się sieci użytkowników.

Jednocześnie zmienia się zapotrzebowanie na rozmiar otrzymywanej przez użytkownika jednostki tekstu, z agregatu treści na poszczególne, niewielkie jego części, a więc z czasopisma lub encyklopedii na pojedynczy artykuł, akapit, wybraną definicję. Można oczekiwać, że serwisy dostarczające taką informację ułatwią eksplorację danych i tekstu. Dzięki tym procesom można odnajdywać stałe wzorce oraz anomalie, występujące w zgromadzonych zasobach, co umożliwi odkrywanie ważnych zależności, prowadzących do często istotnych odkryć naukowych.

Lokalne serwisy, ich zasoby i usługi są integrowane w bibliotekę globalną, dzięki funkcjonowaniu globalnych narzędzi indeksujących i wyszukiwawczych. Szczególną rolę odgrywają tu działania, służące scalaniu technologii, powstałych w bibliotekach tradycyjnych w efekcie komputeryzacji, a pozwalających na tworzenie światowych zasobów metadanych opisowych (OPAC), z technologiami typowymi dla sieci rozległych, głównie globalnymi wyszukiwarkami internetowymi. Rezultatem będzie współdziałanie wielkich zasobów danych i informacji, bez względu na środowisko, w jakim powstawały. Dzięki temu konsument informacji cyfrowej staje się jednocześnie konsumentem treści, cyfrowych lub tradycyjnych. Informacja cyfrowa jest więc impulsem do sekwencji zachowań informacyjnych w globalnym środowisku cyfrowym [Nicholas i in. 2008a, s. 1].

W zakresie usług, służących udostępnianiu danych w GBC, można zauważyć pewne sprzeczności, jak wspomniana sprzeczność między potencjalnie nieograniczonym zasięgiem udostępniania i dużymi możliwościami jego restrykcyjnego ograniczania. Pogodzenie tych sprzeczności, w sposób wygodny dla użytkownika i korzystny dla dostawcy usług, wydaje się ważnym warunkiem pomyślnego rozwoju GBC. Doświadczenia, wynikające z usług komercyjnych świadczonych w Internecie, a polegających na przykład na dostarczaniu muzyki rozrywkowej o ściśle reglamentowanym dostępie, nie mogą i nie powinny być bezpośrednio stosowane w serwisach służących komunikacji naukowej.

Społeczności

Osiągnięcie efektywności funkcjonowania architektury informatycznej infrastruktury nauki wymaga zapewnienia dostępu do treści, usług i narzędzi oraz określenia poziomu spójności i sposobów zarządzania tą infrastrukturą, bowiem nie istnieje pojedyncza instytucja ani organizacja, a tym bardziej osoba, zarządzająca Internetem, infrastrukturą naukową lub globalną biblioteką cyfrową w tym zakresie. Stosowane rozwiązania muszą być elastyczne, aby umożliwić rozproszone dostarczanie zróżnicowanych treści, narzędzi i usług. Ich dostawcy z naturalnych względów dążą do rozszerzenia kontroli nad oferowanymi przez siebie zasobami i produktami; z kolei użytkownicy domagają się prostego i jednolitego, uniwersalnego dostępu do wszelkich zasobów i usług informacyjnych. Sprzeczne potrzeby prowadzą do powstawania konfliktów.

Rozproszone zarządzanie infrastrukturą nauki i komunikacji naukowej przynosi zarówno korzyści, jak i straty. W sytuacji równoważenia się interesów wielu społeczności system funkcjonuje poprawnie. W przypadku utraty równowagi, tak jak to ma miejsce obecnie, niektórzy uczestnicy tych procesów są nieproporcjonalnie uprzywilejowani w stosunku do innych. Funkcjonowanie części ważnych elementów infrastruktury, takich jak instytucje zarządzające przestrzeniami nazw, narzędziami typu tezaurusów i ontologii lub strukturami metadanych, jest niewidoczne z zewnątrz, przez co łatwo może dojść do ich zaniedbania i utraty. Natomiast inne elementy tej infrastruktury, na przykład dostawcy treści (wydawcy), doskonale widoczni dzięki prowadzeniu intensywnego marketingu, uzyskują ponadprzeciętne zyski (zob. p. 1.2).

W systemie nauki i komunikacji naukowej funkcjonuje wiele grup, w obrębie których istnieje silna konkurencja. Uczni konkurują o publikacje i ich czytelników, granty naukowe, zatrudnienie, studentów, nagrody i wyróżnienia. Uczelnie konkurują o uczonych, studentów, granty i dotacje oraz o wysokość budżetu. Wydawcy konkurują o autorów, kontrakty z bibliotekami i wysokość sprzedaży. Agendy finansujące konkurują o projekty badawcze i zasoby na ich realizację. Podczas konkurowania każdego z każdym w obrębie tego samego rodzaju działalności, powstają nieprzewidywalne interakcje, które można opisać przy pomocy teorii gier. Ponieważ trudno jest przewidzieć przebieg każdej z tych gier, niemożliwe jest także przewidzenie efektu wielu możliwych kombinacji interakcji [Borgman 2007, s. 264]. Rozwój IIN i GBC jest naturalnym efektem funkcjonowania wielkiego, skomplikowanego systemu interakcji, który – tak jak każdy wielki system – ma mechanizmy, dające możliwości samoregulacji, zapewniające stały rozwój, chociaż czasem w nieoczekiwanych kierunkach.

Jednym z nich jest wykorzystanie sieci globalnych do tworzenia małych grup na zasadzie wspólnych zainteresowań. Przyczynia się do tego rozwój serwisów Nauki 2.0, u podstaw których znajduje się założenie o zbiorowej inteligencji użytkowników, pozwalającej doskonalić funkcjonowanie aplikacji Web. Użytkownicy oczekują możliwości wykorzystania przestrzeni społecznej, w której mogą funkcjonować, prowadząc dialog na temat wyszukiwania i wyszukanych materiałów. Mają możliwość tworzenia swoich własnych, wyodrębnionych, kastomizowanych miniserwisów GBC, które są wyrazem odwrotu od oficjalnych, zhierarchizowanych struktur na rzecz rozwoju wielokierunkowego. Małe, samoorganizujące się społeczności (gildie) uczonych powstają jako sieci osób o podobnych zainteresowaniach i potrzebach. Świadczone przez nie usługi mogą stanowić wyzwanie dla istniejących struktur serwisów zinstytucjonalizowanych.

Jedną z najważniejszych cech obecnej sytuacji w komunikacji naukowej są zmiany, które starałem się opisać w tej książce. Są one tak wielopłaszczyznowe, a jednocześnie wzajemnie powiązane, że próby przewidywania ich skutków, nawet w nieodległej przyszłości, napotykają na wiele trudności. Niewątpliwie kierunki tych zmian są godne uwagi i dalszych badań, gdyż mogą prowadzić do głębokich przeobrażeń w zasadach uprawiania nauki, co z kolei spowoduje znaczne przyspieszenie postępu.

Zmiany te następują w kierunku integracji wielu rozwiązań związanych z osiągnięciami, znanymi pod różnymi nazwami. Obecnie mówi się o Webie 3.0, jako połączeniu osiągnięć w zakresie Semantycznego Webu, Webu społecznościowego, mikroformatów, przetwarzania języka naturalnego i prowadzenia wyszukiwania w tym języku, eksploracji danych, personalizacji funkcjonowania serwisów GBC, systemów agentowych. Jednocześnie następuje bardzo szybki postęp w zakresie urządzeń: mobilnych, sieciowych, rozproszonego przetwarzania (Grid). To wszystko uzupełniane jest nową organizacją funkcjonowania, opartą na otwartości: protokołów, formatów, oprogramowania, identyfikacji (OpenID) i co najważniejsze z punktu widzenia GBC: treści. Wszystkie te osiągnięcia łącznie tworzą nowe środowisko, wykorzystywane m.in. w komunikacji naukowej, w którym ich współdziałanie powoduje powstawanie nowych jakości i wartości; zapewne więc wkrótce mówić będziemy o Webie 4.0 i następnych.

Biblioteki naukowe, stanowiące dotychczas główne instytucje pamięci, odpowiedzialne za zachowanie i udostępnianie dorobku naukowego, znajdują się w centrum tych przemian. Ich działalność może stać się funkcjonalną częścią tej nowopowstającej, złożonej struktury, nazwanej przeze mnie globalną biblioteką cyfrową, stanowiącą zintegrowane środowisko informacyjne nauki.

Rozwój nowej, cyfrowej infrastruktury komunikacji naukowej powoduje zacieranie się granic pomiędzy społecznościami tradycyjnie wspomagającymi tę komunikację, takimi jak wydawcy, księgarze, bibliotekarze, informatycy. Obecnie konieczna jest znajomość,

a przynajmniej orientacja we wszystkich zagadnieniach związanych z komunikacją naukową. Z tym związane są nowe wyzwania stojące przed zawodem bibliotekarza i innymi, funkcjonującymi w obrębie komunikacji naukowej. W związku z tym bibliotekarze muszą stawać się specjalistami w zakresie sieciowej komunikacji naukowej i jej szczegółowych zagadnień, takich jak protokoły, formaty, narzędzia wyszukiwawcze i innych, przedstawionych w mojej książce.

Z opisu poszczególnych elementów GBC wynika, że jest ona środowiskiem informacyjnym, którego nie można w pełni opisać przez analogię do bibliotek tradycyjnych, czy nawet odrębnych serwisów GBC, wzorowanych na bibliotekach tradycyjnych, takich jak lokalne biblioteki cyfrowe. Internet nie jest biblioteką; nie można takiej metafory traktować dosłownie [Brabazon 2006, s. 163]. Być może biblioteka naukowa, taka jak opisywana w publikacjach i podręcznikach bibliotekarskich przełomu wieków XIX i XX już nie istnieje – cyfryzacja i digitalizacja zmieniły ją na zawsze. Podobnie nie istnieją już dawni bibliotekarze, wydawcy, a przede wszystkim zmieniły się potrzeby użytkowników – oczekują oni innych efektów procesów komunikacji. Uczni chcą i muszą stosować wszelkie, dostępne sposoby komunikacji naukowej, a GBC daje szansę na ich integrację w jedno, przyjazne środowisko. Zachodzące zmiany przypominają tzw. ruchy komputeryzacyjne, opisywane przez Roberta Klinga i Suzanne Iacono, prowadzące do powstawania nowego porządku społecznego [Iacono, Kling 2001]. Pozostaje przystosować się do nowych warunków funkcjonowania, stwarzanych przez globalne środowisko informacyjne Internetu, w którym powstaje nowy, zrównoważony układ interesów, charakterystyczny dla współczesnego i przyszłego sieciowego społeczeństwa informacyjnego, ewoluującego w społeczeństwo wiedzy [Zacher 2006, s. 75].

PUBLIKACJE WYKORZYSTANE*

- Aalberg, Trond (2003). *Supporting relationships in digital libraries*. Oslo: Norwegian Univ. of Science and Technology.
- Abrams, Stephen [dok. elektr.] (2007). File formats. W: S. Ross, M. Day (ed.) *DCC digital curation manual*. <http://www.dcc.ac.uk/resource/curation-manual/chapters/file-formats>.
- Ackerman, Mark [dok. elektr.] (1994). Metaphors along the information highway. <http://www.eecs.umich.edu/~ackerm/pub/94b10/diac.final.html>.
- Adams, Andrew [dok. elektr.] (2008). Re: Explaining and justifying a mandate. W: BOAI Forum. Southampton: School of Electronics and Computer Science, Univ. of Southampton October, 12 2008. boai-forum@ecs.soton.ac.uk.
- Allinson, Julie; Sebastien François; Stuart Lewis [dok. elektr.] (2008). SWORD: Simple Web-service Offering Repository Deposit. *Ariadne*, nr 54. <http://www.ariadne.ac.uk/issue54/allinson-et-al/>
- Allinson, Julie; Pete Johnston; Andy Powell [dok. elektr.] (2007). A Dublin Core application profile for scholarly works. *Ariadne*, nr 50. <http://www.ariadne.ac.uk/issue50/allinson-et-al/>.
- Anderson, Chris [dok. elektr.] (2004). The long tail. *Wired Magazine*, vol. 10, nr 12. <http://www.wired.com/wired/archive/12.10/tail.html>.
- Anderson, James; José Pérez-Carballo (2001). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part 1, Research and the nature of human indexing. *Information Processing and Management*, vol. 37, nr 2, s. 231-254.
- Andriessen, Daniel; Marien Van Den Boom (2009). In search of alternative metaphors for knowledge; inspiration from symbolism. *Electronic Journal of Knowledge Management*, vol. 7, nr 4, s. 397-404.
- Apps, Ann; Ross MacIntyre [dok. elektr.] (2006). Why OpenURL? *D-Lib Magazine*, vol. 12, nr 5. <http://www.dlib.org/dlib/may06/apps/05apps.html>.
- Armbruster, Chris [dok. elektr.] (2008a). Access, usage and citation metrics: what function for digital libraries and repositories in research evaluation? <http://ssrn.com/abstract=1088453>.
- Armbruster, Chris (2008b). Cyberscience and the knowledge-based economy, open access and trade publishing: from contradiction to compatibility with nonexclusive copyright licensing. *International Journal of Communications Law and Policy*, nr 12, s. 20-38.
- Armbruster, Chris [dok. elektr.] (2008c). A european model for the digital publishing of scientific information? <http://ssrn.com/abstract=1106162>.
- Armbruster, Chris (2007). Moving out Oldenburg's long shadow: what is the future for society publishing? *Learned Publishing* vol. 20 nr 4, s. 259-266.
- Arms, William [dok. elektr.] (2008). Cyberscholarship: high performance computing meets digital libraries. *Journal of Electronic Publishing*, vol. 11, nr 1. <http://hdl.handle.net/2027/spo.3336451.0011.103>.
- Arms, William [dok. elektr.] (2005). A viewpoint analysis of the digital library. *D-Lib Magazine*, Vol. 11, Nr 7/8. <http://www.dlib.org/dlib/july05/arms/07arms.html>.
- Arms, William [dok. elektr.] (2002). Quality control in scholarly publishing on the Web. *Journal of Electronic Publishing*, vol. 8, nr 1. <http://www.press.umich.edu/jep/08-01/arms.html>.
- Arms, William (2000). *Digital libraries*. Cambridge, Mass.: MIT Press.
- Arms, William [et al.] [dok. elektr.] (2006). A research library based on the historical collections of the Internet Archive. *D-Lib Magazine*, vol. 12, nr 2. <http://www.dlib.org/dlib/february06/arms/02arms.html>.
- Arunachalam, Subbiah (2008). Open Access to scientific knowledge. *DESIDOC Journal of Library and Information Technology*, vol. 28, nr 1, s. 7-14.

* Dla dokumentów opublikowanych w Internecie, o ile nie wskazano inaczej, aktualność adresów URL została sprawdzona w dniach 4-10 stycznia 2010 r.

- Atkins, Daniel [et al.] (2003). Revolutionizing science and engineering through cyberinfrastructure : report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Arlington: NSF.
- Bacza-Yates, Ricardo; Carlos Castillo (2006). Web search. W: Keith Brown (ed.) Encyclopedia of language and linguistics. Vol. 13. Oxford: Elsevier, s. 527-537.
- Bacza-Yates, Ricardo; Berthier Ribeiro-Neto (1999). Modern information retrieval. New York: ACM Press.
- Bailey, Charles (2006). What is open access? W: Niel Jacobs (ed.) Open Access: key strategic, technical and economic aspects. Oxford: Chandos Publ., s. 13-26.
- Baker, Karen; Florence Millerand [dok. elektr.] (2007). Scientific infrastructure design: information environments and knowledge processes. http://interoperability.ucsd.edu/docs/07BakerMillerand_07asist_KnowledgeProvinces.pdf.
- Baptista, Ana; Migiel Ferreira [dok. elektr.] (2007). Tea for two : bringing informal communication to repositories. *D-Lib Magazine*, vol. 13, nr 5/6. <http://www.dlib.org/dlib/may07/baptista/05baptista.html>.
- Baptista, Ana; Eloy Rodrigues; Altamiro Machado (1999). Online publishing as a support for scholarly communication in dynamic knowledge communities. W: ElPub'99. Redefining the Information Chain. New Ways and Voices. Ronneby, Sweden 12-12 May 1999. Washington: ICC Press, s. 236-249.
- Bard, Alexander; Jan Söderqvist (2006). Netokracja : nowa elita władzy i życie po kapitalizmie. Warszawa: Wydaw. Akad. i Profesjonalnie.
- Barney, Darin (2008). Społeczeństwo sieci. Warszawa: Wydaw. Sic!
- Barta, Janusz; Ryszard Markiewicz (2008). Prawo autorskie. Warszawa: Wolters Kluwer.
- Batorowska, Hanna (2009). Kultura informacyjna w perspektywie zmian w edukacji. Warszawa: Wydaw. SBP.
- Battelle, John (2006). Szukaj : jak Google i konkurencja wywołałi biznesową i kulturową rewolucję. Warszawa: PWN.
- Batty, Michael (1997). Virtual geography. *Futures*, vol. 29, nr 4/5, s. 337-352.
- Bawden, David; Polona Vilar (2006). Digital libraries: to meet or manage user expectations. *ASLIB Proceedings*, vol. 58, nr 4, s. 346-354.
- Beagrie, Neil (2006). Digital curation for science, digital libraries and individuals. *International Journal of Digital Curation*, vol. 1, nr 1, s. 3-16.
- Bednarek-Michalska, Bożena [dok. elektr.] (2002). Ocena jakości bibliotekarskich serwisów informacyjnych udostępnianych w Internecie. *Biuletyn EBIB*, nr 2(31). <http://ebib.oss.wroc.pl/2002/31/michalska.php>.
- Bcit-Arie, Oren [et al.] [dok. elektr.] (2001). Linking to appropriate copy : report of the DOI-based prototype. *D-Lib Magazine*, vol. 7, nr 9. <http://www.dlib.org/dlib/september01/caplan/09caplan.html>.
- Belkin, Nicholas; Robert Oddy; Hester Brooks (1982). ASK for information retrieval. *Journal of Documentation*, vol. 38, nr 2, s. 61-71.
- Bell, Jonathan; Stuart Lewis (2006). Using OAI-PMH and METS for exporting metadata and digital objects between repositories. *Program*, vol. 40, nr 3, s. 268-276.
- Benkler, Yochai (2008). Bogactwo sieci. Warszawa: Wydaw. Akad. i Profesjonalne.
- Bergstrom, Carl; Theodore Bergstrom (2004). The costs and benefits of library site licenses to academic journals. *Proceedings of National Academy of Sciences of the USA*, vol. 101, nr 3, s. 897-902.
- Berners-Lee, Tim (2000). Weaving the Web : The original design and ultimate destiny of the World Wide Web. New York: Harper Business.
- Bertot, John [et al.] (2006). Functionality, usability and accessibility. Iterative user-centered evaluation strategies for digital libraries. *Performance Measurement and Metrics*, vol. 7, nr 1, s. 17-28.
- Besser, Howard [dok. elektr.] (2002). The next stage: moving from isolated digital collections to interoperable digital libraries. *First Monday*, vol. 7, nr 6. http://www.firstmonday.org/issues/issue7_6/besser/index.html.
- Bianco, Simone [et al.] [dok. elektr.] (2007). Towards Open Access publishing in High Energy Physics : Report of the SCOAP Working Party. <http://scoap3.org/files/Scoap3WPRreport.pdf>
- Bieber, Michael [et al.] (2002). Toward virtual community knowledge evolution. *Journal of Management Information Systems*, vol. 18, nr 3, s. 11-35.

- Björk, Bo-Christer; Annikki Roos; Mari Lauri (2008). Global annual volume of peer reviewed scholarly articles and the share available via different Open Access options. W: Leslie Chan, Susanna Mornati (ed.) *Open scholarship : authority, community, and sustainability in the age of Web 2.0 : proceedings of the 12th International Conference on Electronic Publishing* held in Toronto, Canada 25-27 June 2008. Toronto: Univ. of Toronto, s. 178-186.
- Björncborn, Lennart (2004). Small-world link structures across an academic Web space : a library and information science approach. Copenhagen: Royal School of Library and Information Science.
- Boissy, Robert; Tina Feick; Leslie Knapp (2007). Journal pricing ten years later. W: Carol Borchet, Gary Ives (ed.). *Mile high views : surveying the serials vista*. Binghamton: The Haworth Press, s.167-182.
- Bojar, Bożenna (2009). Języki informacyjno-wyszukiawcze wczoraj, dziś... czy jutro? *ZIN*, nr 1, s. 3-24.
- Bojar, Bożenna (oprac.) (2002). *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiawczych*. Warszawa: Wydaw. SBP.
- Borgman, Christine (2008). Data, disciplines, and scholarly publishing. *Learned Publishing*, vol. 21, nr 1, s. 29-38.
- Borgman, Christine (2007). *Scholarship in the digital age: information, infrastructure and Internet*. Cambridge, Mass.; London: MIT Press.
- Borgman, Christine [dok. elektr.] (2004). The interaction of community and individual practices in the design of a digital library. <http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp9.pdf>.
- Borgman, Christine (2003a). From Gutenberg to the global information infrastructure: access to information in the networked world. Cambridge, Mass.: MIT Press.
- Borgman, Christine [dok. elektr.] (2003b). Personal digital libraries: creating individual spaces for innovation. NSF Workshop on Post-Digital Libraries Initiative Directions, 4 June. http://www.sis.pitt.edu/~dlwshop/paper_borgman.pdf.
- Borgman, Christine (1996). Why are online catalogs still hard to use? *Journal of ASIS*, vol. 47, nr 7, s. 493-503.
- Borgman, Christine; Jillian Wallis; Noel Enyedy (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, vol. 7, nr 1-2, s. 17-30.
- Borgman, Christine; Jonathan Furner (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, vol. 36, s. 3-72.
- Borgman, Christine [et al.] (2005). Comparing faculty information seeking in teaching and research : implications for the design of digital libraries. *Journal of ASIST*, vol. 56, nr 6, s. 636-657.
- Borgman, Christine [et al.] [dok. elektr.] (1996). Social aspects of digital libraries : final report. http://is.gscis.ucla.edu/research/dig_libraries/UCLA_DL_Report.html.
- Borgman, Christine [et al.] (1995). Children's searching behaviour on browsing and keyword online catalogs : the science library catalog project. *Journal of the ASIST*, vol. 46, nr 9, s. 663-684.
- Bourne, Philip; Lynn Fink [dok. elektr.] (2008). I am not a scientist, I am a number. *PLoS Computational Biology*, vol. 4, nr 12. <http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1000247>.
- Bowker, Geoffrey; Susan Star (2000). *Sorting things out : classification and its consequences*. Cambridge, Mass.: MIT Press.
- Brabazon, Tara (2006). The Google Effect : googling, blogging, wikis and the flattering of expertise. *Libri*, vol. 56, nr 3, s. 157-167.
- Bradley, Phil [dok. elektr.] (2008). Custom-built search engines. *Ariadne*, nr 55. <http://www.ariadne.ac.uk/issue55/search-engines/>.
- Brin, Sergey; Lawrence Page [dok. elektr.] (1998). The anatomy of a large-scale hypertextual Web search engine. <http://infolab.stanford.edu/~backrub/google.html>.
- Brindley, Gavin; Adrienne Muir; Steve Proberts (2004). Provision of digital preservation metadata: a role for ONIX? *Program*, vol. 38, nr 4, s. 240-250.
- Brody, Tim (2006). Evaluating research impact through Open Access to scholarly communication. [Praca doktorska]. Southampton: Univ. of Southampton.
- Brody, Tim; Stevan Harnad; Les Carr (2006). Earlier Web usage statistics as predictors of later citation impact: research articles. *Journal of ASIST*, vol. 57, nr 8, s. 1060-1072.

- Brooks, Terrence [dok. elektr.] (2004). The nature of meaning in the age of Google. *Information Research*, vol. 9, nr 3. <http://informationr.net/ir/9-3/paper180.html>.
- Brooks, Terrence [dok. elektr.] (2003). Web search: how the Web has changed information retrieval. *Information Research*, vol. 8, nr 3. <http://informationr.net/ir/8-3/paper154.html>.
- Brophy, Jan; David Bawden (2005). Is Google enough? Comparison of an Internet search engine with academic library resources. *ASLIB Proceedings*, vol. 57, nr 6, s. 498-512.
- Burk, Alan; James Kerr; Andy Pope [dok. elektr.] (2000). The credibility of electronic publishing : archiving and text fluidity/version control. <http://web.mala.bc.ca/hssfc/Final/Archiving.htm#Archive>.
- Burke, Mary (2009). The Semantic Web and the digital library. *ASLIB Proceedings*, vol. 61, nr 3, s. 316-322.
- Bush, Vannevar (1945). As We May Think. *The Atlantic Monthly*, vol. 176, nr 1, s. 101-108.
- Byrne, Alex [dok. elektr.] (2005). Promocja globalnych wspólnot informacyjnych. Oświadczenie IFLA przygotowane na Światowy Szczyt Społeczeństwa Informacyjnego w Tunisie. *Biuletyn EBIB*, nr 9 (70). <http://ebib.oss.wroc.pl/2005/70/byrne2.php>.
- Cardoso, Gustavo (2008). From mass to networked communication: communicational models and the informational society. *Intern. Journal of Communication*, nr 2, s. 587-630.
- Chalmers, Alan (1997). Czym jest to, co zwiemy nauką? Wrocław: Siedmioróg, s. 216 s.
- Chandrakar, Rajesh (2006). Digital object identifier system: an overview. *The Electronic Library*, vol. 24, nr 4, s. 445-452.
- Camp, Jean; Yu-Tsun Chien (2000). The Internet as public space: concepts, issues and implications in public policy. *ACM SIGCAS Computers and Society*, vol. 30, nr 3, s. 13-19.
- Candela, Leonardo [et al.] [dok. elektr.] (2007). Setting the foundations of digital libraries. The DELOS manifesto. *D-Lib Magazine*, vol. 13, nr. 3/4. <http://www.dlib.org/dlib/march07/castelli/03castelli.html>.
- Carvalho, Elizabeth [dok. elektr.] (2001). Grey literature and its contribution to knowledge society. 67th IFLA Council and General Conference, Boston 16-25 August. <http://www.ifla.org/IV/ifla67/papers/090-173e.pdf>.
- Centre for Information Behaviour and the Evaluation of Research [dok. elektr.] (2008). Information behaviour of the researcher of the future. A CIBER briefing paper. <http://www.bl.uk/news/pdf/googlegen.pdf>.
- Chan, Lois; Marcia Zeng [dok. elektr.] (2006). Metadata interoperability and standardization – a study of methodology. Part I. *D-Lib Magazine*, vol. 12, nr 6. <http://www.dlib.org/dlib/june06/chan/06chan.html>.
- Chandler, Daniel [dok. elektr.] (2000). The transmission model of communication. <http://www.aber.ac.uk/media/Documents/short/trans.html>.
- Chen, Chia-chen; An-Pin Chen (2007). Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*, vol. 25, nr 6, s. 711-724.
- Chen, Ching-chih [dok. elektr.] (2007). Challenges for developing a global digital library and gateway: from Global Memory Net to World Heritage Digital Library. http://www.fulbrightacademy.org/file_depot/0-10000000/20000-30000/21647/folder/59410/Chen+-+World+Heritage+Digital+Library.pdf.
- Chen, Ching-chih [dok. elektr.] (2003). Toward a Global Digital Library. http://www2.sis.pitt.edu/~dlwshop/paper_chen_ching.pdf.
- Chen, Ching-chih [dok. elektr.] (1998). Next generation Internet and the Global Digital Library. <http://puclink.org/annual/annual1998/1998pdf/chen-cc.pdf>.
- Chen, Ching-chih, Kevin Kiernan [dok. elektr.] (2002). Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials. <http://delos-noe.ici.pi.cnr.it/activities/internationalforum/Joint-WGs/digitalimaging/DigitalImaging.pdf>.
- Chen, Hsinchun (2004). Digital library research in the US: an overview with a knowledge management perspective. *Program*, vol. 38, nr 3, s. 157-167.
- Chien, Yi-Tzue (2004). Disruptive technologies, innovation and digital libraries research – the case of a billion-dollar business. W: International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society DLKC'04, 2-5.03.2004 Tsukuba, Japan. Tsukuba: Univ. of Tsukuba, s. 18-23.
- Choroś, Kazimierz (2002). Efektywność wyszukiwarek internetowych. W: Czesław Daniłowicz (red.) III Krajowa Konferencja nt. Multimedialne i Sieciowe Systemy Informacyjne MISSI 2002. Wrocław: Oficyna Wydaw. Politechn. Wrocławskiej, s. 85-94.

- Choudhury, Sayeed [et al.] [dok. elektr.] (2002). A framework for evaluating digital library services. *D-Lib Magazine*, vol. 8, nr 7/8. <http://www.dlib.org/dlib/july02/choudhury/07choudhury.html>.
- Cisek, Sabina [dok. elektr.] (2008). Badanie zachowań informacyjnych użytkowników bibliotek: metodologia Sense-Making. W: Maria Kocójowa (red.) *Biblioteka : klucz do sukcesu użytkowników [CD-ROM]*. Kraków: IINiB UJ, s. 97-103.
- Cisek, Sabina [dok. elektr.] (2007). Nauka o informacji na świecie w XXI wieku : badania metanaukowe. http://eprints.relis.org/12699/1/Cisek_in_na_swiecie_eng.pdf.
- Cisek, Sabina; Remigiusz Sapa (2007). Komunikacja naukowa w Internecie – mity i rzeczywistość. W: Wiesław Lubaszewski (red.) *Komputer – Człowiek – Prawo*. Kraków: Wydaw. UJ, s. 39-49.
- Clausen, Lars [dok. elektr.] (2004). Handling file formats. <http://netarchive.dk/publikationcr/File-Formats-2004.pdf>.
- Cockerill, Matthew (2006). Business models in open access publishing. W: Neil Jacobs (red.) *Open Access: key strategic, technical and economic aspects*. Oxford: Chandos Publ., s. 111-119.
- Conkling, Jeff (1988). Hypertext: an introductory survey. W: Irene Greif (ed.) *Computer - supported cooperative work : a book of readings*. San Mateo: Morgan Kaufman, s. 423-476.
- Coyle, Karen; Diane Hillmann [dok. elektr.] (2007). Resource Description and Access (RDA) : cataloguing rules for the 20th century. *D-Lib Magazine*, vol. 13, nr 1/2. <http://www.dlib.org/dlib/january07/coyle/01coyle.html>.
- Crane, Gregory [dok. elektr.] (2006). What do you do with a million books? *D-Lib Magazine*, vol. 12, nr 3. <http://www.dlib.org/dlib/march06/crane/03crane.html>.
- Crow, Ryam [dok. elektr.] (2006). Publishing cooperatives: an alternative for non-profit publishers. *First Monday*, vol. 11, nr 9. http://www.firstmonday.org/issues/issue11_9/crow/index.html.
- Cullen, Rowena (2003). Evaluating digital libraries in the health sector. Part 1, Measuring inputs and outputs. *Health Information and Libraries Journal*, vol. 20, nr 4, s. 195-204.
- Daconta, Michael (2007). Information as product : how to deliver the right information, to the right person at the right time. Denver, Co.: Outskirts Press.
- Daconta, Michael; Leo Obst; Kevin Smith (2003). *The Semantic Web : a guide to future of XML, Web Services and knowledge management*. Indianapolis: Wiley Publ.
- Dahl, Mark; Kyle Bancrjce; Michael Spalti (2006). *Digital libraries: integrating content and systems*. Oxford: Chandos Publ.
- Dale, Robin; Bruce Ambacher (2007). *Trustworthy repositories audit & certification: criteria and checklist*. Dublin: OCLC.
- Dankiewicz, Katarzyna [dok. elektr.] (2008). Index Copernicus. *Biuletyn EBIB*, nr 8(99). <http://www.ebib.info/2008/99/a.php?dankiewicz>.
- Das, Sudeshna [et al.] [dok. elektr.] (2009). Scientific publications on Web 3.0. <http://conferences.aepic.it/index.php/clpub/clpub2009/paper/view/149/60>.
- David, Paul [dok. elektr.] (2005). Towards a cyberinfrastructure for enhanced scientific collaboration : providing its 'soft' foundations may be hardest part. <http://www.oii.ox.ac.uk/resources/publications/RR4.pdf>.
- Davidson, Cathy (2009). The futures of scholarly publishing. W: Albert Greco (ed.) *The state of scholarly publishing : challenges and opportunities*. New Brunswick: Transaction Publ., s. 35-47.
- Davis, Hilary; John Vickery (2007). Datasets, a shift in the currency of scholarly communication: implications for library collections and acquisitions. *Serials Review*, vol. 33, nr 1, s. 26-32.
- Day, Michael [dok. elektr.] (2005). Metadata. W: S. Ross, M. Day (ed.) *DCC Digital curation manual*. <http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata/>.
- Day, Michael; Maureen Pennock; Julie Allinson [dok. elektr.] (2007). Co-operation for digital preservation and curation: collaboration for collection development in institutional repository networks. http://ils.unc.edu/digicurr2007/papers/dayPennock_paper_9-3.pdf.
- Delany, Paul; George Landow (2008). *Zarządzanie cyfrowym słowem; tekst w epoce reprodukcji elektronicznej*. W: Andrzej Gwóźdź (red.) *Ekran piśmiennosci*. Warszawa: Wydaw. Akad. i Profesjonalne, s. 75-101.
- Dempsey, Lorcan [dok. elektr.] (2009). Always on: libraries in a world of permanent connectivity. *First Monday*, vol. 14, nr 1. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printer-friendly/2291/2070>.
- Dempsey, Lorcan [dok. elektr.] (2006). Libraries and the Long Tail : some thoughts about libraries in a network age. *D-Lib Magazine*, vol. 12, nr 4. <http://www.dlib.org/dlib/april06/dempsey/04dempsey.html>.

- Derfert-Wolf, Lidia [dok. elektr.] (2006). Elektroniczne usługi informacyjne typu pytaniec-odpowiedź – światowe trendy i doświadczenia bibliotek. *Biuletyn EBIB*, nr 1(71). <http://www.ebib.info/2006/71/derfert.php>.
- Derfert-Wolf, Lidia; Marek Górski; Artur Jazdon [dok. elektr.] (2009). Finansowanie bibliotek uczelni publicznych. Analiza wybranych wskaźników z lat 2002-2007. *Biuletyn EBIB*, nr 8(108). http://www.ebib.info/2009/108/a.php?derfert_gorski_jazdon.
- Dervin, Brenda (1999). On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing and Management*, vol. 34, nr 6, s. 727-750.
- Dobratz, Susanne; Astrid Schoger; Stefan Strathmann [dok. elektr.] (2007). The nestor catalogue of criteria for trusted digital repository evaluation and certification. *Journal of Digital Information*, vol. 8, nr 2. <http://journals.tdl.org/jodi/article/viewArticle/199/180>.
- Dobrowolski, Zdzisław (2004). On the future of the Internet. W: Barbara Sosińska-Kalata, Katarzyna Materska, Wiesław Gliński (red.) *Spółeczeństwo informacyjne i jego technologie*. Warszawa: Wydaw. SBP, s. 67-77.
- Doerr, Martin [dok. elektr.] (2008). Ontologies. W: S. Ross, M. Day (ed.) *DCC Digital Curation Manual*. <http://www.dcc.ac.uk/resource/curation-manual/chapters/ontologies/>.
- Dren, Marek (2008). Czy Google stoi ponad prawem? *Chip*, nr 5, s. 22.
- Dublin Core Collection Description Task Group [dok. elektr.] (2007). Dublin Core Collections Application Profile. <http://dublincore.org/groups/collections/collection-application-profile/index.shtml>.
- Duguid, Paul [dok. elektr.] (2007). Inherence and loss? A brief survey of Google Books. *First Monday*, vol. 12, nr 8. http://firstmonday.org/issues/issue12_8/duguid/index.html.
- Duval, Erik [et al.] [dok. elektr.] (2002). Metadata principles and practicalities. *D-Lib Magazine*, vol. 8, nr 4. <http://www.dlib.org/dlib/april02/weibel/04weibel.html>.
- Edwards, Paul [et al.] (2007). Understanding infrastructure: dynamics, tensions and design. Report of a Workshop: History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructure. Ann Arbor 28.09-1.10. Ann Arbor: Univ. of Michigan, School of Information.
- Escarpit, Robert (1969). *Rewolucja książki*. Warszawa: PWN.
- Esposito, Joseph [dok. elektr.] (2008). Open Access 2.0: access to scholarly publications moves to a new phase. *Journal of Electronic Publishing*, vol. 11, nr 2. <http://hdl.handle.net/2027/spo.3336451.0011.203>.
- Estabrook, Leigh; Bijan Warner [dok. elektr.] (2003). The book as the gold standard for tenure and promotion in the humanistic disciplines. http://msc.mellon.org/research-reports/Book%20as%20the%20Gold%20Standard.pdf/at_download/file.
- Evans, James (2008). Electronic publication and the narrowing of science and scholarship. *Science*, vol. 321, nr 5887, s. 395-399.
- Evans, Woody (2008). Embryonic Web 3.0. *Searcher*, vol. 16, nr 1, s. 12-17.
- Eysenbach, Gunther [dok. elektr.] (2008). Medicine 2.0: social networking, collaboration, participation, apomediation and openness. *Journal of Medical Internet Research*, vol. 10, nr 3. <http://www.jmir.org/2008/3/e22/>.
- Fast, Karl; Kamran Sedig [dok. elektr.] (2005). The INVENT framework: examining the role of information visualization in the reconceptualization of digital libraries. *Journal of Digital Information*, vol. 6, nr 3. <http://jodi.tamu.edu/Articles/v06/i03/Fast/>.
- Feather, John (2003). *Communicating knowledge : publishing in the 21st century*. München: K.G. Saur.
- Filipiak, Marian (2004). *Homo Communicans. Wprowadzenie do teorii masowego komunikowania*. Lublin: UMCS.
- Firlej-Buzon, Aneta (2005). *Systemy zarządzania treścią – możliwości i wykorzystanie*. W: Maria Kocójowa (red.) *Profesjonalna informacja w Internecie*. Kraków: Wydaw. UJ, s. 162-165.
- Fisher, Julian [dok. elektr.] (2008). Scholarly publishing re-invented: real cost and real freedoms. *Journal of Electronic Publishing*, vol. 11, nr 2. <http://hdl.handle.net/2027/spo.3336451.0011.204>.
- Foster, Ian [dok. elektr.] (2003). The Grid: Computing without Bounds. *Scientific American*, April. <http://www.sciam.com/article.cfm?chanID=sa006&colID=1&articleID=000B1833-21DF-1E64-A98A809EC5880105>.
- Foster, Ian [dok. elektr.] (2002). What is the grid? A three point checklist. <http://www-fp.mcs.anl.gov/~foster/Articles/WhatsTheGrid.pdf>.

- Foster, Richard (1986). *Innovation: the attacker's advantage*. New York: Summit Books.
- Foucault, Michel (2005). Inne przestrzenie. *Teksty Drugie* nr 6/2005, s. 117-125.
- Fourie, Ina (2006). Learning from web information seeking studies: some suggestions for LIS practitioners. *The Electronic Library*, vol. 24, nr 1, s. 20-37.
- Fox, Edward [dok. elektr.] (2005). Digital libraries: archaeology, automation, EDTS, and enhancements. The International Advanced Digital Library Conference (IADLC). Nagoya Univ., August 25-26 <http://iadlc.nul.nagoya-u.ac.jp/archives/IADLC2005/FoxNagoya20051.pdf>.
- Fox, Edward; Neill Kipp; Paul Mather [dok. elektr.] (1996). Sets and functions : a foundation for digital libraries, though streams, spaces, structures, and scenarios (S4). <http://citescr.ist.psu.edu/cache/papers/cs/3122/>.
- Francke, Helena (2008). (Re)creations of scholarly journals : document and information architecture in Open Access journals. Boras: VALFRID.
- Franke, Jerzy (2007). *Googletheca Universalis? W: Jadwiga Woźniak-Kasperk, Jerzy Franke (red.) Biblioteki cyfrowe : projekty, realizacje, technologie*. Warszawa: Wydaw. SBP, s. 121-166.
- Frazier, Kenneth [dok. elektr.] (2001). The librarian's dilemma : contemplating the cost of the "Big Deal". *D-Lib Magazine*, vol. 7, nr 3. <http://www.dlib.org/dlib/march01/frazier/03frazier.html>.
- Freeman, Peter [dok. elektr.] (2007). Is „designing“ cyberinfrastructure – or, even, defining it – possible? *First Monday*, vol. 12, nr 6. http://firstmonday.org/issues/issue12_6/freeman/index.html.
- Frost, William (2004). Do we want or need metasearching? *Library Journal*, vol. 129, nr 6, s. 68.
- Fuhr, Norbert [et al.] (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, vol. 8, nr 1, s. 21-38.
- Gantz, John [et al.] [dok. elektr.] (2008). The diverse and exploding digital universe. <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.
- Garfield, Eugene [dok. elektr.] (1994). The Impact Factor. http://wzar.unizar.es/acad/fac/cie/quiorg/catalisis/isi_fi.html.
- Gargouri, Yassine [et al.] [dok. elektr.] (2010). Self-selected or mandated, Open Access increases citation impact for higher quality research. http://arxiv.org/PS_cache/arxiv/pdf/1001/1001.0361v1.pdf.
- Gartner, Richard (2008). *Metadata for digital libraries: state of the art and future directions*. Bristol: JISC.
- Garvey, William; Belver Griffith (1972). Communication and information processing within scientific disciplines: empirical findings for psychology. *Information Storage and Retrieval*, vol. 8, nr 3, s. 123-136.
- Gašević, Dragan; Dragan Djurić, Vladan Devedžić (2006). *Model driven architecture and ontology development*. Berlin; Heidelberg: Springer Verl.
- Gawrysiak, Piotr (2008a). *Cyfrowa rewolucja : rozwój cywilizacji informacyjnej*. Warszawa: PWN.
- Gawrysiak, Piotr (2008b). *Cyfrowe biblioteki a wyszukiwanie informacji. Przegląd Biblioteczny*, R. 76, z. 4, s. 652-661.
- Gawrysiak, Piotr (2008c). *Wolna informacja, wolna kultura*. W: Katarzyna Materska, Ewa Chuchro, Barbara Sosińska-Kalata (red.) *Organizowanie środowiska informacji i wiedzy*. Warszawa: Wydaw. SBP, s. 59-70.
- Gentil-Beccot, Anne; Salvatore Mele; Travis Brookes [dok. elektr.] (2009). Citing and reading behaviors in high-energy physics. <http://arxiv.org/ftp/arxiv/papers/0906/0906.5418.pdf>
- Glombowski, Karol; Helena Szwajkowska (1971). *Książka rękopiśmienna i biblioteka w starożytności i średniowieczu*. Warszawa: PWN.
- Gmiterek, Grzegorz [dok. elektr.] (2007). Library 2.0. Możliwości zastosowania Web 2.0 w bibliotekach polskich. *Biuletyn EBIB*, nr 4 (85). <http://www.ebib.info/2007/85/a.php?gmiterek>
- Goban-Klas, Tomasz (2005). Społeczeństwo masowe, informacyjne, sieciowe czy medialne? *Ethos*, vol. 18, nr 1-2, s. 100-114.
- Godlee, Fiona; Kay Dickerson (2003). Bias, subjectivity, chance and conflict of interest in editorial decisions. W: Fiona Godlee, Tom Jefferson (ed.) *Peer Review in Health Sciences*. London: BMJ Publ. Group, s. 91-117.
- Gold, Anna [dok. elektr.] (2007). Cyberinfrastructure, data and libraries : libraries and the data challenge: roles and actions for libraries. *D-Lib Magazine*, vol. 13, nr 9/10. <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>.
- Goldsmith, Beth; Frances Knudson [dok. elektr.] (2006). *Repository librarian and the next cru-*

- sadc. *D-Lib Magazine*, vol. 12, nr 9. <http://www.dlib.org/dlib/scptember06/goldsmith/09goldsmith.html>.
- Golka, Marian (2008). *Bariery w komunikowaniu i społeczeństwo (dez)informacyjne*. Warszawa: Wydaw. Naukowe PWN.
- Gołębiewski, Łukasz (2009). E-książka : szerokopasmowa kultura. Warszawa: Biblioteka Analiz.
- Gonçalves, Marcos [et al.] (2007). What is good digital library? – A quality model for digital libraries. *Information Processing and Management*, vol. 43, nr 5, s. 1416-1437.
- Gonçalves, Marcos [et al.] (2004). Streams, structures, spaces, scenarios, societies (5S) : a formal model for digital libraries. *ACM Transactions on Information Systems*, vol. 22, nr 2, s. 270-312.
- Górska, Małgorzata (2009). *Książki, nowe media i ich czasoprzestrzeń*. Warszawa: Wydaw. SBP.
- Górska, Małgorzata [dok. elektr.] (2007). Rewolucja Gutenberga – przełom cyfrowy. Próba porównania. <http://www.cbib.info/publikacje/matkonf/mat17/goralska.php>.
- Górny, Mirosław (2005). Biblioteka akademicka jako element globalnej cyfrowej infrastruktury informacyjnej. W: Halina Ganińska (red.) *Biblioteki naukowe w kulturze i cywilizacji: działania i codzienność*. Poznań: Bibl. Gł. Politechn. Poznańskiej, vol. 1, s. 52-58.
- Grabowska, Marta (2007). Biblioteka cyfrowa w środowisku wirtualnym : nowe wyzwanie dla katalogów bibliotek w erze dokumentów elektronicznych. W: Jadwiga Woźniak-Kasperk i Jerzy Franke (red.) *Biblioteki cyfrowe, projekty, realizacje, technologie*. Warszawa: Wydaw. SBP, s. 21-35.
- Greenberg, Jane (2007). Advancing the Semantic Web via library functions. *Cataloging & Classification Quarterly*, vol. 43, nr 3-4, s. 203-227.
- Greenberg, Jane; Kristina Spurgin; Abe Crystal (2005). *Final report for the AMeGA Project*. Washington: Library of Congress.
- Gresham, John (1994). From invisible college to cyberspace college: computer conferencing and the transformation of informal scholarly communication networks. *Interpersonal Computing and Technology*, vol. 2, nr 4, s. 37-52.
- Griffin, Stephen [dok. elektr.] (2005). Funding for digital libraries research. Past and present. *D-Lib Magazine*, vol. 11, nr 7/8. <http://www.dlib.org/dlib/july05/griffin07griffin.html>.
- Griffiths, José-Marie [dok. elektr.] (2000). Back to the future: information science for the new millennium. *Bulletin of ASIS*, vol. 26, nr 4. <http://www.asis.org/Bulletin/May-00/griffiths.html>.
- Gruber, Thomas (2008). Collective knowledge systems: where the Social Web meets the Semantic Web. *Journal of Web Semantics*, vol. 6, nr 1, s. 4-13.
- Guédon, Jean-Claude (2006). Open access: a symptom and a promise. W: Niel Jacobs (ed.) *Open Access: key strategic, technical and economic aspects*. Oxford: Chandos Publ., s. 27-38.
- Guédon, Jean-Claude [dok. elektr.] (2001). In Oldenburg's long shadow: librarians, research scientists, publishers and control of scientific publishing. *ARL Proceedings*, nr 138. <http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml>.
- Gulli, Antonio; Alessio Signorini (2005). The indexable Web is more than 11.5 billion pages. W: 14th Intern. WWW2005 Conference, 10-14 May 2005, Chiba, Japan. New York: ACM, s. 902-903.
- Habermas, Jürgen (1999). *Racjonalność działania a racjonalność społeczna*. Warszawa: PWN.
- Hahn, Karla [dok. elektr.] (2008). Talk about talking about new models of scholarly communication. *Journal of Electronic Publishing*, vol. 11, nr 1. <http://hdl.handle.net/2027/spo.3336451.0011.108>.
- Hakala, Juha (2006). The seven levels of identification. An overview of the current state of identifying objects within digital libraries. *Program*, vol. 40, nr 4, s. 361-371.
- Hakala, Juha (2004). Archiving the Web: European experiences. *Program*, vol. 38, nr 3, s. 176-183.
- Halasz, Frank [dok. elektr.] (1990). The Dexter hypertext reference model. <http://ei.cs.vt.edu/~mm/pdf/dexter.pdf>.
- Halliday, Leah [dok. elektr.] (2001). Scoping study for JISC Scholarly Communication Group. http://www.jisc.ac.uk/uploaded_documents/halliday.pdf.
- Hallmark, Julie [dok. elektr.] (2004). Access and retrieval of recent journal articles : a comparative study of chemists and geoscientists. *Issues in Science and Technology Librarianship*, nr 40. <http://www.istl.org/04-summer/article1.html>.
- Hardin, Garrett (1968). The tragedy of the commons. *Science*, nr 162, s. 1243-1248.

- Harnad, Stevan [dok. elektr.] (2001). Minotaur: six proposals for freeing the refereed literature online : a comparison. *Ariadne*, nr 28. <http://www.ariadne.ac.uk/issue28/minotaur/intro.html>.
- Harnad, Stevan (1998). Learned inquiry and the net: the role of peer review, peer commentary and copyright. *Learned Publishing*, vol. 11, nr 4, s. 183-192.
- Harnad, Stevan (1992). Post-Gutenberg galaxy : the fourth revolution in the means of production of knowledge. *PACS Review*, vol. 2, s. 39-53.
- Harter, Stephen [dok. elektr.] (1997). Scholarly communication and the digital library: problems and issues. *Journal of Digital Information*, vol.1, nr 1. <http://journals.tdl.org/jodi/article/view/jodi-3/4>.
- Hartley, Richard (2009). Folksonomies to ontologies: the changing nature of controlled vocabularies. W: Jillian Griffiths, Jenny Craven (ed.) *Access, delivery, performance the future of libraries without walls*. London: Facet Publ., s. 145-157.
- Harvey, Ross [dok. elektr.] (2007). Appraisal and selection. W: S. Ross, M. Day (ed.) *DCC Digital Curation Manual*. <http://www.dcc.ac.uk/resource/curation-manual/chapters/appraisal-and-selection>.
- Heery, Rachel; Manjula Patel [dok. elektr.] (2000). Application profiles: mixing and matching meta-data schemas. *Ariadne*, nr 25. <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>.
- Henderson, Albert (2002). The growth of printed literature in the twentieth century. Richard Abel, Lyman Newlin (ed.). *Scholarly publishing: books, journals, publishers and libraries in the twentieth century*. New York: Wiley, s. 9-11.
- Herrera, Gail (2007). MetaSearching and beyond: implementation experiences and advice from an academic library. *Information Technology and Libraries*, vol. 26, nr 2, s. 44-52.
- Hess, Charlotte; Elinor Ostrom (2006). Introduction: an overview of the knowledge commons. W: Charlotte Hess, Elinor Ostrom (ed.) *Understanding knowledge as a commons : from theory to practice*. Cambridge: MIT Press, s. 3-26.
- Higgins, Sarah (2007). Draft DCC curation lifecycle model. *International Journal of Digital Curation*, vol. 2, nr 2, s. 82-87.
- Hildreth, Charles [dok. elektr.] (2000). Online catalog design models: are we moving in the right direction? <http://www.ou.edu/faculty/H/Charles.R.Hildreth/clrthree.html>.
- Hine, Christine (2006). Computerization movements and scientific disciplines: the reflexive potential of new technologies. W: Christine Hine (ed.) *New infrastructures for knowledge production: understanding e-Science*. Hershey: Information Science Publ., s. 26-47.
- Hirsch, Jorge (2005). An index to qualify an individual's scientific research output. *Proceedings of the National Acad. of Sciences*, vol. 102, nr 46, s. 16569-16572.
- Hofmokl, Justyna (2009). Internet jako nowe dobro wspólne. Warszawa: Wydaw. Akad. i Profesjonalne.
- Hofmokl, Justyna; Alek Tarkowski [dok. elektr.] (2005). Creative Commons, innowacje w systemie prawa autorskiego i rozwój wolnej kultury. *Biuletyn EBIB*, nr 3 (64). http://ebib.oss.wroc.pl/2005/64/hofmokl_tarkowski.php.
- Holmberg, Kim [et al.] (2009). What is Library 2.0? *Journal of Documentation*, vol. 65, nr 4, s. 668-681.
- Holmström, Jonas [dok. elektr.] (2004). Report on the 8th European Conference on Digital Libraries (ECDL 2004). *D-Lib Magazine*, vol. 10, no 10. <http://www.dlib.org/dlib/october04/holmstrom/10holmstrom.html>.
- Holz, Dayna (2006). Technologically enhanced archival collections: using the buddy system. *Journal of Archival Organization*, vol. 4, nr 1/2, s. 29-44.
- Horgan, John (1996). The end of science: facing the limits of science in the twilight of the scientific age. Reading, Mass.: Addison-Wesley Publ., 322 s.
- Howorka, Bolesław (2009). Udośćpnianie w bibliotece cyfrowej. *Forum Akademickie*, nr 7/8, s. 44-46.
- Hull, Duncan; Steve Pettifer; Douglas Kell [dok. elektr.] (2008). Defrosting the digital library: bibliographic tools for the next generation Web. *PLoS Computational Biology*, vol. 4, nr 10. doi:10.1371/journal.pcbi.1000204.
- Hunter, Jane (2006). Scientific Publication Packages – a selective approach to the communication and archival of scientific output. *Intern. Journal of Digital Curation*, vol. 1, nr 1, s. 33-52.
- Huuhka, Petteri [dok. elektr.] (2006). Google: data structures and algorithms. http://www.cs.helsinki.fi/u/hinden/teaching/irr06/papers/petteri_huuhka_google_paper.pdf.
- lacono, Suzanne; Rob Kling (2001). Computerization movements: the rise of the Internet and distant

- forms of work. W: Joanne Yates, John Van Maanen (ed.) *Information technology and organizational transformation: history, rhetoric, and practice*. Thousand Oaks: Sage Publ. s. 93-136.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional Requirements for Bibliographic Record* : Final Report. München: Saur.
- Isfandyari-Moghaddam, Alireza; Behrooz Bayat (2008). Digital libraries in the mirror of the literature: issues and considerations. *The Electronic Library*, vol. 26, nr 6, s. 844-862.
- Jackson, Steven [et al.] [Dok. elektr.] (2007). Understanding infrastructure: history, heuristics, and cyberinfrastructure policy. *First Monday*, vol. 12, nr 6. http://firstmonday.org/issues/issue12_6/jackson/index.html.
- Janiak, Małgorzata (2007). Grid – dynamiczny dostęp do zasobów. W: Maria Próchnicka, Agnieszka Korycińska-Huras (red.) *Między przeszłością a przyszłością : książka, biblioteka, informacja naukowa – funkcje społeczne na przestrzeni wieków*. Kraków: Wydaw. UJ, s. 226-232.
- Jantz, Ronald; Michael Giarlo (2006). Digital archiving and preservation: technologies and process for a trusted repository. *Journal of Archival Organization*, vol. 4, nr 1/2, s. 193-213.
- Jaskowska, Bożena (2009). Efektywny marketing SEO witryny bibliotecznej w wyszukiwarkach. *Przegląd Biblioteczny*, R. 77, z. 2, s. 229-238.
- Jaskowska, Bożena [dok. elektr.] (2008). O kulturze konwergencji słów kilka. *Biuletyn EBIB*, nr 1 (92). <http://www.ebib.info/2008/92/a.php?jaskowska>.
- Jaskowska, Małgorzata; Agnieszka Korycińska-Huras; Maria Próchnicka (2009). Bibliografia jako źródło informacji naukowej – w opinii studentów. W: Jadwiga Woźniak-Kasparek, Mikołaj Ochmański (red.) *Bibliografia*. Warszawa: Wydaw. SBP, s. 233-259.
- Jenkins, Henry (2007). *Kultura konwergencji: zderzenie starych i nowych mediów*. Warszawa: Wydaw. Akad. i Profesjonalne.
- Johnson, Richard (2007). In Google's broad wake: taking responsibility for shaping the global digital library. *Bimonthly Report on Research Library Issues and Actions*, nr 250. <http://www.arl.org/bm~doc/arlbr250digprinciples.pdf>.
- Johnson, Ben; Charles Oppenheim (2007). How socially connected are citers to those that they cite? *Journal of Documentation*, vol. 63, nr 5, s. 609-637.
- Jose, Antony (2007). Evaluation of digital libraries: a case study. W: ARD Prasad, Devika Madalli (ed.) *ISDC 2007, Proc. of 11th Intern. Conf. on Semantic Web and Digital Libraries*, Bangalore, 21-23 February. Karnataka: DRTC, s. 229-238.
- Juza, Marta (2007). Wiedza ekspercka a mądrość zbiorowa w komunikacji internetowej. *Studia Socjologiczne*, nr 3(186), s. 37-58.
- Kahin, Brian [dok. elektr.] (2007). Cyberinfrastructure and innovation policy. *First Monday*, vol. 12, nr 6. http://www.firstmonday.org/issues/issue12_6/kahin/index.html.
- Kahn, Robert; Robert Wilensky [dok. elektr.] (1995). A framework for distributed digital object services. <http://dx.doi.org/cnri.dlib/tn95-01>.
- Kajewski, Mary Ann (2007). Emerging technologies changing our service delivery models. *The Electronic Library*, vol. 25, nr 4, s. 420-429.
- Kalfoglou, Yannis (2001). Exploring ontologies. W: S. K. Chang (Ed.) *Handbook of software engineering and knowledge engineering*. Singapore: World Scientific Publ., s. 863-887.
- Kalinichenko, Leonid [et al.] (2000). Infrastructure of the subject mediating environment aiming at semantic interoperability of heterogeneous digital library collections. W: 2nd Russian Scientific Conf. Digital libraries advanced methods and technologies, digital collections. 26-28 Sept. Protno: Russian Acad. of Sciences, s. 78-90.
- Keen, Andrew (2007). *Kult amatora : jak internet niszczy kulturę*. Warszawa: Wydaw. Akad. i Profesjonalne.
- Kellogg, David (2006). Toward a post-academic science policy: scientific communication and the collapse of the mertonian norms. *Int. Journal of Communications Law and Policy*, Special Issue, Autumn, s. 1-29.
- King, Donald; Frances Alvarado-Albertorio (2008). Pricing and other means of charging for scholarly journals. *Learned Publishing*, vol. 21, nr 4, s. 248-272.
- King, Donald; Carol Montgomery [dok. elektr.] (2002). After migration to an electronic journal collection: impact on faculty and students. *D-Lib Magazine*, vol. 8, nr 12. <http://www.dlib.org/dlib/december02/king/12king.html>.
- King, Kenneth [dok. elektr.] (2004). Creating a sustainable scholarly communication system. http://ecommons.library.cornell.edu/bitstream/1813/179/2/King_27May04.pdf.
- Kiorgaard, Deirdre; Ebe Kartus [dok. elektr.] (2006). A rose by any other name?: from AACR2 to

- Resource Description and Access. <http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1169/1437>.
- Kisilowska, Małgorzata (2009). Modelowanie rozległych systemów informacyjnych : zdrowie i kultura. Warszawa: Wydaw. SBP.
- Klincewicz, Krzysztof (2008). Polska innowacyjność : analiza bibliometryczna. Warszawa: Wydaw. Nauk. Wydz. Zarządzania Univ. Warszawskiego.
- Kling, Rob [dok. elektr.] (2003). The Internet and unrefereed scholarly publishing. <http://rkcsi.indiana.edu/archive/CSI/WP/WP03-01B.html>.
- Kling, Rob; Lisa Spector; Geoffrey McKim [dok. elektr.] (2002). Locally controlled scholarly publishing via the Internet: the guild model. *Journal of Electronic Publishing*, vol. 8, nr 1. <http://www.press.umich.edu/jep/08-01/kling.html>.
- Kodali, Naren; Csilla Farkas; Duminda Wijesekera (2004). An authorization model for multimedia digital libraries. *International Journal on Digital Libraries*, vol. 4, nr 3, s. 139-155.
- Koehler, Wallace [dok. elektr.] (2006). Public libraries as international repositories and stewards in an historical and ethical context. W: World Library and Information Congress: 72nd IFLA General Conference and Council. 20-24 August, Seoul, Korea. <http://www.ifla.org/IV/ifla72/papers/151-Koehler-en.pdf>.
- Koehler, Wallace [dok. elektr.] (1999). Digital libraries and World Wide Web sites and page persistence. *Information Research*, vol. 4, nr 4. <http://informationr.net/ir/4-4/paper60.html>.
- Kolb, David (2009). Other spaces for spatial hypertext. *Journal of Digital Information*, vol. 10, nr 3. <http://journals.tdl.org/jodi/article/viewArticle/171/485>.
- Konik, Roman (2009). Wirtualność jako rehabilitacja iluzji. *Diametros*, nr 21, s. 78-95.
- Koszowska, Agnieszka (2009). Biblioteka 2.0 – usługi biblioteczne z wykorzystaniem technologii i narzędzi Web 2.0. W: Nowe technologie w bibliotekach publicznych. Materiały z VIII ogólnopolskiej konferencji „Automatyzacja bibliotek publicznych”. Warszawa, 26-28 listopada 2008 r. Warszawa: Wydaw. SBP, s. 75-86.
- Kousha, Kayvan; Mike Thelwall (2007). Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis. *Journal of ASIST*, vol. 58, nr 7, s. 1055-1065.
- Krzysztofek, Kazimierz (2008). Magiczny trójkąt: między techniką, logiką i kulturą. W: Piotr Gawrysiak. Cyfrowa rewolucja : rozwój cywilizacji informacyjnej. Warszawa: PWN, s. 7-16.
- Krzysztofek, Kazimierz (2007). WEBSki świat: mądrość tłumów sieciowych czy zbiorowe nieucucie? W: Andrew Keen. Kult amatora : jak Internet niszczy kulturę. Warszawa: Wydaw. Akad. i Profesjonalne, s. 11-23.
- Krzysztofek, Kazimierz (2005). Między kulturą a antropotechnologią. Wzrę społeczeństwa informacyjnego. *Ethos*, vol. 18, nr 1-2, s. 187-204.
- Kuhn, Thomas (2001). Struktura rewolucji naukowych. Warszawa: Fundacja Aletheia.
- Kuny, Terry [dok. elektr.] (1997). A digital dark ages? Challenges in the preservation of electronic information. <http://www.ifla.org/IV/ifla63/63kuny1.pdf>
- Kyrillidou, Martha; Mark Young (2006). ARL statistics 2004-05. Washington: ARL.
- Lack, Rosalie (2006). The importance of user-centered design: exploring findings and methods. *Journal of Archival Organization*, vol. 4, nr 1/2, s. 69-86.
- Lagoze, Carl (2004). Bridging the Past and Future: Scholarly Communication in the 21st Century. W: DLKC'04 International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society. Tsukuba, 2-5 March. Tsukuba: Univ. of Tsukuba, s. 88-96.
- Lagoze, Carl; David Fielding; Sandra Payette (1998). Making global digital libraries work: collection services, connectivity regions and collection views. W: Ian Witten, Rob Akseyn, Frank M. Shipman (Red.) Proceedings of the 3rd ACM Conference on Digital Libraries (DL-98). Pittsburgh, 23-26.06.1998. New York: ACM, s. 134-143.
- Lakoff, George; Mark Johnson (1988). Metafory w naszym życiu. Warszawa: PIW
- Lancaster, Frederick (2003). Do indexing and abstracting have a future? *Anales de Documentación*, nr 6, s. 137-144.
- Landow, George (2006). Hypertext 3.0 : critical theory and new media in era of globalization. Baltimore: The Johns Hopkins Univ. Press.
- Lankes, David [et al.] [dok. elektr.] (2007). Participatory networks: the library as conversation. *Information Research*, vol. 12, nr 4. <http://informationr.net/ir/12-4/colis/colis05.html>
- Larsen, Ronald; Howard Wactlar (2003). Knowledge lost in information : Report of the NSF Workshop on Research Directions for Digital Libraries. Chatham: NSF.

- Lassila, Ora [dok. elektr.] (1997). Introduction to RDF metadata : W3C note 1997-11-13. <http://www.w3.org/TR/NOTE-rdf-simple-intro>.
- Lavoie, Brian; Lorcan Dempsey [dok. elektr.] (2009). Beyond 1923 : Characteristics of potentially in-copyright print books in library collections. *D-Lib Magazine*, vol. 15, nr 11/12. <http://www.dlib.org/dlib/november09/lavoie/11lavoie.html>.
- Lavoie, Brian; Lorcan Dempsey [dok. elektr.] (2004). Thirteen ways of looking at... digital preservation. *D-Lib Magazine*, vol. 10, nr 7/8. <http://www.dlib.org/dlib/july04/lavoie/07lavoie.html>.
- Lawrence, Steve; Kurt Bollacker; Lee Giles (1999). Indexing and retrieval of scientific Literature. W: 8th international Conference on Information and Knowledge Management CIKM 99, Kansas City, November 2-6. New York: ACM, s. 139-146.
- Lesk, Michael (2005). Understanding digital libraries. Amsterdam: Elsevier.
- Lessig, Lawrence (2005). *Wolna kultura*. Warszawa: WSiP.
- LeVan, Ralph (2006). OpenSearch and SRU: a continuum of searching. *Information Technology and Libraries*, vol. 25, nr 3, s. 151-153.
- Levy, David; Catherine Marshall (1995). Going digital: a look at assumptions underlying digital libraries. *Communications of ACM*, vol. 38, nr 4, s. 77-84.
- Lévy, Pierre (1996). *Second flood : report on cyberculture*. Bruxelles: Council of Europe.
- Lewandowski, Dirk (2008). The retrieval effectiveness of web search engines: considering results descriptions. *Journal of Documentation*, vol. 64, nr 6, s. 915-937.
- Lewandowski, Dirk (2005). Web searching, search engines and Information Retrieval. *Information Services & Use*, vol. 25, nr 3/4, s. 137-147.
- Leydesdorff, Loet (2001). The complex dynamics of scientific communication. W: Colin Grant, Donald McLaughlin (ed.) *Language – meaning – social construction : interdisciplinary studies*. Amsterdam; New York: Rodopi, s. 163-179.
- Leydesdorff, Loet (1998). Theories of citation? *Scientometrics*, vol. 43, nr 1, s. 5-25.
- LIBER [dok. elektr.] (2009). LIBER position statement for the European Commission's hearing on the Google Book US Settlement Agreement. http://www.libereurope.eu/files/LIBER_Position_Statement_Google_Settlement.pdf.
- Li, Charlene; Josh Bernoff (2008). *Groundswell: winning in a world transformed by social technologies*. Boston: Harvard Business Press.
- Library of Congress WG on the Future of Bibliographic Control [dok. elektr.] (2008). On the record. <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.
- Licklider, Joseph; Robert W. Taylor [dok. elektr.] (1968). The computer as a communication device. [repr. of *Science and Technology*, April]. <http://memex.org/licklider.pdf>.
- Liew, Chern Li (2009). Digital library research 1997-2009 : organisational and people issues. *Journal of Documentation*, vol. 65, nr 2, s. 245-266.
- Linde, Peter (2006). Introduction to digital libraries – Memex of the future. W: Klaus Kreulich [et al.] (ed) *Integrating Technology and Culture*. 10th Intern. Conference on Electronic Publishing ELPUB. Supplement to the proceedings. Bansko, 14-16 June 2006. Sofia: POI-Commerce, s. 15-32.
- Liu, Ying; Anita Coleman [dok. elektr.] (2004). Communicating digital library services to scientific communities. *Libres*, vol. 14, nr 1. <http://libres.curtin.edu.au/libres14n1/>.
- Liu, Ziming (2003). Trends in transforming scholarly communication and their implications. *Information Processing and Management*, vol. 39, s. 889-898.
- Lossau, Norbert [dok. elektr.] (2004). Search engine technology and digital libraries : libraries need to discover the academic Internet. *D-Lib Magazine*, vol. 10, nr 6. <http://www.dlib.org/dlib/june04/lossau/06lossau.html>.
- Lourdi, Irene; Christos Papatheodorou, Martin Doerr [dok. elektr.] (2009). Semantic integration of collection description. *D-Lib Magazine*, vol. 15, nr 7/8. <http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>.
- Lyman, Peter (1996). What is digital library? Technology, intellectual property, and the public interest. *Daedalus*, vol. 125, nr 4, s. 1-33.
- Lyman, Peter; Hal Varian [dok. elektr.] (2003). How much information 2003? http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf.
- Lynch, Clifford (2006a). Open computation: beyond human reader-centric views of scholarly literatures. W: Neil Jacobs (ed.) *Open Access: key strategic, technical and economic aspects*. Oxford: Chandos Publ. s. 185-193.

- Lynch, Clifford [dok. elektr.] (2006b). Research libraries engage the digital world: a US-UK comparative examination of recent history and future prospects. *Ariadne*, nr 46. <http://www.ariadne.ac.uk/issue46/lynch/>.
- Lynch, Clifford (2000). Authenticity and integrity in the digital environment: an exploratory analysis of the central role of trust. W: *Authenticity in a digital environment*. Washington: CLIR, s. 32-50.
- Lynch, Clifford [dok. elektr.] (1999). Canonicalization: a fundamental tool to facilitate preservation and management of digital information. *D-Lib Magazine*, vol. 5, nr 9. <http://www.dlib.org/dlib/september99/09lynch.html>.
- Ma, Jin (2007). Metadata SPEC Kit 298. Washington: ARL.
- Mabe, Michael (2009). Scholarly publishing. *European Review*, vol. 17, nr 1, s. 3-22.
- Mabe, Michael; Mayur Amin (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, vol. 51, nr 1, s. 147-162.
- Machlup, Fritz (1983). Epilogue: semantic quirks in studies of information. W: Fritz Machlup, Una Mansfeld (ed.). *The study of information : interdisciplinary message*. New York: Wiley, s. 641-671.
- Madras, Giridhar (2008). Scientific publishing: rising cost of monopolies. *Current Science*, vol. 95, nr 2, s. 163-164.
- Mahner, Martin; Mario Bunge (1997). *Foundations of biophilosophy*. Berlin; New York: Springer Verl.
- Malak, Piotr (2007). Interfejs użytkownika katalogu bibliotecznego online – propozycja usprawnienia. *Przegląd Biblioteczny*, R. 75, z. 2, s. 252-260.
- Mansourian, Yazdan (2008). Web search efficacy: definition and implementation. *ASLIB Proceedings*, vol. 60, nr 4, s. 349-363.
- Mansourian, Yazdan [dok. elektr.] (2004). Similarities and differences between Web search procedure and searching in the pre-web information retrieval systems. *Webology*, vol. 1, nr 1. <http://www.webology.ir/2004/v1n1/a3.html>.
- Marchionini, Gary (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, vol. 49, nr 4, s. 41-46.
- Marchionini, Gary (1999). Augmenting library services: toward the sharium. W: *Proceedings of the Intern. Symposium on Digital Libraries ISDL'99*. Tsukuba, 28-29 September, Tsukuba: Univ. of Library and Inf. Science, s. 40-47.
- Marchionini, Gary; Catherine Plaisant; Anita Komlodi (2003). The people in digital libraries : multifaceted approaches to assessing needs and impact. W: Ann Bishop, Nancy Van House, Barbara Buttenfield (ed.). *Digital library use: social practice in design and evaluation*. Cambridge: MIT Press, s. 119-160.
- Marlino, Mary; Tamara Sumner (2008). From the library to the laboratory: a new future for science librarian. W: Richard Katz (ed.) *The tower and the cloud : higher education in the age of cloud computing*. Washington: Educase, s. 190-196.
- Marshall, Catherine (2005). Reading and interactivity in the digital library: creating an experience that transcends paper. W: Deanna Marcum, Gerald George (ed.) *Digital library development: the view from Kanazawa*. Westport: Libraries Unlimited, s. 127-145.
- Mason, Jon (2009). Knowledge management and Dublin Core. *Proc. of Intern. Conference on Dublin Core and Metadata Applications 2009*, s. 41-50.
- Mason, Robert; Tabitha Hart [dok. elektr.] (2007). Libraries for a global networked world: toward new educational and design strategies. http://www.ifla.org/IV/ifla73/papers/158-Mason_Hart-en.pdf.
- Materska, Katarzyna (2008). Modelowe koncepcje informacji naukowej (information science) na początku XXI wieku. W: Katarzyna Materska, Ewa Chuchro, Barbara Sosińska-Kalata (red.) *Organizowanie środowiska informacji i wiedzy*. Warszawa: Wydaw. SBP, s. 19-40.
- Materska, Katarzyna (2007). *Informacja w organizacji społeczeństwa wiedzy*. Warszawa: Wydaw. SBP.
- Matlak, Andrzej (2001). Dyrektywa Unii Europejskiej dotycząca koordynacji pewnych aspektów prawa autorskiego i praw pokrewnych w społeczeństwie informacyjnym. *Prace z Wynalazczości i Ochrony Własności Intelektualnej : zeszyty naukowe Uniwersytetu Jagiellońskiego*, z. 78, s. 161-185.
- Mayer, Rudolf; Andreas Rauber [dok. elektr.] (2009). Establishing context of digital objects' creation, content and usage. <http://cs.harding.edu/indp/papers/mayer9.pdf>.

- Mayr, Philipp; Anne-Kathrin Walter (2007). An exploratory study of Google Scholar. *Online Information Review*, vol. 36 nr 6, s. 814-830.
- Mazurek, Jolanta [dok. elektr.] (2009). Drogi czytelniku, jak korzystasz z biblioteki cyfrowej? Próba odpowiedzi. *Biuletyn EBIB*, Nr 5 (105). <http://www.ebib.info/2009/105/a.php?mazurek>.
- McCabe, Mark (2002). Journal pricing and mergers: a portfolio approach. *The American Economic Review*, vol. 92, nr 1, s. 259-269.
- McDonough, Jerome (2006). METS: standardized encoding for digital library objects. *International Journal on Digital Libraries*, vol. 6, nr 2, s. 148-158.
- McGregor, Bruce [dok. elektr.] (2008). Facets and hierarchies in scientific search. *Journal of Electronic Publishing*, vol. 11, nr 2. <http://hdl.handle.net/2027/spo.3336451.0011.205>.
- McGuinness, Deborah (2005). Ontologies come of age. W: Dieter Fensel (ed.) *Spinning the Semantic Web : bringing the World Wide Web to its full potential*. Cambridge, Ma: MIT Press, s. 171-195.
- McLuhan, Marshall (2004). *Zrozumieć media : przedłużenia człowieka*. Warszawa: WN-T.
- Meadows, Arthur (1998). *Communicating research*. San Diego: Academic Press.
- Meho, Lokman (2007). The rise and rise of citation analysis. *Physics World*, vol. 20, nr 1, s. 29-31.
- Meho, Loknam; Kiduk Yang (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of ASIST*, vol. 58, nr 13, s. 2105-2125.
- Méndez, Eva [et al.] (2008). DCMF: DC & microformats, a good marriage. W: Jane Greenberg, Wolfgang Klas (red.) *Metadata for semantic and social applications : Proc. of Internat. Conf. on Dublin Core and Metadata Applications*. Berlin, 22-26 Sept. 2008. Göttingen: Universitätsverlag, s. 141-145.
- Merton, Robert (2002). *Teoria socjologiczna i struktura społeczna*. Warszawa: Wydaw. PWN.
- Metzger, Miriam (2007). Making sense of credibility on the Web: models for evaluating online information and recommendations for future research. *Journal of ASIST*, vol. 58, nr 13, s. 2078-2091.
- Meyer, Eric; Ralph Schroeder [dok. elektr.] (2008). The World Wide Web of research and access to knowledge. <http://ssrn.com/abstract=1153922>.
- Meyer, Philip (2004). *The vanishing newspaper : saving journalism in the information age*. Columbia: Univ. of Missouri Press.
- Mikułowski-Pomorski, Jerzy (2006). *Komunikacja wobec procesów fragmentaryzacji*. W: Lesław Haber, Marian Niezgodą (red.) *Spółeczeństwo informacyjne : aspekty funkcjonalne i dysfunkcjonalne*. Kraków: Wydaw. UJ, s. 93-104.
- Miller, Cass; Julianna Harris (2009). *Conflicting agendas for scholars, publishers and institutions*. W: Albert Greco (ed.) *The state of scholarly publishing : challenges and opportunities*. New Brunswick: Transaction Publ., s. 9-25.
- Miller, Eric [dok. elektr.] (1998). An introduction to the Resource Description Framework. *D-Lib Magazine*, May. <http://www.dlib.org/dlib/may98/miller/05miller.html>.
- Miller, Eric; Diane Hillman (2002). *Libraries and the future of the Semantic Web: RDF, XML, and alphabet soup*. W: Wayne Jones, Judith Ahronheim, Josephine Crawford (red.) *Cataloging the Web : metadata, AACR, and MARC 21*. Lanham: The Scarecrow Press.
- Miller, Paul [dok. elektr.] (2000). Interoperability : what it is and why should I want it? *Ariadne*, vol. 24. <http://www.ariadne.ac.uk/issue24/interoperability>.
- Milstread, Jessica; Susan Feldman [dok. elektr.] (1999). Metadata: cataloging by any other name... *Online January 1999*, s. 25-31. <http://www.onlineinc.com/onlinemag/OL1999/milstread1.html>.
- Ministerstwo Nauki i Informatyzacji [dok. elektr.] (2003). *Strategia informatyzacji Rzeczypospolitej Polskiej – ePolska na lata 2004-2006*. <http://www.regiony.nck.pl/download.php?id=75>.
- Mischo, William [dok. elektr.] (2005). Digital libraries : challenges and influential work. *D-Lib Magazine*, vol. 11, nr 7/8. <http://www.dlib.org/dlib/july05/mischo/07mischo.html>.
- Monastersky, Richard [dok. elektr.] (2006). Liczba, która paraliżuje naukę. Impact factor – kiedyś prosta metoda tworzenia rankingów czasopism naukowych, dziś bezwzględne kryterium zatrudniania i przyznawania grantów. *Biuletyn EBIB*, Nr 3(73). <http://www.ebib.info/2006/73/monastersky.php>.
- Morbitz, Janusz [dok. elektr.] (2002). Świat wartości w Internecie. http://www.tuo.agh.edu.pl/Aksjol_Int_AGH_JMorbitz.pdf.

- Morbitzer, Janusz [dok. elektr.] (2000). Od homo sapiens do homo computerus. *Konспект*, nr 4. <http://www.wsp.krakow.pl/konспект/konспект4/homo.html>.
- Morrow, Terry [et al.] (2008). A comparative study of e-journal archiving solutions : a JISC funded investigation. Bristol: JISC.
- Muraszkiewicz, Mieczysław (2005). Esej: nowy paradygmat, czyli od systemu do sieci. W: Barbara Sosińska-Kalata, Maria Przystek-Samokowa (red.) *Od informacji naukowej do technologii społeczeństwa informacyjnego*. Warszawa: Wydaw. SBP, s. 83-86.
- Muraszkiewicz, Mieczysław (2002). Wybrane zagadnienia systemów baz danych. W: Janusz Grabara, Jerzy Nowak (red.) *Efektywność zastosowań systemów informacyjnych*. Warszawa: WN-T, s. 25-34.
- Muraszkiewicz, Mieczysław; Bruno Jacobfeuerborn (2008). An essay on modernisation, innovativeness, informational processes integration, and mobility in high-tech enterprises. W: Katarzyna Materska, Ewa Chuchro, Barbara Sosińska-Kalata (red.) *Organizowanie środowiska informacji i wiedzy*. Warszawa: Wydaw. SBP, s. 11-18.
- Nahotko, Małgorzata; Marek Nahotko [dok. elektr.] (2009). Stopień cyfryzacji polskich czasopism naukowych. W: Maria Kocójowa (ed.) *Biblioteki i ich klienci : między płatnym a bezpłatnym komunikowaniem się w erze zasobów cyfrowych i sieci [CD-ROM]*. Kraków: IINiB UJ, s. 106-111.
- Nahotko, Marek (2007a). Cyfrowa nauka – cyfrowe publikacje – cyfrowe biblioteki. *Przegląd Biblioteczny*, R. 75, z 1, s. 7-28.
- Nahotko, Marek (2007b). *Naukowe czasopisma elektroniczne*. Warszawa: Wydaw. SBP.
- Nahotko, Marek [dok. elektr.] (2007c). Wpływ Open Access na poziom cytowań publikacji. <http://www.ebib.info/publikacje/matkonf/mat18/nahotko.php>.
- Nahotko, Marek (2006a). Bibliotekarze cyfrowi w środowisku cyfrowej nauki, biblioteki i cyfrowych publikacji. *Biuletyn EBIB*, nr 10(80). <http://www.ebib.info/2006/80/a.php?nahotko>.
- Nahotko, Marek (2006b). Opis dokumentów elektronicznych : teoretyczny model i możliwości jego aplikacji. Kraków: Wydaw. UJ.
- Nahotko, Marek (2004). *Metadane : sposób na uporządkowanie Internetu*. Kraków: Wydaw. UJ.
- National Science Foundation (2007). *Cyberinfrastructure vision for 21st century discovery*. Arlington: NSF.
- Neelameghan, A.; Jacques Tocatlian (1985). International cooperation in information systems and services. *Journal of the ASIS*, vol. 36, nr 3, s. 153-163.
- Negroponce, Nicholas (1996). *Being digital*. New York: Vintage Books.
- Nelson, Theodor (1993). *Literary machines*. Sausalito: Mindful Press.
- Nelson, Theodor; Robert Smith (2007) [dok. elektr.]. Back to the future: hypertext the way it used to be. <http://xanadu.com/XanaduSpace/btf.htm>.
- Nentwich, Michael (2003). *Cyberscience : research in the age of the Internet*. Vienna: Austrian Acad. of Sciences Press.
- Neylon, Cameron; Shirley Wu [dok. elektr.] (2009). Article-level metrics and the evolution of scientific impact. *PLoS Biology*, vol. 7, nr 11. doi: 10.1371/journal.pbio.1000242.
- Nicholas, David [et al.] (2008a). The digital consumer: an introduction and philosophy. W: David Nicholas, Ian Rowlands (ed.) *Digital consumers: reshaping the information professions*. London: Facet Publ., s. 1-12.
- Nicholas, David [et al.] (2008b). The information-seeking behaviour of the digital consumer : case study – the virtual scholar. W: David Nicholas, Ian Rowlands (ed.) *Digital consumers: reshaping the information professions*. London: Facet Publ., s. 113-158.
- Nicholas, David; Paul Huntington; Hamid Jamali (2007). Open Access in context : a user study. *Journal of Documentation*, vol. 63, nr 6, s. 853-878.
- Nicholson, Scott (2004). A conceptual framework for the holistic measurement and cumulative evaluation of library services. *Journal of Documentation*, vol. 60, Nr 2, s. 164-182.
- Nielsen, Jakob (1993). *Usability engineering*. San Diego: Academic Press.
- Nielsen, Jakob; Hoa Lorangen (2007). *Optymalizacja funkcjonalności serwisów internetowych*. Gliwice: Helion.
- Niezgoda, Marian (2006). E-styl życia w społeczeństwie informacyjnym. W: Lesław Haber, Marian Niezgoda (red.) *Spółczesność informacyjna : aspekty funkcjonalne i dysfunkcjonalne*. Kraków: Wydaw. UJ, s. 105-115.
- Nilson, Mikael [et al.] [dok. elektr.] (2008). Expressing Dublin Core metadata using the Resource Description Framework (RDF). <http://dublincore.org/documents/2008/01/14/dc-rdf/>.

- NISO Web Services and Practices Working Group (2006). Best practices for designing Web services in the library context. Bethesda: NISO Press.
- Nolan, Mark (2008). Exploring exploratory search. *Bulletin of ASIST*, vol. 34, nr 4, s. 38-41.
- Noonan, Christine; Melissa McBurney (2007). Application of electronic serial usage statistics in a National Laboratory. W: David Fowler (ed.) Usage statistics of e-serials. Binghamton: The Haworth Press 2007, s. 151-160.
- Nowak, Piotr (2006). Bibliometria, webometria : podstawy , wybrane zastosowania. Poznań: Wydaw. Naukowe UAM.
- Nunberg, Geoffrey [dok. elektr.] (2009). Google's Book Search: a disaster for scholars. *The Chronicle of Higher Education*, August 31. <http://chronicle.com/article/Google-Book-Search-A-Dis/48245/>.
- Oblinger, Diana; James Oblinger (red.) (2005). Educating the Net Generation. Washington: Educare.
- OCLC Marketing (2004). 2004 Information format trends: content, not containers. Dublin, Ohio: OCLC.
- Odlyzko, Andrew [dok. elektr.] (1994). Tragic loss or good riddance? The impending demise of traditional scholarly journals. *Journal of Universal Computer Science*, October. http://www.jucs.org/jucs_0_0/tragic_loss_or_good/Odlyzko_A.pdf.
- Olędzki, Jerzy (2001). Komunikowanie w świecie : narzędzia, teorie, unormowania. Warszawa: Aspra-Jr.
- Oppenheim, Charles (2008). Legal issues for information professionals IX : An overview of recent developments in the law, in relation to the Internet. *Journal of Documentation*, vol. 64, nr 6, s. 938-955.
- Oppenheim, Charles; Fytton Rowland (2009). Scoping study on issues relating to quality-control measures within the scholarly communication process. W: Susanna Mornati, Turid Hedlund (ed) Rethinking electronic publishing: innovation in communication paradigms and technologies : EIPub2009 : 13th International Conference on Electronic Publishing. Milano 10-12 June 2009. Milano: Nuova Cultura, s. 263-289.
- O'Reilly, Tim [dok. elektr.] (2005). What is Web 2.0 : design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.
- Orsdel, Lee (2007). The state of scholarly communications : an environmental scan of emerging issues, pitfalls, and possibilities. W: Carol Borchet, Gary Ives (ed.) Mile high views : surveying the serials vista. New York, London, Oxford: The Haworth Press, s.191-209.
- Orsdel, Lee; Kathleen Born [dok. elektr.] (2008). Periodicals price survey 2008 : embracing openness. *Library Journal*, 4/15/2008. <http://www.libraryjournal.com/article/CA6547086.html>.
- Osińska, Veslava [dok. elektr.] (2006). Przybliżenie semantyczne w wizualizacji informacji w Internecie i bibliotekach cyfrowych. *Biuletyn EBIB*, nr 7 (77). <http://www.cbib.info/2006/77/osinska.php>.
- Otlet, Paul (1935). Monde, essai d'universalisme: connaissance du monde, sentiment du monde, action organisée et plan du monde. Brussels: Mundaneum.
- Owen, John Stuart Mackenzie (2007). The scientific article in the age of digitization. Berlin: Springer Verl.
- Owen, John Stuart Mackenzie (2006). Authenticity and objectivity in scientific communication: implications of digital media. W: Sonja Neef, Jose van Dijck, Eric Ketelaar (red.). Sign here! Handwriting in the age of new media. Amsterdam: Amsterdam Univ. Press, s. 60-75.
- Owen, John Stuart Mackenzie (2002). The new dissemination of knowledge: digital libraries and institutional roles in scholarly publishing. *Journal of Economic Methodology*, vol. 9, nr 3, s. 275-288.
- Owen, John Stuart Mackenzie [dok. elektr.] (2001). Tacit knowledge in action: basic notions of knowledge sharing in computer supported work environments. <http://dare.uva.nl/document/14964>.
- Owen, John Stuart Mackenzie [dok. elektr.] (2000). From hybrid to virtual scientific communication: implications for the information professional. <http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/Straatsburg.htm>.
- Owen, John Stuart Mackenzie (1998). Organizing for digital archiving: new distribution models in the scientific information chain. W: I. Butterworth (red.), The impact of electronic publishing on the academic community. London: Portland Press, s. 128-133.

- Palmquist, Ruth; Kyung-Sun Kim (2000). Cognitive style and online database search experience as predictions of Web search performance. *Journal of ASIS*, vol. 51, nr 6, s. 558-566.
- Pamula-Cieślak, Natalia (2006). Typologia zasobów ukrytego Internetu. *Przegląd Biblioteczny*, R. 74, z. 2, s. 153-164.
- Parks, Robert (2007). The Faustian grip of academic publishing. *Journal of Economic Methodology*, vol. 9, nr 3, s. 317-335.
- Pask, Gordon (1976). *Conversation theory: applications in education and epistemology*. Amsterdam, New York: Elsevier.
- Pearce, Judith; Janifer Gatenby [dok. elektr.] (2005). New frameworks for resource discovery and delivery. <http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1216/1501>.
- Penny, Dan (2008). Publishing technologies: what does the future hold? *Learned Publishing*, vol. 21, nr 1, s. 39-47.
- Pierce, Marlon [et al.] (2006). *Cyberinfrastructure and Web 2.0*. W: Lucio Grandinetti (ed.) *High performance computing and grids in action*. Amsterdam: IOS Press, s. 265-287.
- Pikas, Christina [dok. elektr.] (2006). The impact of information and communication technologies on informal scholarly scientific communication: a literature review. http://terpconnect.umd.edu/~cpikas/878/Pikas_The_Impact_of_ICTs_on_ISSC_0506.pdf.
- Pindłowi, Wanda (2008). Foresight dla bibliotek. W: Elżbieta Zybort, Dorota Grabowska (red.) *Książka, biblioteka, informacja w kręgu kultury i edukacji*. Warszawa: Wydaw. SBP, s. 149-155.
- Pindłowa, Wanda (2005). O zjawisku globalizacji coraz bardziej widocznym w pracy bibliotek. W: Halina Ganińska (red.) *Biblioteki naukowe w kulturze i cywilizacji: działania i codzienność*. Poznań: Bibl. Gł. Polit. Poznańskiej, s. 44-51.
- Pindłowa, Wanda (2000). [dok. elektr.] Czy technika, która obecnie wspomaga komunikację między ludźmi, doprowadzi do upadku zawodu bibliotekarza – pracownika informacji? *Biuletyn EBIB*, nr 1(9). <http://oss.wroc.pl/biuletyn/ebib09/pindlowa.html>.
- Plutchak, Scott (2007). What's a serial when you're running on Internet time? W: Carol Borchet, Gary Ives (ed.) *Mile high views : surveying the serials vista*. Binghamton: The Haworth Press, s. 79-90.
- Płoszajski, Grzegorz (red.) (2008). *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*. Warszawa: Bibl. Gł. Polit. Warszawskiej.
- Polański, Przemysław (2006). Zarys autonomicznego prawa Internetu. *Studia Iuridica*, XLV, s. 173-195.
- Polanyi, Michael (1998) *Personal knowledge : towards a post critical philosophy*. London: Routledge.
- Pomerantz, Jeffrey [dok. elektr.] (2008). Digital (library services) and (digital library) services. *Journal of Digital Information*, vol. 9, nr 27. <http://journals.tdl.org/jodi/article/view.227/210>.
- Pomerantz, Jeffrey (2006). Google Scholar and 100 percent availability of information. *Information Technology and Libraries*, vol. 25, nr 2, s. 52-56.
- Pomerantz, Jeffrey; Gary Marchionini (2007). The digital library as place. *Journal of Documentation*, vol. 63, nr 4, s. 505-533.
- Pool, Ithiel de Sola (1983). *Technologies of freedom*. Cambridge: Belknap Press, 299 s.
- Popper, Karl Raimund (2002). *Logika odkrycia naukowego*. Warszawa: PWN, 382 s.
- Poynder, Richard [dok. elektr.] (2008a). Open Access – dyskusja o kosztach. *Biuletyn EBIB*, nr 6(97). <http://www.ebib.info/2008/97/a.php?poynder2>.
- Poynder, Richard [dok. elektr.] (2008b). Open Access – pożegnanie recenzowania? *Biuletyn EBIB*, nr 6(97). <http://www.ebib.info/2008/97/a.php?poynder>.
- Prensky, Marc (2001). Do they really think differently? : digital natives, digital immigrants. Part 2. *On the Horizon*, vol. 9, nr 6, s. 1-6.
- Price, Derek de Solla (1965). *Węzłowe problemy historii nauki*. Warszawa: PWN.
- Program rozwoju informatycznej infrastruktury nauki na lata 2007-2013. [dok. elektr.] (2007). Warszawa: Min. Nauki i Szk. Wyższego. http://www.nauka.gov.pl/fileadmin/user_upload/29/41/29410/SN_2007_06_07_Program_rozwoju_infrastruktury_informatycznej_nauki_na_lata_2007_-_2013.pdf.
- Próchnicka, Maria (2006). Metafory Internetu a strategie poszukiwania informacji w sieci. W: Diana Pietruch-Reizes, Wiesław Babik (red.) *Kierunki i priorytety rozwoju informacji naukowej w kontekście budowania społeczeństwa wiedzy*. Katowice: PTIN, s. 40-49.

- Próchnicka, Maria (2004). Człowiek i komputer : dialogowy model wyszukiwania informacji. Kraków: Wydaw. UJ.
- Rajendiran, P. [dok. elektr.] (2006). Electronic grey literature in accelerator science and its allied subjects: selected Web resources for scientists and engineers. *HEP Libraries Webzine*, nr 12. <http://library.cern.ch/HEPL/12/papers/4/>.
- Ratajewski, Jerzy (1993). Elementy naukoznawstwa i główne kierunki rozwoju nauki europejskiej. Katowice: Uniw. Śląski.
- Rayward, Boyd [dok. elektr.] (2005). The historical development of information infrastructure and the dissemination of knowledge: a personal reflection. *Bulletin of ASIST*, vol. 31, nr 4. <http://www.asis.org/Bulletin/Apr-05/rayward.html>.
- Reddy, Michael [dok. elektr.] (1979). The conduit metaphor: a case of frame conflict in our language about language. <http://citeseer.comp.nus.edu.sg/context/277859/0>.
- Reference Model for an Open Archival Information System (OAIS) (2002). Washington: CCSDS.
- Reichman, Jerome; Paul Uhler (2003). A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law and Contemporary Problems*, vol. 66, s. 315-462.
- Renear, Allen; Carole Palmer (2009). Strategic reading, ontologies, and the future of scientific publishing. *Science*, vol. 325, nr 5942, s. 828-832.
- Revhaim, Radosława [dok. elektr.] (2002). Impact Factor – polubić czy lekceważyć? W: XXI Konferencja Problemowa Bibliotek Medycznych „Biblioteki medyczne wobec problemu bibliometrycznej oceny dorobku naukowego uczelni”. BG AM w Poznaniu, 28-29.11.2002. <http://www2.bg.am.poznan.pl/inne/konferencja/streszczenia.php?i=2>.
- Rifkin, Jeremy (2003). Wiek dostępu : nowa kultura hiperkapitalizmu, w której płaci się za każdą chwilę życia. Wrocław: Wydaw. Dolnośląskie.
- Rijnsoever, Frank; Rogier Donders (2009). The effect of innovativeness on different levels of technology adoption. *Journal of ASIST*, vol. 60 nr 5, s. 984-996.
- Robinson, Mary; Judith Wusteman (2007). Putting Google Scholar to the test: a preliminary study. *Program*, vol. 41, nr 1, s. 71-80.
- Rogers, Everett (2003). Diffusion of innovations. New York: Free Press.
- Ronfeldt, David; John Arquilla [dok. elektr.] (2007). The promise of noöpolitik. *First Monday*, vol. 12, nr 8. http://www.firstmonday.org/issues/issue12_8/ronfeldt/index.html.
- Rosa, Cathy De [et al.] [dok. elektr.] (2005). Perceptions of libraries and information resources : a report to the OCLC membership. <http://www.oclc.org/reports/2005perceptions.htm>.
- Rosenfeld, Louis; Peter Morville (2003). Architektura informacji w serwisach internetowych. Gliwice: Helion.
- Rosenthal, David [et al.] [dok. elektr.] (2005). Requirements for digital preservation systems. *D-lib Magazine*, vol. 11, nr 11. <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.
- Ross, Lyman; Pongracz Sennyey (2008). Library is dead, long live the library! The practice of academic librarianship and the digital revolution. *Journal of Acad. Librarianship*, vol. 34, nr 2, s. 145-152.
- Rothenberg, Jeff (2000). Preserving authentic digital information. W: Authenticity in a digital environment. Washington: CLIR, s. 51-68.
- Rothenberg, Jeff (1995). Ensuring the longevity of digital information. *Scientific American*, vol. 272, nr 1, s. 42-47.
- Rowland, Ian [et al.] (2008). The Google generation: the information behaviour of the researcher of the future. *ASLIB Proceedings*, vol. 60, nr 4, s. 290-310.
- Ryan, Judith [et al.] (2002). The Future of Scholarly Publishing. *Profession*, vol. 1, s. 172-186.
- Rzeczka, Henry; Peter Murray-Rust (2001). A new publishing paradigm: STM articles as part of the semantic web. *Learned Publishing*, vol. 14, nr 3, s. 177-182.
- Salc, Arthur (2006). Researchers and institutional repositories. W: Neil Jacobs (ed.) Open Access: key strategic, technical and economic aspects. Oxford: Chandos Publ., s. 87-98.
- Sapa, Remigiusz (2005). Benchmarking w doskonaleniu serwisów WWW bibliotek akademickich. Kraków: Wydaw. UJ.
- Sapa, Remigiusz [dok. elektr.] (2002). Jakość serwisów WWW bibliotek akademickich – usability test. *Biuletyn EBIB*, nr 2(31). <http://ebib.oss.wroc.pl/2002/31/sapa.php>.
- Saracevic, Tefko (2007). Relevance: a review of the literature and a framework for thinking on the notion in information science. Part 2, Nature and manifestations of relevance. *Journal of ASIST*, vol. 58, nr 13, s. 1915-1933.

- Saracevic, Tefko [dok. elektr.] (2004). Evaluation of digital libraries: an overview. W: DELOS WP7 Workshop on evaluation of Digital Libraries. Padova, 4-5 October 2004. http://www.scils.rutgers.edu/~tefko/DL_evaluation_Delos.pdf.
- Saracevic, Tefko (2000). Digital library evaluation: toward an evolution of concepts. *Library Trends*, vol. 49, nr 2, s. 350-369.
- Sathe, Nila; Jenifer Grady; Nunzia Giuse (2002). Print versus electronic journals : a preliminary investigation into the effect of journal format on research processes. *Journal of Medical Library Assoc.*, vol. 90, nr 2, s. 235-243.
- Saunders, Pam [dok. elektr.] (2002). From the library, any e-time. <http://www.vala.org.au/vala2002/2002pdf/33Saunders.pdf>.
- Schlimgen, Joan; Michael Kronenfeld (2004). Update on inflation of journal prices: Brandon/Hill list journals and the scientific, technical and medical publishing market. *Journal of Medical Library Assoc.*, vol. 92 nr 3, s. 307-314.
- Schroeder, Ralph (2007). E-Research infrastructures and open science: towards a new system of knowledge production. *Prometheus*, vol. 25, nr 1, s. 1-17.
- Selwyn, Neil (2009). The digital native – myth and reality. *ASLIB Proceedings*. vol. 61, nr 4, s. 364-379.
- Sennyey, Pongracz; Lyman Ross; Caroline Mills (2009). Exploring the future of academic libraries. A definitional approach. *Journal of Acad. Librarianship*, vol. 35, nr 3, s. 252-259.
- Shadbolt, Nigel [et al.] (2006). The Open Research Web. W: Niel Jacobs (ed.) Open Access : key strategic, technical and economic aspects. Oxford: Chandos Publ., s. 195-208.
- Shenton, Helen (2005). The future shape of collection storage. W: John Webster (ed.) Where shall we put it? Spotlight on collection storage issues. Papers given at the National Preservation Office Conference held 4 October 2004 at the British Library. London: The British Library, s. 4-15.
- Shotton, David (2009). Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, vol. 22, nr 2, s. 85-94.
- Shotton, David [et al.] [dok. elektr.] (2009). Adventures in semantic publishing: exemplar semantic enhancements of a research article. *PLoS Computational Biology*, vol. 5, nr 4. 17 doi:10.1371/journal.pcbi.1000361.
- Shieber, Stuart [dok. elektr.] (2009). Equity for Open Access journal publishing. *PLoS Biology*, vol. 7, nr 8. doi:10.1371/journal.pbio.1000165.
- Sienkiewicz, Piotr (2006). Społeczeństwo informacyjne jako społeczeństwo ryzyka. W: Lesław Haber, Marian Niezgodą (red.) Społeczeństwo informacyjne : aspekty funkcjonalne i dysfunkcjonalne. Kraków: Wydaw. UJ, s. 61-70.
- Sicńko, Marcin (2002). Człowiek w pajęczynie : Internet jako zjawisko kulturowe. Wrocław: Atut.
- Siess, Judith [dok. elektr.] (2009). Bibliotekarstwo uczestniczące – co to takiego i czy nas bezpośrednio dotyczy? *Biuletyn EBIB* Nr 9(109) <http://www.ebib.info/2009/109/a.php?siess>.
- Singh, Gian; Rekha Mittal; Moin Ahmad (2007). A bibliometric study of literature on digital libraries. *The Electronic Library*, vol. 25, nr 3, s. 342-348.
- Slone, Debra (2005). A bird's eye view of cross-platform Web interaction. *Journal of Documentation*, vol. 61, nr 5, s. 657-669.
- Smarr, Larry; Charles Catlett (2003). Metacomputing. W: Fran Berman, Geoffrey Fox, Anthony Hey (ed.) Grid computing : making a global infrastructure a reality. Chichester: Wiley & Sons, s. 825-836.
- Soergel, Dagobert [dok. elektr.] (2002). A framework for digital library research : broadening the vision. *D-Lib Magazine*, vol. 8, nr 12. <http://www.dlib.org/dlib/december02/soergel/12soergel.html>.
- Soergel, Dagobert [dok. elektr.] (1998). A digital library model with rich semantic structure. <http://citescer.ist.psu.edu/papers/cs/14041/>.
- Solka, Jeffrey (2008). Text data mining: theory and methods. *Statistics Surveys*, vol. 2, s. 94-112.
- Sompel Van de, Herbert [et al.] [dok. elektr.] (2004). Rethinking scholarly communication: building the system that scholars deserve. *D-Lib Magazine*, vol. 10, nr 9. <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html>.
- Søndergaard, Trine; Jack Andersen; Birger Hjørland (2003). Documents and the communication of scientific and scholarly information : Revising and updating the UNISIST model. *Journal of Documentation*, vol. 59, nr 3, s. 278-320.
- Solska, Ewa (2009). Globalizacja – państwo – uniwersytet. Dyskurs utopistyczny. *Kultura – Historia – Globalizacja*, Nr 5, s. 133-142.

- Sosińska-Kalata, Barbara (2008). Ewolucja modeli organizacji wiedzy w systemach organizacyjnych. W: Elżbieta Zybort, Dorota Grabowska (red.) Książka, biblioteka, informacja w kręgu kultury i edukacji. Warszawa: Wydaw. SBP, s. 103-116.
- Sosińska-Kalata, Barbara (2005). Systemy organizacji wiedzy w środowisku sieciowym. W: Barbara Sosińska-Kalata, Maria Przystek-Samokowa (red.) Od informacji naukowej do technologii społeczeństwa informacyjnego. Warszawa: Wydaw. SBP, s. 141-162.
- Sowa, John (2000). Knowledge representation : logical, philosophical and computational foundations. Albany: Brooks/Cole.
- Spink, Amanda; Bernard Jansen (2004). Web search: public searching of the Web. Dordrecht: Kluwer AP.
- Stoll, Clifford (2000). Krzemowe remedium : garść rozważań na temat infostrady. Poznań: Rebis.
- Strotmann, Andreas; Dangzhi Zhao (2009). Bibliometric factor maps for knowledge Discovery in digital libraries. W: Susanna Mornati, Turid Hedlund (ed.) Rethinking electronic publishing: innovation in communication paradigms and technologies. EIPub 2009: 13th Intern. Conference on Electronic Publishing. Milano 10-12 June 2009. Milano: Nuova Cultura, s. 501-512.
- Strózcwski, Władysław (2003). Ontologia. Kraków: Aureus, Znak.
- Suber, Peter [dok. elektr.] (2008). Równie prawa dla autorów i wydawców. *Biuletyn. EBIB*, nr 5(96). <http://www.ebib.info/2008/96/a.php?suber>.
- Subiata, Kazimierz (1999). Słownik terminów z zakresu obiektowości. Warszawa: Akad. Oficyna Wydaw. PLJ.
- Surowiecki, James (2005). The wisdom of crowds. New York: Anchor.
- Svenonius, Elaine (2000). The intellectual foundation of information organization. Cambridge, Mass., London: The MIT Press.
- Swoboda, Izabela (2009). OPAC WWW wizytówką biblioteki. Ocena jakości katalogów komputerowych małych i średnich bibliotek w Polsce. *Przegląd Biblioteczny*, R. 77, z. 1, s. 16-40.
- Szczeptańska, Barbara (2007a). Prawo autorskie – ochrona dzieł elektronicznych. W: Jadwiga Woźniak-Kasperek, Jerzy Franke (red.) Biblioteki cyfrowe : projekty, realizacje, technologie. Warszawa: Wydaw. SBP, s. 51-66.
- Szczeptańska, Barbara (2007b). Zasady publikowania obiektów cyfrowych w repozytoriach europejskich. Zagadnienia prawnautorskie. *Przegląd Biblioteczny*, R. 75, z. 2, s. 193-220.
- Szczygłowska, Lidia (2006). „Cybrary” to wciąż biblioteka. W: Biblioteki XXI wieku. Czy przetrwamy? : II Ogólnopolska Konferencja Biblioteki Politechniki Łódzkiej. Łódź, 19-21 czerwca. Łódź: Bibl. Gł. Polit. Łódzkiej, s. 51-70.
- Szumilewicz-Lachman, Irena (1989). Przedmowa. W: John Watkins, Nauka a sceptycyzm. Warszawa: PWN, s. 1-XXI.
- Święćkowska, Teresa (2008). Dostęp do wiedzy wobec rozwoju rynku publikacji cyfrowych. Zagrożenia i nadzieje. W: Katarzyna Materska, Ewa Chuchro, Barbara Sosińska-Kalata (red.) Organizowanie środowiska informacji i wiedzy. Warszawa: Wydaw. SBP, s. 41-57.
- Tadeusiewicz, Ryszard (2006). Rewolucja społeczeństwa informacyjnego na tle wcześniejszych rewolucji cywilizacyjnych. W: Lesław Haber, Marian Nieczgoda (red.) Społeczeństwo informacyjne : aspekty funkcjonalne i dysfunkcjonalne. Kraków: Wydaw. UJ, s. 31-45.
- Taraborelli, Dario [dok. elektr.] (2008). Soft peer review: social software and distributed scientific evaluation. http://nitens.org/docs/spr_coop08.pdf.
- Taylor, Arlene (2003). The organization of information. Westport, CN: Libraries Unlimited.
- Taylor, Michael; Pandelis Perakakis; Varvara Trachana (2008). The siege of science. *Ethics in Science and Environmental Politics*, vol. 8, nr 1, s. 17-40.
- Tedd, Lucy; Andrew Large (2005). Digital libraries : principles and practice in a global environment. München: K. G. Saur.
- Teilhard de Chardin, Pierre (1993). Fenomen człowieka. Warszawa: PAX.
- Tennant, Roy [dok. elektr.] (2004). Libraries through the looking-glass. <http://www.cdlib.org/inside/news/presentations/rtennant/2004ala/>.
- Tenopir, Carol [et al.] (2009). Electronic journals and changes in scholarly article seeking and reading patterns. *ASLIB Proceedings*, vol. 61, nr 1, s. 5-32.
- Tenopir, Carol; Donald King [dok. elektr.] (2008). Electronic Journals and changes in scholarly article seeking and reading patterns. *D-Lib Magazine*, vol. 14, Nr 11/12. <http://www.dlib.org/dlib/november08/tenopir/11tenopir.html>.
- Tenopir, Carol; Donald King [dok. elektr.] (2000). Towards electronic journals: realities for scienti-

- sts, librarians, and publishers. *Psychology*, vol. 11, nr 84. <http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?11.084>.
- Thorin, Suzanne [dok. elektr.] (2003). Global changes in scholarly communication. E-Workshop on Scholarly Communication in the Digital Era. 12-14 August, Taichung, Taiwan Feng Chia University. <http://www.arl.org/bm~doc/thorin.pdf>.
- Toffler, Alvin (2001). Trzecia fala. Warszawa: PIW.
- Tonkin, Emma [dok. elektr.] (2008). Persistent identifiers: considering the options. *Ariadne*, nr 56. <http://www.ariadne.ac.uk/issue56/tonkin/>.
- Tonkin, Emma; Henk Muller (2008). Keyword and metadata extraction from pre-prints. W: Leslic Chan, Susanna Mornati (ed.) Open scholarship: authority, community and sustainability in the age of Web 2.0. Proc. of the 12th Intern. Conference on Electronic Publishing EIPub'08. Toronto, 25-27 June 2008. Toronto: Univ. of Toronto, s. 30-44.
- Townsend, Robert (2003). History and the Future of Scholarly Publishing. *Perspectives*, vol. 41, nr 7, s. 32-41.
- Tredinnick, Lukc (2007). Post-structuralism, hyperext, and the World Wide Web. *ASLIB Proceedings*, vol. 59, nr 2, s. 169-186.
- Treloar, Andrew [dok. elektr.] (1999). Applying hypertext and hypermedia to scholarly journals enables both product and process innovation. *ACM Computer Survey*, vol. 31, nr 4. <http://www.acm.org/surveys/Formatting.html>.
- Unsworth, John [dok. elektr.] (1997). The importance of failure. *Journal of Electronic Publishing*, vol. 3, nr 2. <http://www.press.umich.edu/jcp/03-02/unsworth.html>.
- Veltman, Kim (2005). Access, claims, and quality on the Internet – future challenges. *Progress in Informatics*, nr 2, s. 17-40.
- Vetulani, Agnieszka (2009). Dzieła osierocone – w poszukiwaniu europejskiego rozwiązania. *Prace z Wynalazczości i Ochrony Własności Intelektualnej : zeszyty naukowe Uniwersytetu Jagiellońskiego*, z. 103, s. 26-50.
- Vickery, Brian [dok. elektr.] (2005). Information science. <http://www.lucis.me.uk/infosci1.htm>.
- Vickery, Brian (2000). Scientific communication in history. Lanham: Scarecrow Press.
- Vickery, Brian (1997). Metatheory and information science. *Journal of Documentation*, vol. 53, nr 5, s. 457-476.
- Waldrup, Mitchell (2008). Science 2.0 – is Open Access science the future? Is posting raw results online, for all to see, a great tool or a great risk? *Scientific American*, vol. 298, May.
- Wang, Bing (1999). A hybrid system approach for supporting digital libraries. *International Journal on Digital Libraries*, vol. 2, nr 2-3, s. 91-110.
- Ware, Mark [dok. elektr.] (2006). Scientific publishing in transition: an overview of current developments. <http://www.alpsp.org/ForceDownload.asp?id=127>.
- Ware, Mark; Michael Mabe (2009). An overview of scientific and scholarly journals publishing. Oxford: Intern. Assoc. of STM Publishers.
- Waters, Donald (2006). Managing digital assets in higher education: An overview of strategic issues. *ARL Bimonthly Report*, 244, s. 1-10.
- Waters, Lindsay (2009). Scholarship and silence. W: Albert Greco (ed.) The state of scholarly publishing : challenges and opportunities. New Brunswick: Transaction Publ., s. 27-33.
- Weaver, Warren (1948). Science and complexity. *American Scientist*, vol. 36, s. 536-544.
- Weibel, Stuart; Renato Iannella; Warwick Cathro [dok. elektr.] (1997). The 4th Dublin Core Metadata Workshop Report. *D-Lib Magazine*, nr 6. <http://www.dlib.org/dlib/june97/metadata/06weibel.html>.
- Weinberg, Alvin (1961). Impact of large-scale science on the United States. *Science*, vol. 134, s. 161-164.
- Weinberger, David (2007). Everything is miscellaneous : the power of the new digital disorder. New York: Holt Paperback.
- Weitzman, Jonathan [dok. elektr.] (2004). Mis(leading) Open Access myths. <http://www.biomed-central.com/openaccess/inquiry/myths/>.
- Welshons, Marlo, red. [dok. elektr.] (2006). Our cultural commonwealth : the report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences. <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>.
- Wenz, Karin (2008). Tekst w dobie jego reprodukcji elektronicznej. W: Andrzej Gwóźdź (red.) *Ekrany piśmiennosci*. Warszawa: Wydaw. Akad. i Profesjonalne, s. 103-117.
- Westbrook, Bradley (2002). Prospecting virtual collections. *Journal of Archival Organization*, vol. 1, nr 1, s. 73-80.

- Wheeler, Brad (2008). E-research is a fad: scholarship 2.0, cyberinfrastructure and IT governance. W: Richard Katz (ed.) *The tower and the cloud : higher education in the age of cloud computing*. Washington: Educuse, s. 108-117.
- Whitworth, Brian; Rob Friedman [dok. elektr.] (2009). Reinventing academic publishing online. Part II: a socio-technical vision. *First Monday*, vol. 14, nr 9. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/rt/printerFriendly/2642/2287>.
- Williams, Peter; Ian Rowlands; Maggie Fieldhouse (2008). The 'Google Generation' – myths and realities about young people's digital information behaviour. W: David Nicholas, Ian Rowlands (ed.) *Digital consumers: reshaping the information professions*. London: Facet Publ., s. 159-192.
- Willinsky, John (2006). *The access principle : the case for Open Access to research and scholarship*. Cambridge, London: The MIT Press.
- Wilson, Tom (2000). Human information behaviour. *Informing Science*, vol. 3, nr 2, s. 49-56.
- Wilson, Tom (1999). Models in information behaviour research. *Journal of Documentation*, vol. 55, nr 3, s. 249-270.
- Witten, Ian; David Bainbridge (2003). *How to Build a Digital Library*. San Francisco: Morgan Kaufman Publ.
- Wojarnik, Grzegorz (2007). Wykorzystanie metod sztucznej inteligencji w zastosowaniach internetowych. W: Agnieszka Szewczyk (red.) *Społeczeństwo informacyjne – problemy rozwoju*. Warszawa: Difin.
- Wojciechowski, Jacek (2008a). Aktualia 1. *Bibliotekarz*, nr 12, s. 34.
- Wojciechowski, Jacek (2008b). Biblioteka w integracji środowiskowej. *Przegląd Biblioteczny*, R. 76, z. 2, s. 226-239.
- Wojciechowski, Jacek (2007). Pomiędzy drukiem a dygitalizacją. W: *Przyszłość bibliotek w Polsce : materiały z ogólnopolskiej konferencji*, Warszawa, 12-13.10.2007 r. Warszawa: Wydaw. SBP, s. 155-169.
- Wojciechowski, Jacek (2006). *Biblioteczna wartość naddana*. Kraków: Wydaw. UJ.
- Wolniewicz, Tomasz [dok. elektr.] (2008). Katalogi wirtualne na przykładzie polskiej wyszukiwarki KaRo. W: Rola katalogu centralnego NUKAT w kształtowaniu społeczeństwa wiedzy w Polsce : międzynarodowa konferencja z okazji 5-lecia istnienia katalogu centralnego NUKAT, Warszawa 23-25 stycznia 2008 r. http://www.nukat.edu.pl/konf2008/mater/54_wolniewicz.pdf.
- Wolniewicz, Tomasz [dok. elektr.] (2002). Analiza standardu Z39.50 pod kątem możliwości przechowywania parametrów konfiguracyjnych w bazie LDAP. <http://projekt-ldap.uci.umk.pl/raporty/ftp/uci/z3950-ldap.html>.
- Woźniak-Kasperek, Jadwiga (2008). System organizacji wiedzy czy folksonomia? W: Elżbieta Zybort, Dorota Grabowska (red.) *Książka, biblioteka, informacja w kręgu kultury i edukacji*. Warszawa: Wydaw. SBP, s. 117-124.
- Wright, Alex (2007). *Glut : mastering information through the ages*. Ithaca: Cornell UP.
- Yang, Chyan; Keng-Chieh Yang; Hsu-Chieh Yuan (2007). Improving the search process through ontology-based adaptive semantic search. *The Electronic Library*, vol. 25, nr 2, s. 234-248.
- Yiotis, Kristin (2005). The Open Access initiative: a new paradigm for scholarly communication. *Information Technology and Libraries*, vol. 24, nr 4, s. 157-162.
- Young, Nancy; Marilyn Seggern (2001). General information seeking in changing times. *Reference and User Services Quarterly*, vol. 41, nr 2, s. 159-168.
- Yu, Holly; Margo Young (2004). The impact of Web search engines on subject searching in OPAC. *Information Technology and Libraries*, vol. 23, nr 4, s. 168-180.
- Zacher, Lech (2006). Społeczeństwo postinformacyjne w kontekście ewolucji społeczeństw i wizji przyszłości. W: Lesław Haber, Marian Niezgodą (red.) *Społeczeństwo informacyjne : aspekty funkcjonalne i dysfunkcjonalne*. Kraków: Wydaw. UJ, s. 71-81.
- Zawado, Hanna (2002). Elektroniczne wydawnictwa ciągłe w bibliotekach. *ZIN*, nr 2(80), s. 86-93.
- Zhang, Wende (2007). Digital library intellectual property right evaluation and method. *The Electronic Library*, vol. 25, nr 3, s. 267-273.
- Zhou, Qian (2005). The development of digital libraries in China and the shaping of digital librarians. *The Electronic Library*, vol. 23, nr 4, s. 433-441.
- Zhu, Qin (2004). Understanding OpenURL standard and electronic resources: effective use of available resources. *Program*, vol. 38, nr 4, s. 251-256.
- Zieliński, Grzegorz (2000). *XML : rozszerzalny język znakowań*. Warszawa: Wydaw. SBP.

Indeks osobowo-przedmiotowy

- AACR 109-111
Abstraktowe bazy danych 114, 136, 171
Académie des Sciences 30
Académie Française 30
Ackerman, Mark 124
Acta Eruditorum 31
Agent (człowiek) 103, 122
Agent (oprogramowanie) 22, 68, 97, 103, 175, 201, 247
Agregacja danych 175
Agregacja informacji 22
Agregacja metadanych 112, 165
Agregacja obiektów cyfrowych 83-85, 134
Agregacja odnośników 60, 118
Agregacja publikacji elektronicznych 85-87, 92
Agregacja treści 83, 87, 177, 246
Agregatorzy 136, 142, 220, 233
AHCI 160
AIFF 72,
AJAX 202
Akademia Umiejętności 30
Algebra Boole'a 159, 161-163, 167
AltaVista 167
Amatorskie użytkowanie informacji naukowej 209, 240
Amazon 62, 79, 98, 139, 177, 222
Amazoogle 61, 163
American Mathematical Society Patrz AMS
Amin, Mayur 43
AMS 36
Analiza cytowań 171, 196, 238
Analiza hiperlinków 152, 238
Analiza językowa 152
Analiza lingwistyczna Patrz Analiza językowa
Analiza ontologiczna 103
Analiza składniowa struktury dokumentu 151
Analiza wykorzystania bazy danych 166
Analizy bibliometryczne 152
Anderson, Chris 168
Anderson, James 152
Annotea 196
Antelman, Kristin 163
Apomediacja 186
Apple 72, 89-90
Aquadbrowser 163
Architektura informacji 101-102
Architektura klient-serwer 77, 117
Archiwizacja danych 43, 52, 55, 65
Archiwizacja obiektów cyfrowych 58, 64, 67, 82, 94,
124, 141, 146, 181, 213, 216, 235
ARL 38, 132
Armbruster, Chris 88
Arms, William 22, 138, 187, 199, 207
ARPA 117
Arpanet 117
Arts and Humanities Citation Index Patrz AHCI
ArXiv 93, 160, 196
Aristoteles 19, 102
ASCII 66, 69, 73
Association of Research Libraries Patrz ARL
Athens 187
Atkins, Daniel 43
Atom 202
AU (format) 72
Autentyczność obiektu cyfrowego 18, 82, 133
Autoarchiwizacja 98, 147, 182, 227, 235
Automatyczne indeksowanie 69, 149, 157, 170, 201,
239
Automatyczne tworzenie metadanych 95, 105, 150, 166,
176, 201
Autonomiczne indeksy cytowań 160
Autopublikowanie Patrz Samopublikowanie
Autorskie prawa majątkowe 89, 92, 216, 225
Autorskie prawa osobiste 225
Autorstwo dzieła 11, 32, 52, 87, 99, 147, 181-182, 226
Autorytety kognitywne 194
Autoryzacja dostępu 48
Autorzy ezoteryczni 212
Autorzy komercyjni 212
AVI 68, 72

Baker, Karen 216
Bard, Alexander 203
Barta, Janusz 224, 226
Batty, Michael 123
Bayat, Behrooz 24
Bazy danych 82, 102, 113-115, 158-161, 166, 176, 228
Bazy wiedzy 50, 102, 113
Beilstein (wydawnictwo) 38
Benchmarking 132
Benkler, Yochai 118, 227
Berners-Lee, Tim 118
Bernoff, Josh 203
Besser, Howard 22
Bib-1 78
Bibliograficzne bazy danych 78, 94, 113-114, 158-161
Bibliografie załącznikowe 99-100, 114, 136, 160, 179
Biblios.net 164, 229, 245
Biblioteka 2.0 204
Bibliotekarstwo uczestniczące 62, 154, 217
Bibliotekarze 40, 61, 94, 100, 110, 146, 149, 154, 163,
166, 180, 200, 206, 214-217
Bibliotekarze cyfrowi 62
Bibliotekarze dziedzinowi 62
Bibliotekarze jako wydawcy 215
Biblioteki cyfrowe 20, 23, 55, 63, 94, 122
Biblioteki hybrydowe 22, 24, 58, 113-114
Biblioteki naukowe 16, 56, 144, 206, 210, 248
Biblioteki wirtualne 59, 122
Bibliotekoznawcy 18, 214, 217
Bieber, Michael 22
Big Deal 88

BIOME 113
 BioMed Central 196
 BioMedNet 38
 Blackwell 191,
 Blogi 9, 15, 48, 59, 91, 98-100, 127, 202-204
 BOAI 146-147
 Borgman, Christine 15, 18, 19, 24, 33, 49, 95, 124, 135,
 146, 155, 184, 185, 188, 227
 Bramka tematyczna Patrz Portale internetowe
 Broker informacji 154
 Brooks, Terrence 27, 107
 BUBL 221
 Burk, Alan 81
 Burke, Mary 204
 Bush, Vannavar 19, 124, 144, 241

CAMiLEON 192
 Cardoso, Gustavo 32
 Carvalho, Elizabeth 91
 Castells, Manuel 42
 Cechy schematów metadanych 107
 CEDARS 192
 Cele kontroli bibliograficznej 110
 Celc prawa autorskiego 225
 Central Processing Unit Patrz CPU
 Ceny czasopism 39
 CGI 135, 165, 168
 Chamo (system) 162
 Chemical Abstracts Service 36
 Chen, Ching-Chih 20
 Cisek, Sabina 50
 Cisza wyszukiwawcza 130
 Citation Linker 137
 CiteSeer 160, 169, 171, 196
 CiteULike 127, 186
 Citizen Science 209
 CMS 162
 Connotea 127, 186
 Content Management System Patrz CMS
 Content-Bearing Clic 130
 Core Scientific MetaData Patrz CSMD
 COUNTER 132
 Coyle, Caren 110
 CPU 42, 222
 Crane, Gregory 84
 Creative Commons 41, 147, 202, 229
 CrossRef 136, 176, 179
 CSMD 111, 112
 CSS 68, 75
 Cullen, Rowena 132
 Cutler, Charles 19
 Cyberinfrastruktura Patrz IIN
 Cyberprzestrzeń 18, 54, 62, 119, 178-179
 Cyfrowa ochrona 141
 Cyfrowe biblioteki osobiste 124, 242, 245
 Cyfrowi imigranci 210
 Cyfryzacja 13, 24, 29, 46, 143, 233, 241
 Cykl wydawniczy 145, 224
 Cykl życia danych 215
 Cykl życia obiektu cyfrowego 68, 141

Cykl życia publikacji 235
 Cytowalność publikacji 126, 171
 Czasopisma elektroniczne 13, 58, 85, 88, 98, 136, 191
 Czasopisma hybrydowe 34
 Czasopisma interaktywne 98
 Czasopisma komercyjne 147, 189, 197, 210, 214, 218,
 223
 Czasopisma naukowe 27, 30-33, 38-40, 85, 88, 92, 96,
 128, 212
 Czasopisma recenzowane 13, 34, 88-89, 195
 Czasopisma STM 32, 38
 Czasopisma subsydiowane 223
 Czasopisma wirtualne 98
 Czytelnictwo strategiczne 100
 Czytelnik aktywny 208

Dahl, Mark 164
 Dane 48, 67, 75, 82, 162, 166, 184, 187, 200, 243
 Dane naukowe 35, 45, 50-52, 56, 60, 96-97, 106, 140,
 216, 225, 245
 DARPA 21, 1117
 DCCAP 110
 Defense Advanced Research Projects Agency Patrz
 DARPA
 Definicja GBC 56
 Delany, Paul 12
 DELOS 20, 23
 Dempsey, Lorean 61
 Derfert-Wolf, Lidia 154
 Deutsche Forschungsgemeinschaft Patrz DFG
 DFG 21
 Dickersin, Kay 41
 Digg 127, 186, 196
 Digital Library Federation Patrz DLF
 Digital Library for Earth System Education Patrz DLE-
 SE
 Digital Library Initiative Patrz DLI
 Digital natives Patrz Użytkownicy rdzennie cyfrowi
 Digital Rights Management Systems Patrz DRM
 Digitalizacja 13, 29, 55, 72, 90, 92, 142, 171, 180, 228,
 234
 DivX 72
 DjVu 68-70
 DLESE 21
 DLF 112
 DLI 21
 dLibra 58, 113
 DLO 81, 105, 166
 Długi ogon (model) 168, 197
 Długotrwała archiwizacja 14, 16, 41, 64, 68, 82, 94, 133,
 141, 213, 216-218
 Dobra prywatne 48, 120
 Dobra publiczne 34, 45, 47, 92, 119, 148, 180
 Dobra wspólne 48, 226, 228
 Document Type Definition Patrz DTD
 Document-like Object Patrz DLO
 DOI 136-137
 Dokumenty rekombinowane 84
 Dokumenty tekstowe 66-68, 87, 98, 110, 156
 Domena publiczna 119-120, 227-229

- Dostawcy baz danych 112, 114, 136, 159, 166
 Dostęp kastomizowany 205
 Dostęp licencjonowany 38, 118, 120, 224, 272, 325
 Dostępność publikacji naukowych 20, 33, 144, 179, 196, 204
 Dostępność serwisu GBC 131
 DRM 226, 227
 Drucker, Peter 15
 DTD 75-76
 Dublin Core 68, 76, 80, 107, 109-113, 150, 161, 184
 Dublin Core Collections Application Profile Patrz DCCAP
 Dynamiczne strony Web 82, 95, 143, 172, 210
 Dysonans poznawczy 14
 Dzieła osierocone 226, 228
 Dzieło 81, 208, 225-228
- Edwards, Paul 57
 eEurope 15
 EEVL 113
 Efekt małego świata 173
 Efekt skali 14, 35
 Efektywność funkcjonalna systemu 130
 Efektywność komunikacji naukowej 17, 19, 125
 Efektywność kosztowa systemu 130, 146
 Efektywność nauczania 51
 Efektywność wyszukiwania informacji 129, 169, 172, 174, 181, 193, 199, 206
 - Kompletność wyszukiwania 125, 129-131, 167, 171, 199
 - Trafność wyszukiwania 125, 129-131, 167, 199
- Egzemplarz autorytarny 99, 141, 183
 El Inc. 38
 EJS 144
 Ekonomia skali 221
 E-książki Patrz Książki elektroniczne
 Eksperyment 35, 50
 Eksploracja danych 52, 107, 140, 246
 Eksploracja tekstu 52, 151, 246
 Electronic Journal System Patrz EJS
 Elsevier 32, 38, 114, 160, 161, 191
 Emulacja 192
 e-Nauka Patrz IIN
 Encore 163, 164
 Endeca 163
 Engineering Index 36
 ENIAC 19
 Enquire 118
 E-platności 65
 ePolska 15
 EQUINOX 132
 Escarpit, Robert 11
 Esposito, Joseph 147
 Etykiety 48, 73-75, 101, 106, 144, 152
 Europeana 90
 eVALUed 132
 Ex Libris 136, 163
 Excerpta Medica 38
 Excite 167
 Extensible Markup Language Patrz XML
- Facebook 203
 Faculty of 1000 196
 Falsyfikacja teorii naukowej 50
 Filmowanie (strategia archiwizacji) 192
 Flash 71
 Flickr 57, 202, 204
 Folksonomic 63, 109, 202
 Formaty hybrydowe 69
 Formaty rastrowe 68, 70
 Formaty wektorowe 68, 71
 Formularze metadanych 137, 150
 Foster, Ian 44
 Foster, Richard 26
 Foucault, Michel 121
 Fragtegracia 58
 Francke, Helena 88, 194
 FRBR 81, 83, 85, 101, 106, 109, 110
 Freebase 175
 Freeman, Peter 44
 Freud, Zygmunt 50
 Friedman, Rob 233
 Friendfeed 186
 Frost, William 167
 Functional Requirements for Bibliographic Records Patrz FRBR
 Funkcja hash 156, 157
 Funkcje komunikacji naukowej 28, 33, 91, 141
 Funkcje metadanych 106, 114
 Funkcje publikacji 29, 93, 143, 146
- Galileusz 28
 Gallica 90
 Gawrysiak, Piotr 114, 115
 GDFR 72
 GenBank 96
 Generacje OPAC 161
 getCITED 160
 GIF 68, 70, 71
 GIGO 200
 Gildie uczonych 220, 247
 g-index 126
 Ginsparg, Paul 93
 Giornale de'letterati d'Italia 31
 Global Digital Format Repository Patrz GDFR
 Globalizacja informacji 209
 Globalizacja nauki 35
 Globalizacja 58, 120, 209, 245
 Głęboki Web Patrz Ukryty Web
 Głębokie linkowanie 224
 GNU GPL 67, 202, 229
 Godlee, Fiona 41
 Golka, Marian 119, 215
 Gołbiewski, Łukasz 244
 Gonçalves, Marcos 22
 Google 61, 92, 94, 110, 126, 139, 168-172, 176, 180, 189, 199, 214, 228, 242, 244
 Google Books 90, 98, 160, 171, 177, 190
 Google Effect 170
 Google Książki Patrz Google Books
 Google Maps 96,

Google Scholar 124, 126, 160, 167, 169-171, 176, 179, 245
 Górska, Małgorzata 29
 Górny, Mirosław 24
 Grafika komputerowa 140, 205
 Graphical User Interface Patrz GUI
 Greenberg, Jane 62
 Gresham, John 53
 Grid 44, 48, 52, 59, 65, 181, 239, 247
 Gruber, Tom 204
 Grygowski, Dariusz 55
 Guédon, Jean-Claude 195
 GUI 59, 138, 155, 161
 Gutenberg, Johann 26, 29, 62

 Habermas, Jürgen 33
 Hakala, Juha 135
 Hallmark, Julie 168
 Handel elektroniczny 42, 65, 77, 168, 181
 Handle 136
 Hardware 42, 45, 53, 132
 Harnad, Stevan 197, 212, 221
 Harvey, Ross 193
 Henderson, Albert 36
 Heterotopia 121
 Hcurystyki 151, 152
 Higgins, Sarah 141
 Hillmann, Diane 110
 h-index 126,
 Hiperlinki 69, 84, 97, 117, 125, 173, 196, 224, 238
 Hipertekst 19, 57, 73, 81, 95, 99, 118, 143, 161, 174, 208, 230
 Historia obiektu cyfrowego 52, 85, 183
 Honoraria autorskie 148
 Horgan, John 34
 HTML 48, 66, 68, 70, 73-75, 98, 108, 111, 150
 HTTP 68, 79, 106, 133, 139, 202
 HUMBUL 113
 Huxley, Julian 179
 Hybrydowy model komunikacji naukowej 235-237
 Hybrydyzacja publikowania naukowego 142
 Hybrydyzacja technologii 14

 Iacono, Suzanne 248
 ICT 14, 15, 42, 214, 218
 Identyfikacja publikacji 58, 84, 96, 105, 134, 247
 Identyfikacja użytkowników 186, 187
 Identyfikatory obiektów cyfrowych 82-84, 108, 183
 IDPF 75, 90
 IF Patrz Impact Factor
 IIN 15, 17, 24, 35, 42-48, 51-53, 65, 84, 107, 131, 148, 189, 203, 212, 217, 229, 235, 239, 247
 IIPC 189
 Impact Factor 32, 126, 147
 Indeks cytowań 32, 99, 100, 114, 160, 166, 170, 176, 179, 196, 218, 238
 Indeksowanie pełnotekstowe 61, 69, 95, 105, 111, 113, 171, 177, 239
 Index Copernicus 128
 Info URI 137
 Information and Communications Technology Patrz ICT
 Informatyczna infrastruktura nauki Patrz IIN
 Infosfera 99, 178, 179
 Infrastruktura 35, 42, 43, 141, 201, 241-243
 Infrastruktura informacyjna 15, 24, 64, 92, 179, 243
 Infrastruktura organizacyjna 43
 Infrastruktura sieciowa 14, 56
 Infrastruktura techniczna 43
 Innovative 136, 163
 Innowacyjność 29, 42, 45, 46, 158, 227, 241
 Instytucje wspomagające naukę 53, 209
 Integracja informacji 100, 172, 178, 201, 241
 Integralność dzieł pochodnych 182
 Inteligencja zbiorowa 109, 170, 179, 181, 202, 204, 214, 247
 Inteligentne serwisy GBC 174, 200
 Intencje użytkownika 115
 Interfejs bazy danych 54
 Interfejs GBC 74, 124, 139, 139, 199
 Interfejs uniwersalny 177, 199, 206, 220
 Interfejs użytkownika 22, 53, 59, 117, 118, 130, 161, 165, 175, 188, 246
 Interfejs wyszukiwarki 177, 239, 245
 Interfejs wyszukiwarki globalnej 170, 173, 176
 Interfejs wyszukiwawczy 71, 159, 166, 179, 245
 International Digital Publishing Forum Patrz IDPF
 International Internet Preservation Consortium Patrz IIPC
 International Standard Bibliographic Description Patrz ISBD
 International Standard Book Number Patrz ISBN
 International Standard Serial Number Patrz ISSN
 Internet Archive 189, 245
 Internet szerokopasmowy 65
 ISBD 109-111
 ISBN 133, 134, 136, 199
 ISO 2709 78, 108, 109, 113
 ISSN 135, 137

 JACKPHY 67
 Jacobfeuerborn, Bruno 46
 Jakość publikacji 29, 31, 41, 89, 93, 107, 125, 145, 193
 JANET 21
 Janiak, Małgorzata 44
 Jenkins, Henry 12, 170
 Język dokumentu 152, 185, 244
 Język opisu strony 69
 Języki Informacyjno-Wyszukiwawcze Patrz JIW
 Języki kodowania 18, 48, 64, 70, 133, 139, 150
 JHOVE 72
 JISC 21, 22
 JIW 102, 111, 133, 161, 166, 201
 Johnson, Ben 238
 Johnson, Mark 122, 123
 Joint Information Systems Committee Patrz JISC
 Journal des Sçavans 30, 31
 Journals STORage Patrz JSTOR
 JPEG 68, 70, 71
 JSTOR 72, 114

- Kahin, Brian 42
 Kahlc, Brewster 189
 Kahn, Robert 57
 Kalfoglou, Yannis 103
 Kalinichenko, Leonid 22
 Kastomizacja wyników wyszukiwania 172
 Kellogg, David 34
 King, Donald 38
 Klient-serwer Patrz Architektura klient-serwer
 Kling, Robert 248
 Kluwer 32,
 Kodeki 71, 72
 Kodowanie semantyczne 175, 201
 Kolaboratoria 17, 44, 53-55, 62, 91, 96, 123, 150, 154,
 207, 217
 Kolektywy wiedzy 59
 Komasaacja rynku wydawniczego 38, 40, 218
 Komentarze użytkowników 99, 110, 127, 194-196, 208,
 245
 Komercjalizacja komunikacji naukowej 88
 Komfort czytania 210
 Kompetencje informacyjne 211, 214
 Kompresja pliku 69, 71, 72
 Komputer osobisty 19, 89, 144, 198
 Komunikacja beczprzewodowa 64
 Komunikacja naukowa
 – Komunikacja formalna 9, 15, 33, 56, 87, 91, 133,
 141, 153, 204, 235-238, 243
 – Komunikacja nieformalna 9, 15, 33, 40, 56, 87, 90-
 92, 133, 147, 153, 195, 204, 235-238, 243
 Komunikacja oralna 33, 230
 Komunikacja utylitarna 243
 Komunikacja zapośredniczona 58, 120
 Koninklijke Bibliotheek 191
 Konkurencja w nauce 247
 Konsorcjum SCOAP Patrz SCOAP
 Konwergencja mediów 12
 Konwergencja ról 208
 Konwergencja technologii 56, 99
 Konwergencja treści 58, 65
 Konwersja do wersji cyfrowej Patrz Digitalizacja
 Konwersja formatów 74, 192, 215
 Konwersja katalogów 206
 Konwersja metadanych 93, 107-109, 149
 Kopernik, Mikołaj 28
 Koszt pierwszego egzemplarza 39, 146, 221
 Koszt pozyskania publikacji 209
 Kryteria oceny GBC 24, 131, 132
 Kryteria wyszukiwawcze 78, 86, 115, 153, 161, 199,
 244
 Krzysztofek, Kazimierz 59
 Książki elektroniczne 83, 87, 89, 90, 95, 98
 Kuhn, Thomas 51
 Kultura wymiany darów 236
 Kyrillidou, Marta 38

 Lagoze, Carl 85, 238
 Lakoff, George 122, 123
 Landow, George 12, 208
 LaTeX 83
 LCSH 162
 LDAP 187
 Lesk, Michael 221
 Lessig, Lawrence 9, 229
 Levy, David 22
 Lévy, Pierre 170
 Lewandowski, Dirk 130, 168
 LexisNexis 38, 159
 Leydesdorff, Loet 28, 238
 Li, Charlene 203
 LibQUAL 132
 Library 2.0 Patrz Biblioteka 2.0
 Library of Congress 21, 51, 77, 78, 109, 112
 Library of Congress Subject Headings Patrz LCSH
 Librarything 245
 Licencja free source 229
 Licencja GPL Patrz GNU GPL
 Licencja na publikowanie 226
 Licencja niewyłączna 228
 Licklider, Joseph 117
 Liew, Chem Li 24
 Lightweight Directory Access Protocol Patrz LDAP
 Listy dyskusyjne 87, 91, 99, 164, 189, 213, 236
 Living review 98
 LOCKSS 192
 Logosfera 14
 Long tail Patrz Długi ogon (model)
 Lossau, Norbert 166
 Lycos 167
 Lyman, Peter 51
 Lynch, Clifford 148, 181, 183

 Łańcuch informacji 62, 204, 231-235, 238
 Łatwość publikowania 11, 86, 99, 218

 Mabe, Michael 36
 Mała nauka 35, 217
 Mansourian, Yazdan 132
 Mapa pojęć 137
 Mapa serwisu 174
 Mapa źródeł 85
 MARC 49, 78, 80, 109-113, 161, 180, 199
 MARC 21 63, 111, 112
 Marchionini, Gary 22, 121, 125, 155, 156, 206
 MARCXML 109, 111, 112
 MARK I 19
 Markiewicz, Ryszard 224, 226
 Marksizm 50
 Marshall, Catherine 22
 Mashup 96, 242
 Materializacje (dzieła) 81, 82, 85, 121, 122, 238
 Materiały konferencyjne 87, 91, 98, 171, 195, 207
 Materska, Katarzyna 214
 MathSciNet 36, 160
 Matlak, Andrzej 224
 Mayr, Philipp 171
 Mądrość tłumów Patrz Inteligencja zbiorowa
 Mądrość zbiorowa Patrz Inteligencja zbiorowa
 McGuinness, Deborah 175
 McLuhan, Marshall 12

Meadows, Arthur 15
 Mediacja 104, 140, 175, 215
 Mediacyjne usługi informacyjne 214
 Meho, Loknam 171
 Memex 19, 124, 144, 241
 Mendeley 127
 Merton, Robert 34, 226
 Metadane 23, 35, 47, 56, 58, 61, 67, 73-80, 87, 92-95, 104, 124, 133-137, 158, 165, 176, 183, 201, 209, 213, 216, 244, 246
 Metadane „trzeciej strony” 108, 113
 Metadane administracyjne 105, 141, 149, 244
 Metadane nieustrukturyzowane 149
 Metadane opisowe 56, 105, 106, 109, 112, 133, 141, 149, 244, 246
 Metadane osadzone 67, 74, 108
 Metadane strukturalne 56, 82, 105, 141, 149, 244
 Metadane tworzone przez człowieka 105-107, 158
 Metadane ustrukturyzowane 107, 150, 158, 177, 199, 244
 Metadane wyodrębnione 76, 108
 Metadata Encoding and Transmission Standard Patrz METS
 Metadata Object Description Schema Patrz MODS
 Metafora 101, 122, 142, 231, 248
 Metafora biblioteki cyfrowej 124
 Metafora komunikacji naukowej 242
 Metafora miejsca 123-125
 Metafora przestrzeni Internetu 119, 242
 Metafory orientacyjne 123
 Metafory strukturalne 123
 Metainformacja 22, 65, 82, 139, 142, 209, 215
 Metakomputer 53, 59
 Metamodel 23, 104
 Metawyszukiwanie Patrz Wyszukiwanie sfederowane
 Metawyszukiwarki 79, 167, 169, 179
 Metoda dedukcyjno-hipotetyczna 50
 Metoda indukcyjna 50
 Metodologia 5S 18, 23, 64, 117, 141, 242
 Metodologia WSE 132
 Metody modelowania systemów 231
 METS 111, 112
 Meyer, Philip 33
 Microsoft 72
 Microsoft Office 114
 Microsoft Word 70
 Middleware 42, 43, 45, 53, 165
 Migracja (strategia archiwizacji) 192
 Migracja do postaci cyfrowej 12, 34
 Mikroformaty 150, 247
 Mikropłatności 65, 222, 243
 Mikrotreści 65
 Miksowanie danych Patrz Mashup
 Milgram, Stanley 173
 Milion Book Project 90
 Miller, Paul 188
 Millerand, Florence 216
 MIME 68
 MIX 111, 112
 Model 5S Patrz Metodologia 5S
 Model biblioteki cyfrowej 22, 132, 242
 Model cykliczny 142
 Model komunikacji naukowej 15, 143, 231, 237, 239
 Model łańcucha informacji 62
 Model prowincji wiedzy 216
 Model przewodu 231, 234
 Model transmisji danych 231
 Model wiedzy 103
 Modele ciągnące 231
 Modele ekonomiczne 33, 47, 147, 218, 220
 Modele finansowania 40, 221-223
 Modele mentalne 123, 231
 Modele pchające 231, 233
 Modele publikowania 13, 19, 33, 41, 142, 147, 154, 231
 Modele wyszukiwania informacji 153, 154, 159
 MODS 111, 112
 Moghaddam, Alireza 24
 Moore, Gordon 42
 Moving wall 216
 MP3 68, 72, 83
 MPEG 72
 MPEG-4 72
 Muller, Henk 151
 Multimedia 20, 36, 58, 68, 72, 82, 95, 206
 Mundaneum 19
 Muraszewicz, Mieczysław 12, 24, 46
 Murray-Rust, Peter 97
 MyExperiment 186
 MySpace 203

 Nahotko, Marck 55
 NASA 21, 192
 National Aeronautics and Space Administration Patrz NASA
 National Digital Information Infrastructure and Preservation Program Patrz NDIPP
 National Institute of Health, USA Patrz NIH National Library of Medicine (USA) Patrz NLM
 National Science Digital Library Patrz NSDL
 National Science Foundation Patrz NSF
 Nauka 2.0 17, 48, 127, 186, 197, 203, 207
 Nauka obliczeniowa 35, 42-45, 97
 Nawigacja 59, 69, 93, 101, 105, 131, 139, 143, 155, 165, 173, 199
 Nazwa uniwersalna 134
 NDIPP 189
 Negroponte, Nicholas 11
 Nelson, Theodor 19, 118, 144, 222
 Nentwich, Michael 54, 113
 Net Generation Patrz Pokolenie Sieci
 Newton, Isaac 28, 31
 Nicholson, Scott 200
 Nieformalne kanały komunikacji 91
 Nielsen, Jakob 130
 Niewidzialne uniwersytety 53, 54, 154, 236
 Niewidzialny Web Patrz Ukryty Web
 NLM 21, 79
 Noosfera 178, 179
 North-Holland 38
 Nowe rodzaje publikacji 87, 97, 143

NSDL 21
 NSF 20, 21, 43
 Nukat 176, 179

OAI 77, 146
 OAI-ORE 85
 OAI-PMH 79, 80, 85, 179
 OAI 192, 193
 Obiektywne języki programowania 200
 Obiekty cyfrowe 13, 58, 85, 105, 124, 146, 149, 176, 189, 218, 239
 Obiekty dynamiczne 82, 84, 87, 143, 144
 Obiekty multimedialne Patrz Multimedia
 Obiekty transgraniczne 60
 Obligatoryjny Open Access 93, 219, 224
 Oblinger, Diana 210
 Oblinger, James 210
 Obserwacja (met. bad.) 16, 48, 50, 51, 199, 209
 Ocena funkcjonalna systemu 128, 130
 Ocena funkcjonalna wyszukiwania 129, 132
 Ocena heurystyczna interfejsu 130
 Ocena holistyczna potrzeb użytkowników 200
 Ocena holistyczna serwisów GBC 200
 Ocena holistyczna uczonych 194
 Ocena jakości publikacji 196, 219
 Ocena oceniających 196
 Ocena serwisów GBC 193, 198
 OCF 90
 OCLC 61, 72, 112, 159, 167, 176, 177, 180, 210
 Odległość semantyczna 151
 Odlyzko, Andrew 36
 Odnośniki 19, 60, 73, 74, 84, 95, 115, 117, 130, 133, 135-137, 165, 170, 173, 222, 224
 Ogburn, William 14
 Ograniczenia ontologii 102
 Oldenburg, Henry 31, 195
 Online Public Access Catalog Patrz OPAC
 Ontologia dziedziny 100, 102-104
 Ontologie 63, 102, 107, 152, 175, 188, 201
 OPAC 56, 58, 61, 63, 76, 94, 113, 136, 155, 158, 161-164, 166, 177, 180, 187, 204, 206, 214, 245
 Open Access 13, 17, 34, 40, 41, 55, 63, 89, 92, 114, 136, 146, 147, 155, 167, 170, 176, 180, 197, 203, 210, 213, 216, 218, 219, 221, 223, 227, 229, 243
 Open Archival Information Systems Patrz OAIS
 Open Archives Initiative Patrz OAI
 Open Container Format Patrz OCF
 Open Data Commons 164, 229
 Open eBook Forum Patrz IDPF
 Open Library 245
 Open Packaging Forma Patrz OPF
 Open Publication Structure Patrz OPS
 Open Systems Interconnection Patrz OSI
 OpenID 137, 245
 OpenSearch 77, 79
 OpenURL 136, 137, 176, 179
 OPF 90
 Oppenheim, Charles 238
 OPS 90
 Organizacja informacji 100, 215

OSI 77
 Osińska, Veslava 140
 Osobiste biblioteki cyfrowe 242, 245
 Ostrom, Elinor 47, 228
 Otwarta nauka 28, 47, 186, 203
 Owen, John Mackenzie 29, 31, 82, 182, 233
 OWL 85, 201

PageRank 126, 170
 PAL 71
 Palmer, Carole 100
 Pamiętnik Historyczno-Polityczny 31
 Paradygmat konstruktywistyczny 234, 236
 Paradygmat sieciowy 12, 24
 PDDL 164, 229
 PDF 68-70, 83, 98, 114, 151
 Pełnotekstowe bazy danych 88, 98, 114, 158-160, 165
 Percepcja informacji 140, 143
 Pérez-Carballo, José 152
 Pergamon Press Personalizacja 21, 98, 164, 247
 Pertynencja 129, 130, 167, 193
 Philosophical Transactions of the Royal Society of London 31
 Physics Abstracts 36
 Pikas, Christina 15
 Pindłowa, Wanda 55
 Plagiat 182, 207
 Plemiona uczonych 52
 Pliki odwrócone 156-159
 Pliki sygnaturowe 157
 PLoS 41, 88, 147, 197
 Podcast 91, 204
 Podpis cyfrowy 182, 183
 Pokolenie ekranowe 210
 Pokolenie Google 210, 214
 Pokolenie Sieci 210
 Pokolenie X 210
 Polanyi, Michael 49
 Polskie Towarzystwo Matematyczne 36
 Pomerantz, Jeffrey 57, 121, 125
 Pool, Ithiel Solo de 227
 Popper, Karl 50
 Portale biblioteczne 164, 165, 177, 187, 215
 Portale internetowe 113, 158, 164, 222
 Portale proste 165
 Portale zintegrowane 165
 PostScript 69, 151
 Poszukiwanie informacji 131, 152-154, 167, 173, 208, 242
 Potrzeby informacyjne użytkowników 60, 114, 153, 155, 173, 176, 181, 241
 Potrzeby użytkowników 21, 25, 56, 85, 86, 101, 116, 124, 127, 128, 130, 152, 165, 185, 199, 248
 Powerset 175
 Poynder, Richard 224
 Poziom archiwizacji 190
 Poznańskie Towarzystwo Przyjaciół Nauk 30
 Prawo autorskie 40, 41, 71, 89, 213, 224, 234
 Prawo Gordona Moore'a 42
 Prawo opóźnienia kulturowego 14, 143

Prawo własności intelektualnej 93, 119, 199, 224
 Prawo zwyczajowe 224
 Prensky, Marc 210
 Prenumerata czasopism 32, 39, 88, 136, 148, 191, 210, 219, 221-223
 Preprint 91, 93, 154, 211, 227, 235, 236
 Prezentacja informacji 59, 60, 74-76, 78, 123, 138, 155, 179, 202
 Price, Derck Solla de 35, 53
 Produkcja informacji cyfrowej 51
 Profil użytkownika 139, 186, 209, 233
 Profile aplikacyjne 108-110, 134, 184
 Programmable Search Engine 172
 Projekt Gutenberg 89, 92, 113
 Proste obiekty cyfrowe 80
 Prosument 208, 233
 Protein Information Resource 96
 Protokoły sieciowe 117, 133, 188, 199, 201
 Proweniencja obiektu cyfrowego 52, 85, 106, 183, 226
 Prowinccje wiedzy 216
 Próchnicka, Maria 153, 173
 Przeglądanie 32, 37, 95, 139, 156, 161, 173
 Przeglądanie fasetowe 174
 Przeglądowe bazy danych 114
 Przeladowanie informacją 25
 Przenoszalnosc oprogramowania 188
 Przechowanie nazw 18, 108, 137, 188, 201, 246
 Przestrzeń informacyjna 102, 117, 122, 123
 Przestrzeń osobista 124
 Przetwarzanie gróidowe 181
 Przetwarzanie języka naturalnego 150, 152, 247
 Przetwarzanie rozproszone 65, 181, 247
 PSIGate 113
 Public Domain Dedication & License Patrz PDDL
 Public Library of Science Patrz PLoS
 Publikacje elektroniczne 81, 99, 142, 145, 218-220
 Publikowanie naukowe 32, 37, 92, 197, 223
 Publikowanie nieprzerwane 145
 Publikowanie semantyczne 100

QuickTime 72

Radomskie Towarzystwo Naukowe 30
 Rajendiran, P. 91
 RAM 42
 Raporty naukowo-techniczne 30, 87, 91, 97, 207, 219, 228, 236
 RDA 109-111
 RDF 73, 76, 108, 111, 118, 134, 137, 201
 RDN 113, 164
 Realizacje (działa) 81, 83, 85, 135
 Reblogging 203
 Recenzenci 13, 32, 62, 126, 145, 195, 204, 213, 233, 235
 Recenzowanie 40, 99, 145, 147, 193-195, 214, 219, 234,
 Recenzowanie otwarte 195, 214
 Redaktor 13, 30, 32, 62, 145, 195, 204, 223
 Reddy, Michael 231
 Reguły wnioskowania 201

Rejestry metadanych 108, 152
 Relacyjny wskaźnik cytowań 127
 Relevance 125, 128, 129, 163, 166, 170, 193, 244
 Relevanceja pragmatyczna Patrz Pertynencja
 Relevanceja techniczna
 Renear, Allen 100
 RePEc 113, 160, 196
 Replikacja 191, 192
 Repozytoria centralne 216
 Repozytoria dziedzinowe 41, 196
 Repozytoria instytucjonalne 41, 59, 189, 192, 198
 Repozytoria obiektów cyfrowych 15, 17, 22, 35, 58, 79, 85, 87, 93, 94, 107, 113, 147, 154, 179, 212, 228, 233, 236, 238, 241, 243
 Reprezentacja cyfrowa materiałów drukowanych 158
 Reprezentacja informacji 101, 143, 200, 231
 Reprezentacja treści 66, 169
 Reprezentacja wiedzy 49-52, 102, 113, 152
 Reprezentacja wizualna informacji 140
 Reputacja czasopisma 32, 38, 88, 194, 213
 Reputacja instytucji naukowej 33, 87, 194
 Reputacja uczonego 33, 87, 194, 207, 213, 225
 Reputacja wydawcy 87, 194
 Resolver 136, 137, 176, 179
 Resource Discovery Network Patrz RDN
 Resources Description and Access Patrz RDA
 Rewolucja naukowa 28-30, 51, 204
 Rich Site Summary Patrz RSS
 Richelieu, Armand 30
 Rifkin, Jeremy 223
 Rogers, Everett 14
 Role biblioteczny 13, 15, 37, 41, 58, 61, 62, 122, 148, 154, 185, 190, 214-217
 Role uczonych 13, 33, 56, 62, 186, 202, 208-211
 Role wydawców 13, 33, 38, 148, 218-220
 Rothenberg, Jeff 189
 Rozwiązywanie problemów 25, 140, 153
 RSS 79, 100, 145, 162, 165, 202, 204, 233
 RTF 70
 Ruchy komputeryzacyjne 248
 Rusbridge, Chris 206
 Rynek wydawniczy 37, 207, 218, 220
 Rzepa, Henry 97

Sage 191
 Sallo, Denis de 30
 Samoorganizujące się sieci 59, 118, 246
 Samoorganizujące się społeczności 247
 Samoorganizujące się struktury 117
 Samoorganizujące się systemy 60
 Samoorganizujące się topologie 118
 Samopublikowanie 98, 208, 212, 233, 235
 Sapa, Remigiusz 132
 Saracovic, Tefko 125, 198
 Satysfakcja użytkownika 129, 131, 138
 Saunders, Pam 90
 Scalable Vector Graphics Patrz SVG
 Schematy metadanych 77, 104, 108, 110, 114, 134, 150, 188
 Scholar Factor Patrz SF

Scholarly skywriting 197
 Scholarly Works Application Profile Patrz SWAP
 SCI 160
 Science Citation Index Patrz SCI
 Science Commons 229
 ScienceDirect 114, 161
 SciFinder Scholar 161
 SciLink 186
 Scirus 160, 169
 Scitation 161
 Scivee 186
 SCOAP 13, 223
 Scopus 161, 171
 Search engine optimization Patrz SEO
 Selekcja materiałów 24, 41, 141, 145, 146, 169, 173, 185, 186, 193
 Semantyczny Grid 53
 Semantyczny Web 17, 53, 62, 85, 97, 100, 111, 118, 172, 175, 247
 SEO 131
 Serial Item and Contribution Identifier Patrz SIC1
 Serwer zapytań 115
 Serwisy apomediacyjne 186
 Serwisy GBC 40, 56, 59-61, 64, 65, 87, 101, 105-107, 114, 121-125, 127-132, 136-140, 155, 158-162, 174-179, 197-200, 206, 215, 233, 243-245
 Serwisy społecznościowe Patrz Web 2.0
 SF 126
 SFX 136, 137
 SGML 73-75, 150
 Shannon, Claude 49, 50, 231
 Shibboleth 187
 Shotton, David 205
 SIC1 135
 Sieci interpersonalne 154
 Sieci małych światów 239
 Sieci semantyczne 102
 Sieci społeczne 45, 60, 91, 118, 143, 202
 Sieci społeczne uczonych 238
 Sienkiewicz, Piotr 42
 Sienko, Marcin 14
 Singapore Framework 110
 Sirsi/Dynix 163
 SIW 56, 102, 129, 154, 158, 193, 199, 201, 244
 Skalowalność 213
 Skywriting 99, 197
 Słowa kluczowe 25, 49, 109, 112, 114, 124, 151, 161, 167, 174, 196
 Słownik reprezentacji wiedzy 103
 Słowniki kontrolowane 103, 163, 175, 201
 SOAP 79
 Social OPAC Patrz SOPAC
 Social Science Citation Index Patrz SSCI
 Söderqvist, Jan 203
 Soergel, Dagobert 22
 Sompel, Herbert 85, 93
 Søndergaard, Tritne 90
 SOPAC 162, 204
 SOSIG 113
 SOW 100, 102, 124, 151, 152, 230
 Spam 115, 168, 172
 Spires-IIEP 161
 Społeczeństwo informacyjne 12, 17, 42, 248
 Społeczeństwo sieciowe 12, 120, 173, 214
 Społeczeństwo wiedzy 248
 Społeczność wiedzy 236
 Sprawność kosztowa systemu 130, 146
 Springer 32, 191
 SpringerLink 114
 SRU 78, 79
 SRU/W 78, 79
 SRW 78, 79
 SSCI 160
 SSO 186
 SSRN 196
 Standard Generalized Markup Language Patrz SGML
 Stoll, Clifford. 242
 Stop-lista 151, 157, 158
 Strategie archiwizacji 191, 192
 Strategie Open Access 147
 Strategie wyszukiwawcze 132, 170
 Strona zbiorcza 84
 Strony Web 74, 87, 94, 120, 164, 211, 223
 Struktura dokumentu 73-75, 151
 Struktura odnośników 117, 173
 Strumień kliknięć 170, 172
 Suber, Peter 217
 Subito 221
 Sun 72
 SUTRS 78
 SVG 68, 71
 SWAP 110
 System informacyjno-wyszukiwawczy Patrz SIW
 System komunikacji naukowej 28, 32, 41, 56, 125, 142, 147, 207, 233, 238
 System konceptualny 122
 System mediacyjny 175
 Systemy biblioteczne 124, 161, 163, 187, 230
 Systemy biblioteki cyfrowej 22, 23, 59
 Systemy ekspertowe 169
 Systemy inteligentne 103, 174, 200
 Systemy Organizacji Wiedzy Patrz SOW
 Systemy społecznościowe Patrz Web 2.0
 Systemy zarządzania biblioteką cyfrową 22, 23
 Szara literatura 15, 20, 87, 90-92, 195
 Sześć stopni oddzielenia 239
 Sztuczna inteligencja 20, 52, 102, 200, 204
 Śniadecki, Jędrzej 28
 Światowy katalog książek 245
 Tablica kodowania mieszanego 157
 Tadeusiewicz, Ryszard 27
 Tagi Patrz Etykiety
 Taksonomie 102, 103, 201
 TCP 78
 TCP/IP 117, 139
 Techniki wyszukiwania informacji 35, 86, 156, 158, 160, 167, 172-176, 181, 245
 Technologia informacyjna Patrz ICT

- Technologiczne krzywe S (teoria) 26
 Technosfera 14
 Teilhard de Chardin, Pierre 179
 Tekst otwarty 66
 Tenopir, Carol 38
 Teoria gier 247
 Teoria konwersacji 122
 Teoria neuroplastyczności 210
 Tezaurusy 61, 102-104, 152, 166, 246
 The Royal Society 30, 31, 53
 Thomson Scientific 32, 126, 159-161, 196
 Thorin, Suzanne 52
 TIFF 68, 70
 Toffler, Alvin 233
 Tomasz z Akwinu 19
 Tonkin, Emma 151
 Topologia Webu 117
 Towarzystwo Biblioteki Słuchaczy Prawa Uniwersyte-
 tu Jagiellońskiego 30
 Towarzystwo Naukowe Krakowskie 30
 Towarzystwo Naukowe Płockie 30
 Towarzystwo Naukowe w Toruniu 30
 Towarzystwo Warszawskie Przyjaciół Nauk 30
 Tradycyjna teoria wyszukiwania 153
 Transformacja informacji 242, 245
 Transkluzja 222
 Transmission Control Protocol Patrz TCP
 Transmission Control Protocol/Internet Protocol Patrz
 TCP/IP
 Turing, Alan 19
 Typy MIME Patrz MIME
- Ukryty Web 80, 90, 94, 107, 112, 114, 166, 168, 171,
 178
 Ulrichs 36, 136
 Unicode 66
 Universal Bibliographic Control 20
 Uniwersalizm 34, 55
 Uniwersalna Książka 19
 URI 49, 84, 108, 133, 137
 URL 27, 78, 114, 134-137, 164, 190
 Urządzenia mobilne 10, 12, 64, 75, 99, 139
 Usługi Gridowe 53
 Usługi serwisu GBC 49, 57, 138, 200, 243
 Usługi techniczne 57
 Usługi użytkowników 57
 Uwierzytliczenie Patrz Weryfikacja autorstwa
 Użyteczność serwisu GBC 130, 131
 Użytkownicy aktywni 234
 Użytkownicy bibliotek 14, 88, 101, 123, 125, 136, 177,
 223
 Użytkownicy bibliotek cyfrowych 24, 59, 177
 Użytkownicy GBC 18, 86, 105, 120-123, 130, 163, 173,
 199, 206, 217, 245
 Użytkownicy informacji 13, 50, 98, 130, 138, 167, 175,
 185
 Użytkownicy instytucjonalni 209, 222
 Użytkownicy Internetu 65, 70, 100, 125, 170, 208
 Użytkownicy jednocześnie 131
- Użytkownicy końcowi 43, 46, 69, 137, 146, 158, 198,
 212, 223, 231, 243
 Użytkownicy niewidomi 86, 95
 Użytkownicy pasywni 202
 Użytkownicy potencjalni 222
 Użytkownicy rdzenni cyfrowi 210
 Użytkownicy Webu 84, 115, 117, 171, 174, 202, 205
- Varian, Hal 51
 Vetulani, Agnieszka 228
 Vickery, Brian 236
 VITAL 112
 Vivisimo 163
 Vodcast 91
 VTLS 112, 162
- W3C 76, 79, 196, 200
 Waga wpływu 127
 Walter, Anne-Kathrin 171
 Wang, Bing 22
 Wartość dodana agregatorów 220
 Wartość dodana biblioteki 61, 122
 Wartość dodana dokumentów elektronicznych 95, 98
 Wartość dodana serwisu GBC 50, 57
 Wartość dodana wydawców 217, 219
 Warwick Framework 76, 112
 WAV 68, 72
 Wayback Machine 190
 Weaver, Warren 31, 49, 50, 231
 Web 2.0 13, 17, 45, 48, 87, 99-101, 124, 143, 164, 170,
 181, 196, 230
 Web 3.0 205, 247
 Web Impact Factor Patrz WIF
 Web of Science 160, 161, 171
 Web Ontology Language Patrz OWL
 Web społecznościowy Patrz Web 2.0
 Weblog Patrz Blogi
 Webometria 125, 127, 196
 Weinberger, David 93
 Wersjonowanie obiektów cyfrowych 82
 Wersjonowanie publikacji 92
 Weryfikacja autentyczności 82, 182, 183
 Weryfikacja autorstwa 181
 Weryfikacja teorii naukowej 50
 Whitworth, Brian 233
 Wiedza jawna 49, 50, 103, 201
 Wiedza konceptualna 155
 Wiedza obiektywna 103
 Wiedza semantyczna 155
 Wiedza syntaktyczna 155
 Wiedza ukryta 49, 92, 103, 175
 Wielka nauka 35, 241, 247
 Wieloźyczny dostęp do informacji 107, 130, 185
 Wicłoznacznik 156
 WIF 126
 Wiki 48, 91, 99, 127, 202-204
 Wikia Search 181
 Wikipedia 175, 177, 202
 Wilensky, Robert 57

Willinsky, John 212, 227
 Wilson H.W. (wydawnictwo) 159
 Wirtualne instytuty 45, 54
 Wirtualne laboratoria 44, 53, 54, 62
 Wizualizacja danych 35, 48, 52, 71, 96, 107, 140, 143
 Wizualizacja informacji 19, 140
 Wizualizacja treści 58, 74, 151
 Wojcicchowski, Jacek 10, 58, 210
 Wolniewicz, Tomasz 199
 WorldCat 61, 139, 176, 177, 179
 Woźniak-Kasperek, Jadwiga 100
 Wskaźnik cytowań 86, 126, 127
 Wskaźnik odbicia 127
 Wskaźnik popularności 127
 Wskaźnik starzenia się czasopisma 127
 Wskaźnik szybkości oddziaływania 127
 Wskaźnik wpływu Patrz Impact Factor
 Wskaźnik znaczenia czasopisma 127
 Wskaźniki efektywności sprzętu komputerowego 42
 Wskaźniki funkcjonowania serwisów 129-132
 Wskaźniki popularności 163
 Wskaźniki rozwoju nauki 36, 46, 51
 Wspólne filtrowanie 203
 Współdziałanie 18, 22, 42, 45, 47, 53, 73-80, 85, 103, 111, 130, 135, 175, 188, 199, 202, 242, 246
 Współdziałanie międzynarodowe 188
 Współdziałanie polityczne/społeczne 119, 188
 Współdziałanie prawne 188
 Współdziałanie semantyczne 107, 108, 188
 Współdziałanie techniczne 188
 Współpraca uczonych 17, 21, 34, 35, 45, 51-55, 62, 88, 207, 212, 237
 Wulf, William 53
 Wydawcy 13, 33, 40, 45, 87, 90, 106, 142, 147, 176, 190, 194, 207, 212, 226, 232, 233, 246
 Wydawcy czasopism 87, 88, 92, 159
 Wydawcy komercyjni 31, 32, 38, 39, 41, 55, 90, 93, 114, 170, 217-223, 227-229, 243
 Wydawcy niekomercyjni 39, 41, 62, 220
 Wydawcy pseudokomercyjni 38, 39, 220
 Wydawcy uczelniani 32, 38
 Wydobywanie wiedzy z baz danych 52, 75
 Wykluczenia 48, 224
 Wymiana danych 42, 54, 57, 60, 106-108, 184, 188
 Wymiana informacji 12, 16, 27, 187
 WYSIWYG 69
 Wyszukiwanie aktywne 153
 Wyszukiwanie automatyczne 19, 160
 Wyszukiwanie badawcze 156, 171, 210
 Wyszukiwanie ciągłe 153
 Wyszukiwanie manualne 160
 Wyszukiwanie otwarte 181
 Wyszukiwanie pasywne 153
 Wyszukiwanie pełnotekstowe 110, 160, 169, 171, 179
 Wyszukiwanie porównawcze 175
 Wyszukiwanie proste 139, 162, 169
 Wyszukiwanie rozmyte 175
 Wyszukiwanie sekwencyjne 156
 Wyszukiwanie sfederowane 165-167, 175, 179, 181, 245
 Wyszukiwanie uniwersalne 25, 95, 181, 245
 Wyszukiwanie wielojęzyczne 185, 247
 Wyszukiwanie wspomagane 169
 Wyszukiwanie wspomagane ontologią 174
 Wyszukiwanie wspólne 154, 208
 Wyszukiwanie złożone 115, 139, 162, 169
 Wyszukiwarki globalne 106, 124, 126, 149, 164, 167, 169-171, 176, 211, 240, 245
 Wyszukiwarki internetowe 57, 61, 68, 75, 80, 94, 98, 106-109, 114-116, 129, 138, 150, 162, 164-174, 176, 190, 239, 244
 Wyszukiwarki semantyczne 175
 Wyszukiwarki sfederowane Patrz Metawyszukiwarki
 Wyszukiwarki tematyczne 169
 Wyszukiwarki wyspecjalizowane 169, 171
 Wzorce zachowań 154, 172, 217
 Wzorce zapytań 115

 Xanadu 118, 144, 222
 XHTML 75, 111, 139, 150
 XLink 137
 XML 48, 66, 70, 73-80, 85, 97, 100, 109, 111-113, 133, 151, 200-202
 XSD 75

 YADDA (repozytorium) 245
 Yahoo 115, 130, 166, 169, 171, 177
 Yang, Kiduk 171
 Young, Mark 38
 YouTube 87, 91, 202

 Z39.50 77-80, 161
 Zachowania informacyjne użytkowników 19, 123, 153, 154, 210
 Zalety publikacji elektronicznych 86
 Zalew danych 31, 35, 41, 52
 Zarządzanie projektami 101, 215
 Zasada 90-9-1 127
 Zasięg praw autorskich 225
 Zbiór Tygodniowy Wiadomości Uczonych 31
 Zestawy danych 83, 87, 96, 98, 219
 Zielony Open Access 147
 Zintegrowane systemy biblioteczne 109, 161, 163
 Złoty Open Access 147
 Złożone obiekty cyfrowe 18, 58, 67, 80-85, 95, 104, 239
 Zotero 127
 Zróżnicowanie ról uczonych 62, 209, 232

SUMMARY

Scholarly communication in digital environment. Global digital library in e-Research.

In the past few years the Internet has become an integral part of scientific and academic life. Observations made for last years show that it is precisely here locates the mainstream of informal and increasingly formal scholarly communication. Both scientists and students establish and maintain interpersonal relations through Internet-based services, which have tremendous geographic scope widened. Constantly growing collections of high quality network of scientific publications and, consequently, Internet information resources are becoming an increasingly important source of data, information and knowledge for the purposes of research and academic education. Habit of everyday usage of the Internet as a communication channel and its digital resources in research work deepens the ubiquity of this medium in daily activities: the use of information management in education and research institutions, the search for address data, the use of dictionaries and encyclopaedias, organization of leisure and social gatherings, finding information needed in everyday life. As aptly Christine Borgman stated on one of the most important in the information science and librarianship books of recent years - *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (MIT Press, 2007) - in a short period of time for researchers and students the Internet communication environment changed from which they had login from time to time, to become a communications environment which is constantly active.

In the development of any complex technologies such as Internet-related ones, crucial decisions are taken in the first phase of its design. Observing changes in the practice of scientific activities, which are a consequence of the development of modern information technologies (ICT), we should be aware about what technologies are now being developed, why and to whom they serve and how they are used, as it will have a great influence on the forms and principles of science and research activities to be shaped in the future. This applies to a great degree of scholarly communication channels and centres of its organization, which were previously research libraries and centres of scientific information. Investment in so-called Cyberinfrastructure or e-Science, like all modern information infrastructures is global, and therefore gradually being shaped today, in the future it will be difficult to change. At the same time it is worth remember that the development of ICT is still essentially uncontrolled nature and is much faster than the development of social perception and understanding of its current and future consequences. In this context, the research and discussion, which deepen understanding and help to formulate directives for the optimal development and implementation of advanced information technologies appropriate to the needs of users applying them, has a special value. Such research is presented in this book.

Author of the book analyzes the transformation of scholarly communication, driven by information and communication technologies. He shows the change in attitudes, behaviour and activities of the scientific community, in particular relating to finding the scientific information and dissemination of research results. He also discusses the transformation of the principles and scope of activities of the institutions already involved in the organization of scholarly communication, which were triggered by new information technology and the Internet in particular. The analysis of contemporary forms of scholarly communication system is a documentation of the view that the current result of the development of ICT tools is the hybridization of technology in use, which is the transition state of the system, which seeks to replace the existing technology by tools of networked information technology. The technology, becoming a basic infrastructure of social life of the future, implies profound changes in organization and methods of activity in different areas of life, and therefore

also in the way of research activities. To these changes must be adapted new methods and forms of organization of scholarly communication. At the heart of these changes are research libraries, which were until now the main institutions responsible for maintaining the memory of science and providing access to it. Author argues that their activities may become a functional part of a complex structure, called by him a global digital library (GDL), which will be the integrated information environment of science.

The book presents the results of a study on three key topic areas for discussion: 1) electronic publishing and its new forms, including mainly digital libraries, repositories, electronic journals, self-archiving, scholarly blogs, and gray literature, 2) the role of metadata in the description of electronic documents for their efficient search and management, 3) changes in scholarly communication roles of participants, mainly librarians, and ontological modelling processes for scholarly communication.

Author set himself two goals: to know the current problems and trends in the electronic scholarly communications, formal and informal, and developing the aforementioned model of scholarly communication, called a global digital library (GDL), in the global information infrastructure of science, taking into account the latest trends and phenomena, resulting from the implementation of ICT.

The author presents in the book the scholarly communication as "a set of formal and informal processes, for to share knowledge, the newly-created as a result of the scholar's own research, based on observation and verifiable facts, which creates a platform for scientific discourse, by documenting, assessing, editing, formatting, distribution, organizing, sharing, archiving, use and transformation of that knowledge". Indisputable is the author's conviction that the analysis and modelling of these phenomena are a prerequisite for adaptation of scholarly communication to current and future needs of scientific research. Five basic arguments in the book also feature the essence of changes taking place in the system of scholarly communication and set the frame for the discussion:

1. The future of scholarly communication is a digital environment, but for its full exploitation it is necessary to solve many problems (e.g., improvement of methods of long-term archiving);
2. Digitization of scholarly communication will bring significant changes in its organization;
3. Changes in scholarly communication are associated with changes in the organization of science, particularly with the construction of e-Science infrastructure and so-called Science 2.0, based on Web 2.0;
4. Change will include formation of uniform (in terms of meeting the users needs) information resources that underpin the global digital library (GDL);
5. As a result of digitization of scholarly communication are blurred differences between formal and informal communication, articulated in the GDL.

These allegations are verified by answering the following research questions:

1. What kind of changes is caused by the implementation process of scholarly communication in the digital environment?
2. Whether a global digital library as a tool for the implementation of scholarly communication, will allow the unification of different scientific publishing practices, leading to a combination of highly formalized practice of reviewed publishing and more informal and personalized forms of communication, based on various and internal rules adopted by the scientific environment?
3. What is the place of information and library technologies, such as metadata, ontologies, identifiers in the future model of scholarly communication?
4. What is the future of electronic publications and how to search it in global networks?

The considerations are divided into two parts: introducing, in which the author discussed the foundations of global digital library as a model of scholarly communication for complex e-Science infrastructures and exemplification part, where detailed functioning of GDL is presented, using the 5S metamodel of digital library. In this model, digital libraries are analyzed in five aspects: Structures, Streams, Spaces, Scenarios of processes, and the communities to be supported (Societies).

Introducing part forms the first chapter, which presents the functional principles of modern science and its e-Science infrastructure and the author's concept of a global digital library, and assumptions about its place as a communication tool in the scientific infrastructure. Exemplification part compose of the five remaining chapters, dealing successively GDL structures (systems and markup languages, data formats, protocols, digital objects, electronic publications, organization of knowledge tools and metadata); GDL functioning spaces (a metaphor for the Internet as a public digital space, as well a metaphor GDL as place in this space, evaluation criteria and measures used in

this space, objects identification, and user's communication interfaces); GDL processes (publishing, metadata creation, searching and retrieving, ensuring integrity and authenticity of digital objects, the availability of information and authorizing access, interoperability, long-term archiving, quality control, and new developments in the access organization to digital content - Semantic Web, social networks and the vision of its integration in the form of so-called Web 3.0); GDL communities (various categories of individual and institutional users as well as economic and legal aspects of creating and using a global digital library). In the last part was presented a model of scholarly communication in the GDL, which is an assessment of the actual formation of global online digital library structures in contemporary Internet. Author presented also models of the future evolution of GDL and its implications for contemporary structures of scholarly communication organization.



**Seria wydawana przez Wydawnictwo
STOWARZYSZENIA BIBLIOTEKARZY POLSKICH
we współpracy
Z INSTYTUTEM INFORMACJI NAUKOWEJ
I STUDIÓW BIBLIOLOGICZNYCH UNIWERSYTETU
WARSZAWSKIEGO**