NAUKA – DYDAKTYKA – PRAKTYKA

Aneta Januszko-Szakiel

ARCHIWISTYKA CYFROWA

Długoterminowa ochrona dziedzictwa nauki i kultury

WYDAWNICTWO SBP

Archiwistyka cyfrowa

Digital Archiving

SBP STOWARZYSZENIE BIBLIOTEKARZY POLSKICH 100 lat!

Polish Librarians Association
SCIENCE-DIDACTICS-PRACTICE

Aneta Januszko-Szakiel

Digital Archiving

Long-term preservation of scientific, scholarly and cultural heritage

Warsaw 2017

Stowarzyszenie Bibliotekarzy Polskich
NAUKA-DYDAKTYKA-PRAKTYKA

Aneta Januszko-Szakiel

Archiwistyka cyfrowa

Długoterminowa ochrona dziedzictwa nauki i kultury

Warszawa 2017

Publikacja dofinansowana przez
Instytut Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego

Recenzent
Prof. dr hab. Barbara SOSIŃSKA-KALATA

Redakcja techniczna i korekta
Elżbieta MATUSIAK

Projekt okładki
Piotr GÓRSKI

Autor zdjęcia na okładce
© envfx – fotolia.com

ISBN 978-83-65741-02-8

CIP – Biblioteka Narodowa
Januszko-Szakiel, Aneta
Archiwistyka cyfrowa : długoterminowa
ochrona dziedzictwa nauki i kultury / Aneta
Januszko-Szakiel. – Warszawa : Wydawnictwo
Stowarzyszenia Bibliotekarzy Polskich, 2017.
- (Nauka, Dydaktyka, Praktyka ; nr 181)

Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich
00-335 Warszawa, ul. Konopczyńskiego 5/7, tel. 22 827 52 96
www.sbp.pl; wydawnictwo@sbp.pl, biuro@sbp.pl
Warszawa 2017. Wyd. I. Ark. wyd. 17. Ark. druk. 15
Łamanie: Piotr Górski
Druk i oprawa: Mazowieckie Centrum Poligrafii
ul. Piłsudskiego 2A, 05-270 Marki
www.c-p.com.pl, e-mail: biuro@c-p.com.pl, tel. 22 497 66 55

Spis treści

Contents

Wstęp

Jest rok 2045. Wnuki Jeffa Rothenberga[1] odnajdują na strychu odręcznie napisany list z 1995 r. wraz z załączoną płytą CD-ROM. W liście Rothenberg powiadamia wnuki, iż CD-ROM zawiera zakodowaną informację o pozostawionym przez niego spadku i wyjaśnia, w jaki sposób należy ją rozszyfrować. Odkrycie, choć uszczęśliwiające, może okazać się kłopotliwe, bowiem – jak przypuszcza autor wizji – wnuki nie miały okazji widzieć wcześniej małego srebrnego krążka – chyba że w starych filmach. Nawet jeśli uda się im znaleźć odpowiedni czytnik CD, który dokona konwersji rozmieszczonych na krążku wgłębień na ciąg bitów, to kolejnym celem poszukiwań musi stać się program do ich prezentacji [Rothenberg, 1995, s. 42-47]. Pytanie, czy wnukom Rothenberga uda się skorzystać z nieoczekiwanego bogactwa, pozostaje otwarte.

Fikcyjny scenariusz uzmysławia zasadnicze słabości cyfrowego zapisu treści wobec zapisu tradycyjnego. Zdaniem Rothenberga, nawet za pięćdziesiąt i więcej lat napisany ręcznie list będzie można odczytać bezpośrednio, podczas gdy odczyt zapisów elektronicznych będzie znacznie utrudniony lub niemożliwy z racji szybkiego tempa rozwoju w dziedzinie sprzętu i oprogramowania komputerowego. Wszelkie cyfrowe zapisy obecnego pokolenia, tak chętnie i masowo użytkowane, ulegną dużo szybszemu zniszczeniu aniżeli te na papierze. Zawartość mediów cyfrowych staje się niemożliwa do odtworzenia znacznie szybciej aniżeli słowa zapisane dobrym tuszem na dobrym papierze. Do utraty danych cyfrowych dochodzi najczęściej nie tyle z powodu fizycznego zniszczenia nośnika, co z racji

[1] Jeff Rothenberg – informatyk korporacji RAND w Kalifornii. Autor wielu publikacji na temat długoterminowego archiwizowania publikacji elektronicznych. Prekursor metody emulacji. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington, 1999; Jeff Rothenberg's Professional Website: http://www.panix.com/~jeffr/Prof/prof.html [Dostęp: 10.07.2017].

wprowadzania wciąż nowych, niekompatybilnych z wcześniejszymi, generacji nośników i dedykowanych im urządzeń odczytu danych. Najlepszym tego przykładem jest zastąpienie dyskietki 8-calowej kolejnymi generacjami nośników danych cyfrowych. Dla celów długoterminowej archiwizacji konieczne jest przechowywanie, wraz z dokumentem elektronicznym, informacji o wymaganym do jego odczytu otoczeniu sprzętowo-programowym. Najczęściej informacje te zawiera dokument wydrukowany na papierze i dołączony do cyfrowego medium – dokładnie jak w scenariuszu Rothenberga; przewidujący i troskliwy dziadek pozostawia wnukom papierowy list z opisem sposobu odkodowania informacji z krążka. Zdaniem autora naiwne jest myślenie, że teraźniejszy sposób cyfrowego kodowania treści będzie możliwy do odczytu za pomocą oprogramowania używanego w przyszłości. Technologie informacyjne będą wprowadzać rozwiązania, które prawdopodobnie nie będą ani kontynuacją rozwiązań dotychczasowych, ani nie powstaną na ich bazie. Będą to nowe, oryginalne produkty, umiejętnie wdrożone na rynek i rozpowszechnione. Dokładnie tak, jak w przypadku taśm magnetycznych, kaset wideo tudzież dyskietek magnetycznych, które z czasem zostały zastąpione dyskami optycznymi. Czy zatem w obliczu tych zmian nie należałoby zatroszczyć się o utrzymanie dostępu do treści zapisanych na mediach wychodzących z obiegu, w szczególności, gdy treści te są świadectwem istotnej naukowej i kulturalnej działalności człowieka? Czy brak świadomości dotyczącej starań o ich użyteczność nie doprowadzi do bezpowrotnej utraty cennych artefaktów? [Januszko-Szakiel, 2008, s. 121-130; Rothenberg, 1995, s. 42-47]. Wizja Rothenberga, roztoczona i opublikowana w 1995 r., stanowi asumpt do podjętych w niniejszej publikacji rozważań o archiwistyce cyfrowej i jej szczególnie trudnym zagadnieniu – trwałej ochronie dziedzictwa narodowego.

Tematem książki jest długoterminowa archiwizacja zasobów cyfrowych, postrzegana jako skomplikowane zadanie wpisujące się w spektrum działalności instytucji pamięci oraz rozszerzenie zadań archiwistów, bibliotekarzy i muzealników cyfrowych. *Długoterminowa archiwizacja* oznacza nie tylko gromadzenie i przechowanie materiału cyfrowego w długim czasie, ale przede wszystkim zapewnienie jego długoterminowej użyteczności, rozumianej jako dostępność, możliwość odczytania i prezentacji jego treści w formie zrozumiałej dla użytkownika, zarówno obecnie, jak i w najbardziej odległej przyszłości. Dodatkowo procesy ochrony służą zabezpieczeniu istotnych cech dokumentów cyfrowych, takich jak: autentyczność, integralność, poufność, a także zachowaniu informacji kontekstowych, czyli metadanych archiwizowanych dokumentów. Zabiegi te z jednej strony muszą przyczynić się do sprostania oczekiwaniom użytkowników, którzy (tak obecnie, jak i w przyszłości) będą chcieli zidentyfikować, wyszukać,

odczytać i interpretować interesujące ich zasoby informacji zapisane w formie cyfrowej. Będą oczekiwać, że te zasoby będą dostępne, kompletne oraz wiarygodne. Z drugiej strony archiwizacja zasobów cyfrowych musi też spełnić wymagania deponentów, którym zależy na długoterminowej użyteczności przekazanych w depozyt dokumentów.

Działania skoncentrowane wokół trwałej ochrony zasobów cyfrowych określa się mianem *archiwistyki cyfrowej*. Tocząca się w 2009 r. dyskusja [Goldenline, 2009] na temat zasadności stosowania terminu archiwistyka cyfrowa nie przyniosła jednoznacznych ustaleń. Termin przyjął się, został zdefiniowany i jest stosowany w ustnych oraz pisemnych wypowiedziach dotyczących gromadzenia, udostępniania i archiwizowania cyfrowych zasobów [Archiwistyka, 2007; Bończa-Tomaszewski, 2014; Supruniuk, b.d.]. Zgodnie z definicją zaproponowaną przez pracowników Narodowego Archiwum Cyfrowego, *archiwistyka cyfrowa* to „dziedzina zajmująca się zarządzaniem cyfrowymi materiałami archiwalnymi, szczególnie ich gromadzeniem, przechowywaniem i udostępnianiem. Do jej zadań należy m.in.: projektowanie, tworzenie i zarządzanie archiwalnymi systemami informatycznymi, infrastrukturą komputerową i archiwalnymi bazami danych; archiwizacja materiałów cyfrowych, w tym dokumentów elektronicznych, obiektów zdigitalizowanych, systemów bazodanowych i stron internetowych; digitalizacja analogowych materiałów archiwalnych i inne" [Dudek i Kowalska, 2010, s. 71]. *Archiwista cyfrowy* natomiast to „specjalista realizujący zadania związane z zarządzaniem cyfrowym materiałem archiwalnym" [Dudek i Kowalska, 2010, s. 72].

Pomimo że autor definicji odnosi to pojęcie do archiwów i archiwaliów, należy zauważyć, iż również inne instytucje pamięci, głównie biblioteki i muzea, oraz placówki o charakterze naukowym i badawczym, także prowadzące działalność kulturalną, podejmują działania na rzecz długotrwałego zabezpieczenia dziedzictwa narodowego. Opisane w książce strategie, metody oraz projekty długoterminowej archiwizacji, choć wyprowadzone ze studiów i badań głównie instytucji bibliotecznych, mogą w równym stopniu znajdować zastosowanie w innych instytucjach tworzących, gromadzących i zarządzających cyfrowymi zasobami nauki i kultury. Stosowanie zatem w tekście określenia *instytucje archiwizujące* bądź *instytucje pamięci* odnosi się do wszystkich podmiotów sektora nauki i kultury realizujących zadania zabezpieczenia narodowej pamięci w cyfrowej postaci.

Rozważania niniejsze odnoszą się głównie do dokumentów, które są świadectwem dorobku nauki i kultury, stanowią dziedzictwo narodowe, bywają określane jako *narodowy zasób cyfrowy*. Wciąż brakuje jednak zarówno jednoznacznej, uzgodnionej definicji tego terminu, jak i precyzyjnych kryteriów kwalifikowa-

nia obiektów cyfrowych do narodowego zasobu cyfrowego. W niniejszej książce proponuje się zatem ujęcie, w którym

> narodowy zasób cyfrowy to zbiór obiektów – wyrazów myśli i działań ludzkich, w formie publikowanych i niepublikowanych dokumentów, materiałów, eksponatów – zapisanych w postaci kodu zero-jedynkowego, powstających bądź w procesach publikowania elektronicznego, określanych mianem oryginalnych lub pierwotnych dokumentów cyfrowych (ang. *born digital objects*), bądź w procesach konwersji obiektów analogowych do postaci cyfrowej, określanych mianem obiektów zdigitalizowanych lub cyfrowych odwzorowań obiektów tradycyjnych (ang. *digitised objects*). Ich treść jest odczytywana przez elektroniczne urządzenia komputerowe.
>
> Z technicznego punktu widzenia zarówno obiekty cyfrowe oryginalne, jak i zdigitalizowane podlegają tym samym rygorom długoterminowej archiwizacji w repozytoriach, archiwach, depozytach cyfrowych. Ważne jest jednak uwzględnienie faktu, że obiekty cyfrowe oryginalne nie mają swoich odpowiedników (surogatów w postaci drukowanych, analogowych, tradycyjnych oryginałów), co jest domeną obiektów stanowiących cyfrowe odwzorowania. To powinno obligować archiwistów cyfrowych do zapewnienia oryginalnym obiektom cyfrowym szczególnej ochrony. W świecie znane są praktyki świadomego niszczenia oryginalnych dokumentów drukowanych po ich przekształceniu do postaci cyfrowej. Taka strategia zarządzania zasobami w instytucjach pamięci wynika jednak z bardzo starannie zaplanowanej i pieczołowicie realizowanej strategii ochrony zasobu cyfrowego, bazującej na wysokiej technologicznej jakości. Uznano, że narodowy zasób cyfrowy stanowią obiekty znajdujące się w kolekcjach instytucji sektora kultury i nauki, głównie w archiwach, bibliotekach, muzeach, galeriach, uczelniach, instytutach badawczo-rozwojowych, studiach filmowych, w nieco mniejszej liczbie także w operach, teatrach i filharmoniach. Są składowane, zarządzane i udostępniane przez cyfrowe archiwa, biblioteki, repozytoria, wirtualne galerie i muzea.
>
> Z zapisów aktów prawnych o bibliotekach, muzeach oraz o narodowym zasobie archiwalnym i archiwach wyprowadzono wniosek, że do narodowego zasobu cyfrowego należy kwalifikować obiekty przedstawiające wyjątkową wartość i znaczenie dla dziedzictwa narodowego, będące źródłem informacji, o wartości historycznej, dotyczącymi szeroko pojętej działalności Państwa Polskiego, jego stosunków z innymi państwami, życia społecznego, gospodarczego i politycznego. Narodowy zasób cyfrowy powinien być zbiorem świadectw organizacji i rozwoju nauki, kultury i sztuki. Powinny tworzyć go obiekty o treści ponadprzeciętnej, świadczące o wybitnej kulturalnej i naukowej działalności narodu, dokumentujące jego dzieje. Powinien być ukonstytuowany z materiałów umożliwiających upowszechnianie podstawowych wartości historii, nauki i kultury polskiej oraz światowej, a także kształtowanie wrażliwości poznawczej i estetycznej. Narodowy zasób cyfrowy ma

również informować o treściach, wartościach, cechach unikatowych gromadzonych w nim zbiorów[2].

Przekazywana w ręce czytelników książka powstała na podstawie analizy głównie angielsko- i niemieckojęzycznych źródeł wiedzy o długoterminowej archiwizacji zasobów cyfrowych. Uwzględniono publikacje instytucji amerykańskich, angielskich, australijskich, także holenderskich i niemieckich. Szczególnie wnikliwie rozpoznano i często odwoływano się do dorobku Deutsche Nationalbibliothek [DNB, b.d.].

Treść książki została podzielona na trzy rozdziały. Rozdział pierwszy zawiera definicje pojęć podstawowych z zakresu długoterminowej archiwizacji zasobów cyfrowych. Szczegółowo opisano model OAIS (Open Archival Information System), który jest standardem ISO w zakresie organizacji i funkcjonowania archiwów cyfrowych. Scharakteryzowano również system DSEP (Deposit System for Electronic Publication), opracowany na podstawie modelu OAIS w ramach projektu NEDLIB zrealizowanego w holenderskiej bibliotece narodowej – Koninklijke Bibliotheek (KB). Ponadto przedstawiono niemiecki system depozytowy Kopal-Archivsystem. Odniesiono się również do zagadnień wiarygodności i certyfikacji archiwów cyfrowych.

W drugim rozdziale, poświęconym technicznym, ekonomicznym i prawnym zagadnieniom archiwizacji zasobów cyfrowych, omówione zostały wybrane kwestie wymagające uwzględnienia w projektowanych programach ochrony.

Podjęto próbę ustalenia, którym technicznym rozwiązaniom przyznaje się status mających szansę zapewnić dostępność i użyteczność cyfrowego materiału obecnie i w przyszłości. Opisano rekomendowane w piśmiennictwie przedmiotu metody migracji oraz emulacji; wskazano na bardzo ważny problem trwałości cyfrowych nośników danych. Dodatkowo przybliżone zostały zagadnienia identyfikatorów trwałych oraz metadanych – jako elementów niezbędnych w efektywnym zarządzaniu i użytkowaniu archiwalnych obiektów cyfrowych. Odniesiono się również do zagadnień bezpieczeństwa archiwalnych repozytoriów cyfrowych. Następnie wymienione zostały najważniejsze grupy kosztów związanych z długoterminową ochroną dokumentów cyfrowych oraz możliwe źródła finansowania programów archiwizacyjnych. Przedmiotem tego rozdziału stały się także niedostatki regulacji prawnych oraz przepisy kłopotliwe z punktu widzenia procesów

[2] Do treści książki wprowadza się definicję, opracowaną na potrzeby raportu zleconego przez Ministerstwo Kultury i Dziedzictwa Narodowego RP. Raport jest dokumentem niepublikowanym. Ma charakter roboczy i powstał dla wewnętrznych celów operacyjnych Departamentu Mecenatu Państwa Ministerstwa Kultury i Dziedzictwa Narodowego RP [Opracowanie w sprawie, 2017].

archiwizacji, uniemożliwiające bądź ograniczające określone zabiegi konserwatorskie na archiwizowanych obiektach cyfrowych. Przybliżono zagadnienie tzw. *preservation policy*, czyli zbioru dokumentów o charakterze umów, zarządzeń, rozporządzeń, postanowień, ustaw, regulujących i dostarczających podstawy dla wszelkich czynności związanych z długoterminową ochroną zasobów cyfrowych w instytucjach pamięci.

Ostatni, trzeci rozdział dotyczy zagadnień organizacyjnych ochrony zasobów cyfrowych. Jego istotny element stanowi przegląd badań dotyczących m.in. poziomu świadomości pracowników polskich instytucji pamięci, odnośnie do potrzeby podjęcia działań na rzecz ochrony polskiego dziedzictwa cyfrowego. Zauważono, że odpowiedzialność za ochronę narodowego dziedzictwa cyfrowego nie spoczywa tylko i wyłącznie na instytucjach pamięci, lecz jest odpowiedzialnością zbiorową instytucji rządowych, pozarządowych, państwowych i prywatnych, a także twórców i wydawców zasobów cyfrowych. W trzecim rozdziale uwzględniono procesy oceny i selekcji, czyli tworzenia archiwalnych kolekcji cyfrowych, oraz rolę twórców zasobów cyfrowych w procesach ich ochrony.

Publikację zamyka przedstawienie podstawowych założeń autorskiej propozycji programu długoterminowej ochrony polskich zasobów cyfrowych. Program jest połączeniem przedstawionych w tekście rozwiązań, stosowanych w wielu instytucjach pamięci na świecie, oraz autorskich sugestii dotyczących organizacji rodzimych przedsięwzięć na rzecz zachowania polskiego zasobu cyfrowego. Uwzględniono stan polskich materiałów cyfrowych, warunki polskich instytucji pamięci, poglądy ich przedstawicieli rozpoznane w toku prac badawczych podejmowanych w ostatnich dziesięciu latach.

Książka nie pretenduje do miana wyczerpującej publikacji opisującej wszystkie zagadnienia i problemy metodyczne długoterminowej archiwizacji. Przedstawiono jedynie najważniejsze kwestie natury organizacyjnej, technicznej, prawnej i ekonomicznej archiwistyki cyfrowej. Z pewnością każda z nich zasługuje na analizę i mogłaby stanowić temat odrębnej monografii. Opracowania takie są bez wątpienia bardzo potrzebne z uwagi na słabą reprezentację tego tematu w polskiej literaturze przedmiotu. Z oczywistych względów przedstawiony w książce problem nie może być uznany za ostatecznie rozstrzygnięty i zamknięty. Szybki postęp technologiczny i nowe światowe osiągnięcia w archiwistyce cyfrowej będą wymuszać bezustanną aktualizację programów i strategii długoterminowej archiwizacji zasobów cyfrowych.

1. Długoterminowa archiwizacja zasobów cyfrowych – definicje, teorie, normy, projekty

Zapewnienie przyszłym pokoleniom dostępu do cyfrowej kolekcji dorobku nauki i kultury naszych czasów wymaga opracowania programu postępowania w zakresie długoterminowej archiwizacji zasobów cyfrowych. W tym celu konieczne jest dokładne rozpoznanie istoty tego zagadnienia i jego precyzyjna charakterystyka.

1.1. Publikowanie elektroniczne i digitalizacja jako procesy powstawania dokumentów cyfrowych

Terminy *publikować* i *publikowanie* pochodzą od łacińskiego słowa *publico*, *publicare* o znaczeniu „ogłaszać, podawać do publicznej wiadomości, oddawać do publicznego użytku" [Pisarek, 2006, s. 171; Plezia, 2007, s. 376]. Publikowanie jest postrzegane jako proces ogłaszania określonych treści drukiem lub w wersji elektronicznej [Dubisz, 2003, s. 843]. Mogą być one również rozpowszechniane wśród publiczności za pomocą wykładu, wystąpienia, wystawy, audycji radiowej lub telewizyjnej [Hiller i Füssel, 2002, s. 334]. Na mocy ustawy z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych status publikacji przyznaje się dziełom zwielokrotnionym dowolną techniką w celu rozpowszechnienia, w szczególności utworom piśmienniczym, graficznym i graficzno-piśmienniczym, audiowizualnym, zapisanym na nośnikach elektronicznych oraz programom komputerowym [Ustawa, 1996]. Przełomem w procesach publikowania okazały się przekazy cyfrowe. Rozwój technologii komputerowych dostarczył doskonalsze narzędzia, sprzęt i oprogramowanie, służące usprawnieniu i przyspieszeniu procesów wydawniczych wszelkiego typu dokumentów. Przy użyciu odpowiednio wyposażonych komputerów oraz programów edytorskich i graficznych wydawcy dokonują składu dokumentów, a także ich przygotowania do druku [Brauner i in., 1997, s. 67]. Autorzy mają możliwość samodzielnego

tworzenia i rozpowszechniania dokumentów cyfrowych. Za powszechne należy uznać zjawisko publikowania oraz samopublikowania treści, także tych o charakterze naukowym, na przenośnych mediach cyfrowych i w sieciach [Adamczewski, 2005, s. 50; Czapnik, 2009, s. 15-62; Deja, 2015; Grossmann, 1997, s. 86-87; Hacker, 1992, s. 105].

Dokumenty powstające w procesach publikowania elektronicznego określa się fachowo mianem oryginalnych lub pierwotnych publikacji elektronicznych (ang. *digitally born electronic publications*), a nazewnictwo to stosuje się w celu odróżnienia publikacji elektronicznych pierwotnych od zdigitalizowanych, czyli pochodzących z procesów konwersji dokumentów analogowych do postaci cyfrowej (ang. *digitised publications*) [Steenbakkers, 2000, s. 7-8].

Digitalizacja, z ang. *digitizing, digitalization*, tłumaczona również jako *dyskretyzacj*a bądź *cyfryzacja* [Czerni i Skrzyńska, 1986, s. 122; Król, 2004, s. 25-33; Szaniawski, 1997, s. 103], jest rozumiana jako przekształcanie dokumentów analogowych do postaci cyfrowej. Proces takiego przekształcania bywa też określany jako konwersja, w której wyniku otrzymuje się cyfrowe surogaty dokumentów analogowych [Lee, 2003, s. 117; Pawska i Szymorowska, 2001]. W niniejszej książce terminy te są używane zamiennie.

Dokumenty zdigitalizowane deponowane w systemach archiwalnych dzieli się na trzy kategorie, tj. cyfrowe odwzorowania treści dokumentów, cyfrowe odwzorowania treści i cech formalnych dokumentów oraz cyfrowe kopie dokumentów [Radwański, 2005, s. 103].

W przypadku pierwszej kategorii dominującą cechą jest czytelność warstwy tekstowej; każda jakość obrazu i zastosowany format umożliwiający przeczytanie treści dokumentu są w tym przypadku wystarczające. Istotne mogą okazać się także łatwość dostępu i „przeszukiwalność" tekstu osiągana dzięki zastosowaniu takiego formatu zapisu, który umożliwia automatyczne indeksowanie pełnotekstowe. Drugą kategorię archiwizowanych obiektów stanowią cyfrowe odwzorowania dokumentów, przy których sporządzaniu oprócz treści zachowuje się także cechy formalne, na przykład układ typograficzny, ilustracje, fakturę papieru itp.; w tym przypadku zastosowanie znajdują formaty pozwalające na czytelność tychże cech. Natomiast w przypadku wykonywania cyfrowych kopii dokumentów celem jest uzyskanie takiej jakości odwzorowania, która pozwala na pełne zastąpienie oryginału, w tym reprodukcję typograficzną [Radwański, 2005, s. 103].

Biorąc pod uwagę fakt, iż istotnym celem procesu długoterminowej archiwizacji jest zachowanie autentyczności dokumentów, należałoby w trakcie pozyskiwania ich cyfrowych wersji dołożyć wszelkich starań, aby w procesie technologicznym digitalizacji odwzorować wszelkie możliwe atrybuty dokumentu źródłowego. Należy dążyć do uzyskania takiej cyfrowej kopii, która zagwarantuje użytkowni-

kom dostęp do autentycznej treści oryginalnego dokumentu, wiernie przedstawi jego cechy formalne, będzie stanowić jego reprodukcję wraz z odzwierciedleniem stanu fizycznego oryginału. Osiąga się to przez unikanie retuszowania mankamentów technicznych, takich jak np.: rysy, odciski palców, marszczenia, zagięcia papieru. Cechy oryginałów, których z różnych powodów nie udaje się odzwierciedlić w procesie konwersji, zwykle opisuje się w dokumentacji stanowiącej uzupełnienie powstających surogatów cyfrowych.

Na początku obecnego stulecia zwrócono uwagę, aby digitalizacji nie traktować jako formy długotrwałego zabezpieczania dokumentów analogowych. Uznano, że sprawdza się ona głównie jako technika ułatwiająca dostęp do zagrożonych zniszczeniem dokumentów tekstowych i wizualnych. Owszem, wiele zyskuje się poprzez przekształcenie do postaci cyfrowej, lecz zapewnienie trwałości i autentyczności dokumentów – tak istotnych z punktu widzenia procesu długoterminowej archiwizacji – nie znajdowały się wówczas po stronie korzyści [Smith, 2003, s. 108-116]. Należy jednak uwzględnić fakt, że rozwój technologiczny dostarczył nowsze, wydajniejsze narzędzia. Liczne projekty wniosły wiedzę i doświadczenie w obszarze zarówno digitalizacji, jak i trwałej ochrony materiału cyfrowego. W opiniach polskich ekspertów[1] digitalizacja jest właściwą formą długotrwałego zabezpieczania dokumentów analogowych, podobną do starszej technologii zabezpieczania treści na mikrofilmach, lecz nowoczesną. Z racji złego stanu zachowania wiele dokumentów wydrukowanych na kwaśnym papierze wymaga zachowania właśnie metodami cyfrowymi. Z doświadczeń rozmówców wynika, że dokumenty analogowe i niektóre dzieła (typu grafika, fotografie, płaskie formy ulotne) znikają; proces ten jest pewny i nieodwracalny, a jedyną metodą zachowania przekazu jest poprawnie wykonana digitalizacja. Zwrócono uwagę, że sztuki wizualne i formy przekazu ulegają ciągłemu rozwojowi. Gazety analogowe, na przykład, odchodzą w zapomnienie. Świat zmienia się i nie ma już – wg zebranych opinii – innej drogi, jak tylko cyfrowa. Digitalizacja może zabezpieczyć dokument oraz nadać nową funkcjonalność analogowym obiektom. W jednej z wypowiedzi digitalizację uznano za najlepszą metodę ochrony zasobów analogowych, ale i początek drogi zabezpieczenia materiału cyfrowego. W odniesieniu do tematu autentyczności dokumentów, eksperci przyznali, że nie ma, nie było i chyba nigdy nie będzie metody wytworzenia w pełni autentycznej kopii. Zawsze dokument zdigitalizowany będzie

[1] W 2017 r. wypowiedzieli się na temat wybranych zagadnień długoterminowej archiwizacji zasobów cyfrowych eksperci z Biblioteki Uniwersytetu Wrocławskiego, Jagiellońskiej Biblioteki Cyfrowej, Małopolskiej Biblioteki Cyfrowej, Muzeum Narodowego w Krakowie, Śląskiej Biblioteki Cyfrowej. Zgromadzone opinie zostały wkomponowane w treść książki, w różnych jej częściach.

surogatem obiektu analogowego, którego poziom autentyczności zależy od użytej techniki. Na obecnym poziomie techniki digitalizacja nie pozwala na odwzorowanie rzeczywistego wyglądu oryginału. Jeden z rozmówców zauważył, że żywotność informacji jest coraz krótsza. W świecie informatycznym dane są produkowane na potrzeby czasu liczonego w sekundach. Autentyczność potrzebuje nowej definicji i być może jest to definicja przydatności.

Zgromadzone wypowiedzi polskich specjalistów wskazują jednoznacznie, że jest obecnie w środowisku instytucji pamięci świadomość, że digitalizacja to forma długoterminowej ochrony zbiorów analogowych, jednak z określonymi obostrzeniami. W środowisku tym występuje zrozumienie potrzeb i nowych zadań, które są rezultatem digitalizacji. Przyznano w jednej z rozmów, że nieodpowiednia jakość procesu cyfryzacji oraz brak zabezpieczeń materiału cyfrowego, głównie nośników, prowadzą do poważnych strat, oraz że problem jest znany i typowy; „kto w porę nie zareagował, ten nic nie ma". W omawianym przypadku udało się skopiować w porę większość starych płyt i dysków, ale odnotowano również straty – w obrębie nośników. Oryginały poddano powtórnej digitalizacji.

O tym, że digitalizacja jest obecnie pojmowana szerzej niż na początku XXI w. i nie stanowi już tylko wyodrębnionego procesu przekształcenia formy analogowej do cyfrowej, lecz jest zadaniem kompleksowym, świadczy wypowiedź, której autorzy określają terminem digitalizacja proces ucyfrowienia obrazu fizycznego obiektu na potrzeby przetwarzania, udostępniania i archiwizacji za pomocą urządzeń teleinformatycznych. Digitalizację rozumieją jako zbiór wszystkich procesów, które towarzyszą ucyfrowieniu dokumentów. Są to czynności, które przyczyniają się do stworzenia dokumentacji cyfrowej obiektu począwszy od typowania, prac konserwatorskich, po opracowanie metadanych, obróbkę cyfrową, publikację w sieci oraz archiwizację przygotowanych danych cyfrowych. W rozważaniach o digitalizacji uwzględnia się archiwizację jako element składowy procesu i definiuje ją jako „ogół metod i praktyk służących zabezpieczaniu i długoterminowemu przechowywaniu wytwarzanych zasobów cyfrowych. Na archiwizację składają się: infrastruktura techniczna, oprogramowanie, sprzęt, interfejsy użytkowników, mechanizmy wymiany danych, mechanizmy zabezpieczania i migracji danych, obsługa techniczna" [Kalota i Szala, 2012, s. 437-446]. Digitalizację podejmuje się w różnych celach. Najczęściej wymieniane, to ochrona oraz udostępnianie. W zakresie ochrony zwraca się uwagę na: (1) zachowanie zawartości zbiorów, tj. trwałe zachowanie cyfrowego odwzorowania obiektu należącego do dziedzictwa narodowego na wypadek destrukcji oryginału oraz (2) ochronę zbiorów, tj. ochronę oryginału przed nadmierną eksploatacją, zniszczeniem, kradzieżą i wszelkimi skutkami niekorzystnych działań środowiska zewnętrznego. Zamiast oryginału udostępnia się jego kopię cyfrową [Paradowski, 2010, s. 27].

Istotne jest rozumienie, że postrzeganie dyskretyzacji jako panaceum na wszelkie bolączki związane z długoterminowym przechowywaniem i udostępnianiem dokumentów może doprowadzić do ich nieodwracalnej utraty. Dyskretyzacja była przyjmowana na całym świecie (w wielu krajach nadal jest) z entuzjazmem; pracowników instytucji pamięci przekonuje się do podejmowania działań na rzecz przekształcania na postać cyfrową tekstów i materiałów wizualnych, tak jakby to przetworzenie było samo w sobie oczywistym dobrem. Tymczasem należy ich jednocześnie uświadamiać, że jedną z najistotniejszych cech dokumentów cyfrowych jest to, że z natury nie są one na stałe utrwalone na nośniku i tak trwałe jak tekst drukowany na papierze. Nie są też w sposób pewny utrwalone pod względem treści i formy, gdyż można je łatwo zmienić, nie pozostawiając śladów modyfikacji, z jednym wyjątkiem – wydruku. Elastyczność dokumentów cyfrowych jest uznawana za zaletę tylko przez ich twórców, natomiast bibliotekarzom, archiwistom, muzealnikom usiłującym zgromadzić i zachować dokument skończony i ostateczny przynosi znaczące skomplikowanie działań. Z racji tej elastyczności i podatności na zmiany zabezpieczenie dokumentu cyfrowego i jego przechowanie w długim okresie czasu jest zadaniem trudnym. Nierozstrzygnięte są pytania: którą cyfrową wersję dokumentu lub jak wiele wersji należy zarchiwizować i jaką techniczną procedurę zastosować, by zapewnić trwałość i stabilny dostęp? Kto powinien zająć się planowaniem i organizacją projektów długoterminowego przechowania? Skąd wreszcie czerpać fundusze na ich realizację?

Nie udało się dotychczas jednoznacznie odpowiedzieć na powyższe pytania, pomimo że długoterminowa archiwizacja zasobów cyfrowych jest od ponad dwudziestu lat przedmiotem żywych dyskusji w kręgach bibliotekarzy, bibliotekoznawców, informatologów, archiwistów, muzealników i informatyków z niemal całego świata, tematem licznych opracowań, hasłem przewodnim wielu konferencji i sympozjów, a dokumenty cyfrowe poddawane są różnorodnym praktycznym eksperymentom. Dotychczas nie opracowano zadowalającej strategii postępowania z dokumentami cyfrowymi, gwarantującej ich stabilną, długoterminową użyteczność.

1.2. Długoterminowa archiwizacja zasobów cyfrowych

Termin *długoterminowa archiwizacja* (ang. *long term archiving*, niem. *langfristige Archivierung*, *Langzeitarchivierung*) identyfikowany jest przede wszystkim z ochroną, zabezpieczaniem i przechowywaniem dokumentów cyfrowych (ang. *long term protection, long term preserving*, niem. *langfristige Erhaltung, Langzeiterhaltung, dauerhafte Sicherung*), w celu zagwarantowania ich długoterminowej użyteczności. Słowo *długoterminowy* należy rozumieć jako nieograniczony

w czasie lub możliwie najbardziej odległy w przyszłości. Dodatkowego wyjaśnienia znaczenia słowa *długoterminowy* dostarczają autorzy wypowiedzi na temat systemów archiwalnych, projektowanych i implementowanych w bibliotekach oraz archiwach narodowych. Zadaniem takich systemów ma być długoterminowa, tj. trwająca ponad sto lat, archiwizacja zbiorów cyfrowych [Borghoff, 2005]. W wypowiedziach pojawia się również koncepcja, według której archiwizacja długoterminowa oznacza pełnienie zadań ochrony zbiorów cyfrowych przez określony zespół ludzi do czasu, kiedy zadania te przejmie zespół następczy, zdolny do ich kontynuacji [Neuroth i in., 2009].

Do polskojęzycznej terminologii przedmiotu wprowadzono wyrażenie *ochrona informacji cyfrowych* i zdefiniowano je jako „zbiór rozwiązań służących zapewnieniu ciągłości dostępu do materiałów należących do dziedzictwa cyfrowego w okresie, w którym ciągłość taka jest pożądana" [National Library of Australia, 2003, s. 30, 44]. Przy czym dostępność materiału cyfrowego jest utożsamiana z zachowaniem środków, które będą w stanie zapewnić dostęp do autentycznej treści dokumentu i umożliwią jego użytkowanie zgodnie z pierwotnym celem. W dalszej części definicji zwraca się uwagę, że termin ten nie oznacza wykorzystania technik obrazowania cyfrowego w celu stworzenia cyfrowych kopii źródeł niecyfrowych. Jest to potwierdzenie prezentowanej wcześniej opinii, podającej w wątpliwość zasadność utożsamiania procesów digitalizacji z procesami archiwizacji dokumentów. Digitalizację bowiem można traktować jako metodę archiwizacji zbiorów analogowych wówczas, gdy jej naturalnym przedłużeniem jest stosowna dbałość o surogaty cyfrowe. Samo skanowanie i zgromadzenie materiału cyfrowego w jednym miejscu to znacznie za mało, by mówić o archiwizacji, zwłaszcza długoterminowej.

Użyteczność obiektów archiwalnych oznacza przede wszystkim ich dostępność oraz możliwość korzystania z nich, głównie czytania, oglądania zapisanych w nich treści oraz słuchania przez osoby upoważnione, w ramach posiadanych przez nie praw dostępu [Bilski, 2008, s. 423-425; Kriterienkatalog, 2006]. Użyteczność wiąże się z efektywnym korzystaniem z dokumentów cyfrowych, co staje się możliwe wówczas, gdy użytkownik ma pewność, że treści, które odbiera, a także na które powołuje się w swych opracowaniach, są autentyczne i niezafałszowane, tj.: pochodzą od ich autorów i od dnia opublikowania nie uległy zmianie, przedstawiają dokładnie to, co było zamierzeniem ich twórców [Coy, 2006; Kriterienkatalog, 2006]. Obok zagwarantowania dostępności treści obiektów za podstawowy cel procesu archiwizacji należy uznać zapewnienie ich autentyczności i integralności [Attributes, 2001]. Integralność obiektów archiwalnych wiąże się przede wszystkim z ich kompletnością [Kriterienkatalog, 2006]. System archiwalny powinien być odporny na wszelkiego rodzaju nieupoważnione i niewłaściwe modyfikacje

obiektów archiwalnych. Z technicznego punktu widzenia pojęcie modyfikacji jest opisywane jako zmiana wartości danych – wstawienie bądź ich usunięcie [Bilski, 2008, s. 423-452]. Integralność może zostać naruszona przez zarówno celowe działanie nieuprawnionego użytkownika, jak i błędy oraz zaniedbania uprawnionego użytkownika, a także wirusy komputerowe lub inne programy szkodliwe, awarie sprzętu, kanałów komunikacyjnych, zasilania czy błędy oprogramowania [Bilski, 2008, s. 424]. Z pojęciem integralności związane jest pojęcie niezmienności danych, czyli brak możliwości dokonania jakichkolwiek zmian, tak nieautoryzowanych, jak i autoryzowanych [Bilski, 2008, s. 424]. Wśród celów, którym ma służyć długoterminowa archiwizacja, wymienia się również ochronę poufności. Pod pojęciem poufności należy rozumieć stan, w którym dokument nie jest i nie może być ujawniony osobom nieupoważnionym. Zapewnienie poufności dokumentu może wynikać z takich przesłanek, jak: ochrona prywatności i interesów własnych deponenta, ochrona interesów instytucji archiwizującej (czyli depozytariusza), obowiązujące akty prawne [Bilski, 2008, s. 424; Kriterienkatalog, 2006].

Warunkiem koniecznym dla wszelkich czynności zapewniających dostępność i użyteczność obiektów cyfrowych jest utrzymanie ich substancji. Substancja to ciąg bitów (kod zero-jedynkowy) zapisany na medium elektronicznym. Utrzymanie substancji jest uzależnione od dwóch zasadniczych czynników, tj. od ograniczonej trwałości nośników zapisu oraz szybkich zmian zachodzących w dziedzinie formatów zapisu danych. Pomimo że wymienione czynniki znacznie utrudniają proces archiwizacji, nie można pozwolić, aby były one powodem utraty dostępu i możliwości korzystania z dokumentów cyfrowych. Istotne znaczenie dla właściwego przebiegu archiwizacji ma przestrzeganie określonych zasad postępowania z dokumentami zapisanymi w formie cyfrowej [Liegmann, 2001, s. 100-105].

Trwałość medium jest zwykle dłuższa niż dostępność sprzętu i oprogramowania potrzebnych do odczytu zapisanych na nim danych. Dlatego też niezbędne jest stałe obserwowanie zmian zachodzących w technologii i odpowiednio wczesne reagowanie na te zmiany. W anglojęzycznym piśmiennictwie przedmiotu proces ten nazwano *technology watch*, zaś w opracowaniach pochodzących z obszaru niemieckojęzycznego został on określony jako *Frühwarnsystem*. Odpowiednik w języku polskim to *system wczesnego ostrzegania*.

Podstawowe zasady systemu ostrzegania wynikają z instrukcji dotyczących trwałości mediów cyfrowych, na których są zapisane ważne treści, oraz kontrolowania w ustalonych odstępach czasu, czy zapis cyfrowy na poziomie kodu zero-jedynkowego może być odczytany. Względy bezpieczeństwa nakazują, aby przed upływem granicy trwałości nośnika przekopiować zapisane na nim dane na nowy nośnik tego samego typu, np. z dysku na dysk – zabieg taki określany jest jako odświeżenie nośnika (ang. *refreshing,* niem. *Wiederauffrischen*). Natomiast

w przypadku, gdy nośnik przestaje być powszechnie stosowany i zastępuje go nowa generacja, treść dokumentu należy przekopiować na nośnik nowej generacji, np. z płyty CD na płytę DVD; zabieg ten można nazwać zmianą generacji nośnika (ang. *reformating,* niem. *Wechsel der Trägergeneration*). Obecnie najczęściej stosuje się metodę polegającą na odłączeniu treści dokumentu cyfrowego od oryginalnego nośnika zapisu, a następnie umieszcza się ją (w celu ochrony) w archiwalnym repozytorium cyfrowym, magazynie danych cyfrowych, systemie depozytowym.

Samo zabezpieczenie substancji dokumentu cyfrowego zapewniłoby użytkownikom jedynie dostęp do kodu zero-jedynkowego. Potrzebne są zatem odpowiedni sprzęt i oprogramowanie umożliwiające odczytanie treści zakodowanej w postaci zer i jedynek. W historii tworzenia dokumentów cyfrowych znajdowały zastosowanie różne platformy programowo-sprzętowe, toteż w instytucjach pamięci zgromadzono dokumenty cyfrowe, których odczyt z nośnika i odszyfrowanie za pomocą aktualnego sprzętu oraz oprogramowania jest obecnie utrudnione, często niemożliwe. Przykładem mogą być publikacje zapisane na powszechnych w latach 90. XX w. dyskietkach 5,25 czy 3,5 cala. Ich odczyt jest możliwy jedynie przy użyciu stacji dysków, które wyszły z użycia. Jeśli uda się zdobyć odpowiednią stację dysków, to kolejną barierą może stać się dostępność platformy programowo-sprzętowej niezbędnej do zdekodowania treści publikacji. Pomocne w takiej sytuacji okazuje się „zachowanie technologii" [Czermiński, 2002, s. 93]. Przechowywanie sprzętu i oprogramowania, które wyszły z powszechnego użycia, i są wykorzystywane jedynie w celu odczytywania treści zapisanych w formatach specyficznych dla tych platform, jest traktowane jako jedna z możliwych metod długoterminowej archiwizacji zasobów cyfrowych. Jednak tworzenie tzw. muzeów komputerowych nie spotyka się z uznaniem specjalistów. Wypowiadając się na temat istotnych metod długoterminowej archiwizacji, wymieniają migrowanie treści dokumentów i emulowanie systemów [Borghoff i in., 2003].

Obok bezpośredniej dbałości o obiekty archiwalne konieczne jest zapewnienie bezpieczeństwa miejsca, w którym są one przechowywane. Niezbędne są określone regulacje organizacyjne, techniczne oraz prawne. Najistotniejsze kwestie, które powinny zostać rozstrzygnięte to, m.in.: uprawnienia dotyczące dostępu do archiwum, uprawnienia dotyczące przeprowadzania określonych prac na obiektach cyfrowych (przez przygotowanych i upoważnionych pracowników), mechanizmy zabezpieczeń dostępu do archiwum oraz do obiektów archiwalnych (m.in. systemy identyfikacyjne, hasła). Strategia ochrony archiwum i zasobów archiwalnych powinna uwzględniać także zagadnienia dotyczące kopii zapasowych chronionych dokumentów. Ze względów bezpieczeństwa i ochrony przed ewentualnymi katastrofami, należy wykonać kopie zapasowe wszystkich archiwizowanych obiek-

tów i przechowywać je w miejscu terytorialnie oddalonym od instytucji, w której znajduje się archiwum główne.

Pozyskanie i zapis danych cyfrowych to etap, który w dużej mierze decyduje o powodzeniu dalszego przebiegu procesu archiwizacji. Przy zapisie danych cyfrowych konieczne jest rozpatrzenie możliwości ich odczytu nie tylko za miesiąc, rok, ale także za lat pięćdziesiąt, sto i więcej [Rohde-Enslin, 2004, s. 5-6]. Zalecane jest odejście od idei, że zapis materiału w postaci cyfrowej oznacza jego długoterminową dostępność i użyteczność. Niezbędna jest świadomość, ile zależnych od siebie elementów wchodzi w skład otoczenia użytkowania dokumentu cyfrowego. Są to: format zapisu danych, nośnik danych, stacja dysków, napęd, system operacyjny. Każda różnica pomiędzy programowo-sprzętowym otoczeniem[2] powstania i zapisu a otoczeniem odczytu i interpretacji danych cyfrowych zmniejsza prawdopodobieństwo zagwarantowania użyteczności dokumentów cyfrowych w przyszłości. Przed każdą konieczną i uzasadnioną zmianą platformy programowo-sprzętowej lub tylko wybranych jej elementów należy upewnić się, że nowa platforma umożliwi odczyt oraz interpretację danych cyfrowych zapisanych w starszym otoczeniu. Natomiast po zmianie otoczenia wymagana jest kontrola, czy rzeczywiście dane z otoczenia starszej generacji są czytelne w nowym otoczeniu. Przy wszelkich zmianach należy zadbać o kompatybilność nowych elementów sprzętu i oprogramowania z elementami starszymi, tym samym zapewnić dostępność i użyteczność dokumentów cyfrowych. Przestrzega się przed największym zagrożeniem, jakim jest utrata dostępu do danych cyfrowych. Innymi słowy, materiały cyfrowe nie zostały zarchiwizowane, jeśli zostały utracone niezbędne środki dostępu do ich treści [National Library of Australia, 2003, s. 44].

Zasygnalizowane aspekty techniczne stanowią bardzo ważne, lecz nie jedyne elementy procesu długoterminowej archiwizacji zasobów cyfrowych. Przeprowadzanie jakichkolwiek prac na obiektach cyfrowych, w celu zapewnienia ich nieustającej użyteczności, wymaga wpierw ich zgromadzenia. Proces gromadzenia dokumentów cyfrowych, podobnie jak dokumentów analogowych, wiąże się ze stworzeniem stosownych przepisów prawnych, na mocy których instytucje archiwizujące będą uprawnione do otrzymywania dokumentów cyfrowych. Poważnym wyzwaniem jest rejestracja i zachowanie dokumentów sieciowych. Opracowania wymagają procedury oceny i selekcji, które jednoznacznie określałyby cechy dokumentów stanowiących naukowe oraz kulturowe dziedzictwo cyfrowe i tym

[2] Pod pojęciem *otoczenie dokumentu cyfrowego* rozumie się w tej publikacji platformę sprzętowo-programową, w której dokumenty cyfrowe powstają, są zapisywane oraz odczytywane. Zamiennie stosowane są terminy: środowisko i otoczenie dokumentu cyfrowego (cyfrowego obiektu archiwalnego).

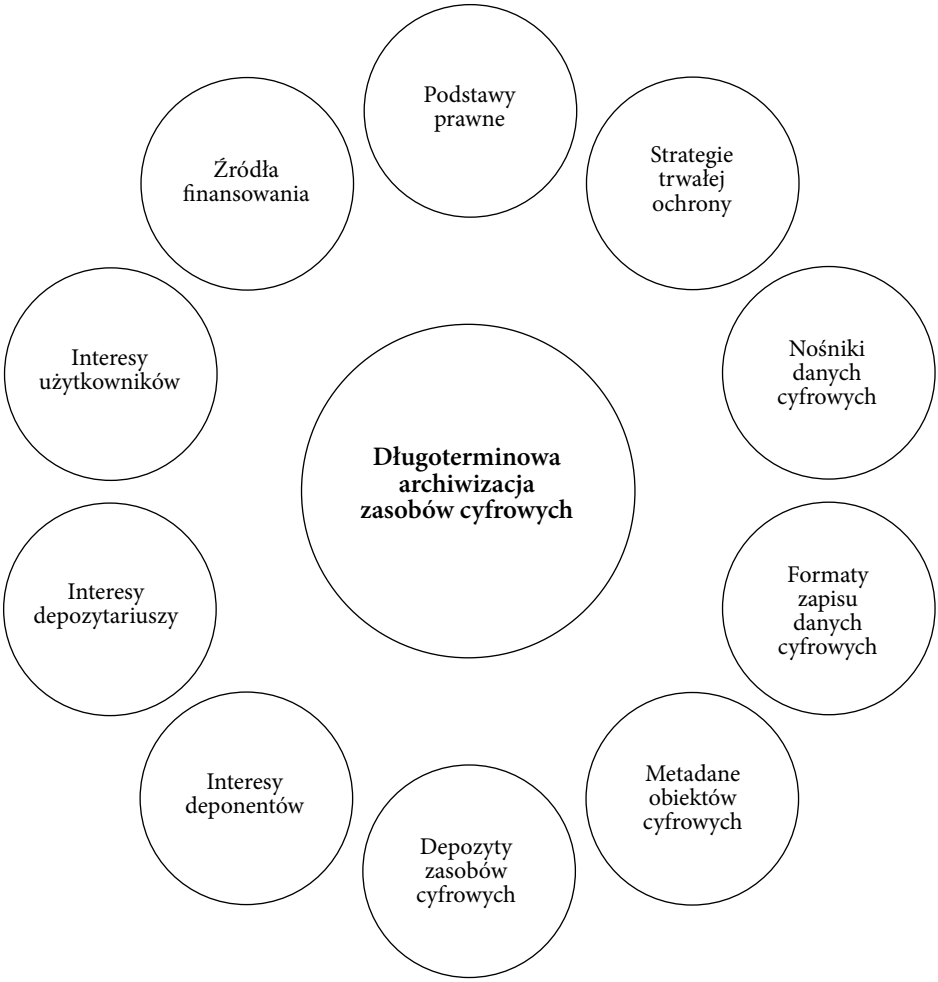
samym zasługujących na włączenie do archiwalnej kolekcji cyfrowej. Procedury selekcyjne powinny uwzględniać nie tylko wartość merytoryczną dokumentów, lecz również oceniać ich parametry techniczne. Gwarancją długoterminowej ochrony użyteczności dokumentów cyfrowych jest ich zapis w otwartych, standardowych formatach danych. Inaczej mówiąc, konieczne jest ustalenie technicznych wytycznych tworzenia i zapisu dokumentów cyfrowych oraz ich konsekwentne stosowanie.

Realizacja wszelkich zadań związanych z długoterminową archiwizacją wiąże się ze zbudowaniem stosownych struktur organizacyjnych, stworzeniem podstaw prawnych i przede wszystkim ustaleniem źródeł finansowania. Powołanie zespołu specjalistów, stworzenie im odpowiednich warunków pracy, a przede wszystkim zaprojektowanie, budowa, implementacja i bieżące utrzymanie systemu archiwalnego łączy się ze znaczącymi nakładami finansowymi. Celem nadrzędnym staje się więc ustalenie źródła stałego finansowania prac na rzecz długoterminowej ochrony dziedzictwa cyfrowego.

W konsekwencji przedstawionych dotychczas rozważań pojęcie *długoterminowa archiwizacja zasobów cyfrowych* należy rozumieć jako proces trwałego, czyli nieograniczonego w czasie, zachowania dostępności dokumentów cyfrowych wraz z ochroną ich użyteczności, czyli autentyczności, integralności oraz poufności. Proces ten polega na zachowaniu niezmienionej substancji dokumentu w postaci kodu zero-jedynkowego oraz na zapewnieniu platformy programowo-sprzętowej, która będzie w stanie zdekodować dane cyfrowe i przedstawić je w postaci czytelnej dla użytkownika. Obok zabiegów technicznych na proces archiwizacji wpływ ma szereg ustaleń oraz czynności natury organizacyjnej, niezbędnych dla utworzenia stosownej infrastruktury (złożonej z ludzi, miejsc i narzędzi ich pracy), przyjęcie niezbędnych aktów prawnych (aby prace w zakresie długoterminowej archiwizacji zasobów cyfrowych nie kolidowały z obowiązującym ustawodawstwem) oraz czynności natury ekonomicznej, w ramach których powstawałyby plany finansowania projektów archiwizacji, pozyskiwałoby się potrzebne fundusze oraz przeprowadzało kontrolę dotyczącą gospodarowania nimi. Istotne jest postrzeganie długoterminowej archiwizacji jako procesu nieustającego, którego powodzenie zależy od poprawności opracowanego planu działania, jego wdrożenia oraz konsekwentnego realizowania w długim czasie. W fazie projektowania takiego procesu uwzględnienia wymaga szereg aspektów (Rys. 1) [Preserving, 1996; Suchodoletz, 2008, s. 6].

Długoterminowa archiwizacja zasobów cyfrowych została niejednokrotnie określona mianem wyzwania [Börsenverein, 1996, passim]. Pomimo wielu lat intensywnej pracy rozmaitych instytucji pamięci na całym świecie oraz poważnych nakładów finansowych, nie udało się stworzyć uniwersalnej, właściwej strategii ochrony zasobów cyfrowych. Nadal szereg aspektów archiwistyki, dotyczącej

Rysunek 1. **Komponenty problematyki długoterminowej archiwizacji zasobów cyfrowych**



Źródło: oprac. własne na podstawie: Suchodoletz, von Dirk (2008); Preserving, 1996

trwałej ochrony danych cyfrowych, postrzega się jako zadanie bardzo trudne, złożone, czaso-, praco- i kosztochłonne. Dodatkową trudność rodzi towarzysząca ochronie zasobów cyfrowych presja czasu, związana z tempem zmian technologicznych. Postępujący rozwój w dziedzinie sprzętu i oprogramowania komputerowego powoduje, że zadania długoterminowej archiwizacji należy traktować jako kategorię pilnych i stałych czynności instytucji pamięci.

Szczególna trudność ochrony wynika z ogromnej ilości heterogenicznych zasobów. Mnogość i różnorodność stosowanych formatów zapisu danych, nośników,

a także występowanie wielu wartościowych treści w postaci publikacji sieciowych, które nie mają surogatów w wersji analogowej, powoduje, iż w dziedzinie archiwistyki cyfrowej pojawiają się nowe, poważne dylematy.

Problematyka długoterminowej ochrony zasobów cyfrowych jest rozpatrywana w kategoriach wyzwania także z racji swej złożoności. Właściwie każdy projekt archiwizacji materiałów cyfrowych wymaga dogłębnego poznania tematu, umiejętności projektowania, dobrej organizacji, wsparcia specjalistów branży IT, podstawy prawnej i najważniejszego zapewne, stabilnego źródła finansowania.

Spektakularne i zadowalające, choć nie rozwiązujące ostatecznie problemu trwałej ochrony zasobów cyfrowych, są efekty współpracy kilkunastu instytucji i organizacji, specjalizujących się w tematyce długoterminowej archiwizacji dokumentów cyfrowych. Największe z nich to:

- DCC – Digital Curation Centre (Wielka Brytania),
- DPC – Digital Preservation Coalition (Wielka Brytania),
- LDP – Centre for long-term digital preservation (Szwecja),
- Fondazione Digitale Rinascimento – Nuove Tecnologie per i Beni Culturali (Włochy),
- NCDD – Nationale Coalitie Digitale Duurzaamheid (Holandia),
- NDIIPP – National Digital Information Infrastructure and Preservation Program (USA),
- Österreichisches Wissensnetzwerk Digitale Langzeitarchivierung (Austria),
- PIN – Groupe Pérennisation des Informations Numériques (Francja),
- European Alliance for Permanent Access (Europa).

Na uwagę zasługują również osiągnięcia National Digital Stewardship Alliance (NDSA), Open Preservation Foundation, UK Web Archive.

1.3. Dokumenty cyfrowe – typologia i charakterystyka w kontekście archiwizacji długoterminowej

Określenia *dokumenty/publikacje elektroniczne* oraz *dokumenty/publikacje cyfrowe* są stosowane synonimicznie. Jest to jednak praktyka budząca wątpliwości, bowiem nie wszystkie dokumenty elektroniczne są zapisane w postaci kodu numerycznego, charakterystycznego dla zapisów cyfrowych, a i nie wszystkie zapisy cyfrowe są jednocześnie elektronicznymi [Czermiński, 2002, s. 169-175]. Wprowadza się rozróżnienie pomiędzy dokumentami elektronicznymi, dokumentami cyfrowymi oraz elektronicznymi dokumentami cyfrowymi. Przyporządkowując dokumenty do określonego typu, należy skupić się na formie ich zapisu i odczytu oraz stosowanych do tego urządzeń. Przymiotnik *elektroniczne* powinien być stoso-

wany w przypadku publikacji, do których zapisu i odczytu używane są urządzenia elektroniczne, przy czym forma zapisu (analogowa bądź cyfrowa) nie ma w tym wypadku znaczenia. Przez dokumenty cyfrowe natomiast należy rozumieć te, które zostały zapisane w postaci kodu numerycznego (jego szczególnym przypadkiem jest kod binarny), przy czym technologia zapisu (elektroniczna bądź mechaniczna) jest nieistotna. Z terminologicznego punktu widzenia elektroniczne dokumenty cyfrowe są przypadkiem szczególnym. Powinny spełniać wymagania stawiane zarówno dokumentom elektronicznym, jak i dokumentom cyfrowym. Większość wypowiedzi, w których synonimicznie stosuje się terminy *dokumenty elektroniczne* oraz *dokumenty cyfrowe*, dotyczy *elektronicznych dokumentów cyfrowych*. Ich przykładem może być plik tekstowy (zredagowany w wybranym edytorze), do którego odczytu konieczne jest posłużenie się urządzeniem elektronicznym [Czermiński, 2002, s. 169-175]. To właśnie te elektroniczne dokumenty cyfrowe stanowią obiekty archiwalne i podlegają procesom trwałej ochrony w archiwach cyfrowych.

Typ dokumentów określany jako *dokumenty/publikacje elektroniczne* bądź *dokumenty/publikacje cyfrowe* definiowany jest w piśmiennictwie przedmiotu rozmaicie. W Polskiej Normie dotyczącej opisu bibliograficznego dokumentu elektronicznego przyjęto, iż są to dokumenty istniejące w postaci elektronicznej, dostępne za pomocą techniki komputerowej [PN, 2000]. Podobną definicję przyjęto w polskim tłumaczeniu dla potrzeb normy PN-N-01152-13:2000, dotyczącej międzynarodowego standardu opracowania bibliograficznego dokumentów elektronicznych ISBD(ER). *Dokument elektroniczny* jest w nim definiowany jako zakodowany dokument (dane i/lub program), który można odczytać przy użyciu komputera. Zalicza się tu dokumenty, które wymagają użycia urządzeń peryferyjnych podłączonych do komputera (np. czytnika CD-ROM), ale także usługi dostępne online (np. listy dyskusyjne lub strony WWW) [Grygrowski, 2001, s. 59; ISBD(ER), 1997]. W rodzimym ustawodawstwie *dokumenty elektroniczne* są opisywane jako zbiory danych, stanowiące odrębną całość znaczeniową, uporządkowane w określonej strukturze wewnętrznej i zapisane na informatycznym nośniku danych [Ustawa, 2005]. Mogą być także udostępniane na pamięciach taśmowych, dyskach magnetycznych, optycznych, w sieci lub w innej elektronicznej formie [Feather i Sturges, 2003, s. 126-132; Prytherch, 1996, s. 225]. W lapidarnym ujęciu słownikowym, dokument elektroniczny jest przechowywany na nośniku czytelnym maszynowo, dostępnym zdalnie lub lokalnie [Czapnik i Gruszka, 2011, s. 68]. W terminologii technicznej dokumenty elektroniczne to ciągi bitów, także kody zero-jedynkowe zapisane na mediach elektronicznych; nazywane dokumentami bądź obiektami elektronicznymi, digitalnymi, także cyfrowymi [Borghoff i in., 2003, s. 11; Schanze, 2002, s. 64-65].

W zbiorach instytucji przechowujących dorobek nauki i kultury występują rozmaite typy dokumentów cyfrowych. Kolekcje cyfrowe mogą obejmować treści

zarówno zapisane na nośnikach przenośnych, jak i udostępniane w sieci, publikacje łączące różne sposoby dostępu do materiałów, np. dokumenty zapisane na nośnikach przenośnych, zawierające dodatkowo linki do witryn internetowych. Mogą to być publikacje z różnych gatunków wydawniczych, tj.: monografie, serie wydawnicze, czasopisma i magazyny internetowe, także publikacje stanowiące skończone całości oraz ulegające stopniowym zmianom w wyniku ich modyfikacji przez autorów korzystających z interaktywnego potencjału internetu. Na dziedzictwo cyfrowe składają się również publikacje różnego typu wydawców oraz miejsc wydawniczych[3], także osób i instytucji, które nie są wydawcami, jednak rozpowszechniają w sieci treści mogące stanowić dziedzictwo kultury i nauki. Do dziedzictwa cyfrowego należą materiały ogólnie dostępne oraz o ograniczonym zasięgu, takie jak: preprinty czy prace naukowe użytkowane w ograniczonym zakresie przez określone środowiska. Warta zachowania może być również instytucjonalna oraz osobista dokumentacja działalności, transakcji, zbiory korespondencji, zapisy poczty elektronicznej, wypowiedzi z list dyskusyjnych, dzienników internetowych, zapisy obrazu i dźwięku z kamer internetowych (np. materiał cyfrowy z coraz popularniejszych naukowych tele- lub wideokonferencji). Instytucje pamięci przechowują również: programy i gry komputerowe, oprogramowanie narzędziowe, produkty z branży filmowej, muzycznej, radiowej, cyfrowo wygenerowane dzieła sztuki, zdjęcia dokumentalne, cyfrowe reprodukcje niecyfrowych oryginałów [National Library of Australia, 2003, s. 39-40].

Heterogeniczność dokumentów cyfrowych sprawia, że zadaniem niełatwym jest ich klasyfikacja. Jednak ze względów praktycznych specjaliści zajmujący się gromadzeniem i archiwizacją zbiorów cyfrowych proponują ich pogrupowanie w następujące klasy [Bide, 2000, s. 5-6; Steenbakkers, 2000, s. 7-8]:

1.3.1. Dokumenty/publikacje cyfrowe przenośne, hybrydowe i sieciowe

Dokumenty cyfrowe offline, określane także jako *dokumenty przenośne* (ang. *hand-held electronic publications*), występują najczęściej na przenośnych mediach, takich jak: dyskietka, CD lub DVD. Odczytywanie zapisanych w nich treści odbywa się najczęściej za pomocą komputera z odpowiednim czytnikiem oraz programem,

[3] Pod pojęciem *miejsce wydawnicze* rozumie się w niniejszej książce instytucję, bądź osobę, która nie ma statutu oficyny wydawniczej, ale publikuje dokumenty cyfrowe, będące wynikiem jej działalności. Miejsce wydawnicze będą stanowić również *twórcy*, czyli autorzy publikacji elektronicznych, publikujący samodzielnie w Internecie, np. autorzy popularnych dzienników internetowych; w przypadku niektórych z nich, wartość dla nauki i edukacji jest niekwestionowana. Warte trwałej ochrony mogą być również treści publikowane na platformach zdalnego nauczania, prywatne kolekcje cyfrowych fotografii, relacje wideo ważnych wydarzeń.

który potrafi dekodować i interpretować dane cyfrowe. Cechą charakterystyczną dokumentów offline jest możliwość korzystania z nich bez połączenia z siecią. W zbiorach instytucji archiwizujących, szczególnie w bibliotekach, istnieje wiele dokumentów typu offline, których treść może być odczytana wyłącznie za pośrednictwem właściwych dla producenta formatów, programów i urządzeń. Przykładem takich dokumentów są pierwsze generacje książek elektronicznych (ang. *e-books*). Nierozerwalne powiązanie treści dokumentu z charakterystycznym dla producenta formatem, nośnikiem danych oraz programem i urządzeniem udostępniającymi treść, znacznie utrudnia, często uniemożliwia długoterminową użyteczność publikacji, bowiem – jak już wspomniano – odczyt zapisanych treści zależy od trwałości nośnika danych oraz urządzenia dekodującego dane i prezentującego je w formie zrozumiałej dla użytkownika. Niestety większości bibliotek nie stać na zakup krótkoterminowo przydatnych programów i urządzeń, zaś wydawcy, choć świadomie integrują treść publikacji ze specyficznymi mediami, oprogramowaniem i urządzeniem, z różnych powodów nie dostrzegają potrzeby przekazywania bibliotekom publikacji w formie gwarantującej jej długoterminowe użytkowanie. W celu uniknięcia takich sytuacji konieczny jest dialog pomiędzy producentami publikacji elektronicznych i specjalistami odpowiedzialnymi za ich długoterminową archiwizację. Instytucje archiwizujące powinny dokładnie określić parametry dokumentów cyfrowych, które umożliwią ich długotrwałą użyteczność. Natomiast firmy wydawnicze i indywidualni producenci powinni, w trosce o zapewnienie długoterminowej użyteczności własnych produktów, wyjść naprzeciw możliwościom instytucji archiwizujących. Istnieje więc potrzeba określenia i konsekwentnego stosowania standardów tworzenia dokumentów cyfrowych. Zaleca się używanie uniwersalnych formatów zapisu danych oraz nośników umożliwiających późniejsze migrowanie treści dokumentów do nowego otoczenia i ich odczyt przy pomocy powszechnie dostępnego oprogramowania i sprzętu [Hägele, 2000, s. 17-21]. Nieliczne instytucje archiwizujące organizują szkolenia w zakresie tworzenia dokumentów cyfrowych – kierowane głównie do wydawców i indywidualnych twórców – mające na celu popularyzację takich metod generowania dokumentów, które usprawniłyby przebieg procesów archiwizacji oraz wpłynęły na ich efektywność.

Obok dokumentów offline instytucje pamięci włączają do swoich zasobów *dokumenty online*, określane jako *sieciowe* (ang. *network publications*). Dokumenty online są umieszczane na serwerach bibliotek, archiwów, muzeów, galerii, wydawnictw, a następnie udostępniane przez Internet. Mogą być również prezentowane tylko w intranetach czy sieciach lokalnych. Do najbardziej powszechnych dokumentów online należą czasopisma elektroniczne (ang. *e-journals*), magazyny (zazwyczaj tematyczne) publikowane w Internecie (ang. *e-ziny*) oraz kolekcje rozpraw naukowych (ang. *e-thesis*).

W obrębie dokumentów elektronicznych online rozróżniono typy statyczne, kumulatywne oraz dynamiczne [Bide, 2000, s. 5-6; Steenbakkers, 2000, s. 7-8]. Cechą charakterystyczną dokumentów statycznych jest stabilność ich substancji – innymi słowy, raz ustalona forma i treść nie ulega zmianom od momentu utworzenia przez cały cykl ich istnienia. Ustalona treść i struktura dokumentu nie powinna też zmieniać się podczas użytkowania. Przykładem statycznych dokumentów są monografie opublikowane w formie elektronicznej. Z kolei dokumenty kumulatywne to takie, których zawartość treściowa zmienia się, rozrasta w trakcie funkcjonowania dokumentu w przestrzeni publicznej, przy czym substancja elementu dodanego jest stabilna. Typowym przykładem zasobów cyfrowych kumulatywnych są dzienniki internetowe (blogi lub weblogi). Dokumenty dynamiczne natomiast cechują się tym, że ich forma oraz treść ulegają ciągłym zmianom. Treść dokumentu cyfrowego i forma jego prezentacji są dynamiczne, ustalane w trakcie ich użytkowania (ang. *on the fly*, „w locie"). Typowym przykładem dynamicznych dokumentów są obrazy z kamer internetowych, należą do nich także systemy bazodanowe, prezentujące generowane „w locie" rezultaty zapytań wyszukiwawczych (jednak to nie rezultat zapytania, lecz system bazodanowy jest obiektem archiwizacji). Z czasem na znaczeniu zyskały również obiekty interaktywne posadowione w środowisku Internetu, np. platformy e-learningowe albo instalacje wirtualnych muzeów i galerii sztuki. Są to przykłady kolejnych typów cyfrowych zasobów nauki i kultury, wartych zachowania na przyszłość, ale równocześnie stanowiących kolejne obszary do zagospodarowania przez archiwistów cyfrowych.

Archiwizacja dokumentów typu online wymaga w pierwszej kolejności ich zgromadzenia i zapisania na serwerze instytucji archiwizującej. W tym celu niezbędne jest ustalenie wymogów, które musi spełnić dokument, by został przyjęty i trwale archiwizowany. Podstawowym kryterium powinna być treść dokumentu, w szczególności jej znaczenie dla nauki i kultury. Ustalenia wymaga również używany format zapisu danych oraz sposób przekazania dokumentu z miejsca wydania do instytucji archiwizującej. Najczęściej dokumenty typu online przesyłane są przez Internet, jednak w szczególnych przypadkach zdarza się ich dostarczenie na elektronicznym medium przenośnym w celu umieszczenia na serwerze instytucji archiwizującej i udostępniania w trybie online.

W celu archiwizacji dokumentów cyfrowych typu online konieczne jest stworzenie odpowiednich przepisów prawnych, regulujących proces ich gromadzenia oraz udostępniania. Na mocy obecnie obowiązujących ustaw instytucje pamięci gromadzą i przechowują publikacje wydane na przenośnych mediach elektronicznych. Jednak w związku z rosnącą popularnością tworzenia sieciowych kolekcji różnego typu dokumentów, w szczególności kolekcji rozpraw naukowych, w niektórych krajach znowelizowano ustawy o bibliotekach narodowych. Głównym

celem zmian jest włączenie dokumentów online do dóbr narodowych oraz nałożenie na wydawców obowiązku ich zgłaszania i transferu do biblioteki narodowej w celu długoterminowej archiwizacji, natomiast obowiązkiem bibliotek narodowych staje się ich gromadzenie, udostępnianie i długoterminowa archiwizacja [DNBG, 2009].

Za typ pośredni – pomiędzy dokumentami przenośnymi oraz sieciowymi – uznawane są *hybrydowe dokumenty elektroniczne* (ang. *hybrid electronic publications*). Zasadniczo są one zaliczane do grupy dokumentów przenośnych, jednak ich cechą charakterystyczną, odróżniającą je od tradycyjnych dokumentów przenośnych, są zawarte w nich linki do materiałów dostępnych w Internecie.

Dokumenty hybrydowe stanowią szczególnie kłopotliwy typ obiektów archiwalnych, gdyż w celu zapewnienia ich użyteczności należałoby w procesie archiwizacji zatroszczyć się nie tylko o zachowanie danych cyfrowych oraz narzędzi odczytu i prezentacji dokumentu właściwego, lecz także o dostępność i czytelność powiązanych z nim dokumentów sieciowych. Optymalnym rozwiązaniem byłaby równoczesna archiwizacja właściwego dokumentu offline oraz wszelkich powiązanych z nim dokumentów online. Realizacja takiego zadania, choć niełatwa, wydaje się możliwa z racji postępów prac w dziedzinie rejestracji i archiwizacji zasobów sieciowych (ang. *web harvesting*) [Day, 2003; Dąbrowska, 2017; Derfert-Wolf, 2012; Konopa, 2017].

Jeszcze innym typem zasobów cyfrowych wartych długoterminowego zachowania są strony WWW. Publikacje umieszczane bezpośrednio na stronach WWW oraz „podpinane" do nich (na zasadzie linków) zawierają nierzadko warte zachowania treści (ang. *Web-based publications*). Jednak zanim biblioteki lub inne instytucje archiwizujące zdecydują się na rejestrację i długoterminową archiwizację stron WWW, musiałyby poradzić sobie z problemem określenia cech dokumentów mających stać się obiektem procesu archiwizacji. Poważną trudność stanowi także ustalenie granic stron internetowych, które miałyby być odrębnymi obiektami podlegającymi archiwizacji. W tym celu konieczne jest stworzenie mechanizmów jednoznacznie rozpoznających zawartość pojedynczego dokumentu oraz rozróżniających linki wewnętrzne (należące do danej strony WWW) od linków zewnętrznych (odsyłających do dokumentów z innych stron).

Dokumentacja i zachowanie tzw. dziedzictwa internetowego, czyli opublikowanych w Internecie dokumentów cyfrowych, wydaje się zadaniem szczególnie pilnym. Po pierwsze, z uwagi na ulotność zasobów WWW – zmiany w jednej części cyberprzestrzeni mogą zdekompletować odwołujący się do niej zbiór materiałów – po wtóre, z racji dużego popytu na nowość. Wydawcy bowiem, aby wyjść naprzeciw oczekiwaniom użytkowników sieci, stale zmieniają „repertuar" udostępnianych materiałów, nie zawsze planując ich długoterminowe przechowanie.

Dodatkowo globalny dostęp do sieci i hakerskie zdolności niektórych użytkowników są w stanie doprowadzić do utraty wielu istotnych dokumentów lub ich zmian, a ponowne pozyskanie lub odtworzenie pierwotnej wersji nie zawsze jest możliwe [National Library of Australia, 2003, s. 42].

1.3.2. Publikacje profesjonalnych wydawców oraz małych, efemerycznych firm wydawniczych

Z uwagi na fakt, że w większości krajów gromadzenie i archiwizacja publikacji elektronicznych, zwłaszcza publikacji sieciowych, bazuje na dobrowolnych umowach pomiędzy bibliotekami narodowymi i wydawnictwami, bardzo duże znaczenie ma nawiązywanie kontaktów i współpraca z wydawcami. Doświadczenia zaawansowanych w tym zakresie instytucji bibliotecznych pokazują, że dużo łatwiej nawiązać kontakt i efektywnie współpracować z profesjonalnymi wydawnictwami, które są obecne na rynku wydawniczym przez długi czas i którym zależy na długoterminowej dostępności oraz użyteczności własnych produktów. Zdecydowanie trudniej o współpracę i dotarcie do oferty wydawców działających spontanicznie i krótkoterminowo.

W krajach, w których biblioteki narodowe traktują długoterminową archiwizację publikacji elektronicznych jako rozszerzenie zakresu swoich dotychczasowych zadań, współpraca taka jest inicjowana głównie przez biblioteki. Starania bibliotekarzy koncentrują się głównie na informowaniu wydawców o istniejącej możliwości zgłaszania i przesyłania publikacji online do biblioteki narodowej w celu ich długoterminowej archiwizacji oraz na stworzeniu infrastruktury dla celów zgłoszenia i transferu dokumentów cyfrowych [Bide, 2000, s. 5-6; Steenbakkers, 2000, s. 7-8].

1.3.3. Dokumenty cyfrowe digitalne i zdigitalizowane

Dokumenty digitalne, nazywane też oryginalnymi dokumentami elektronicznymi bądź cyfrowymi, powstają w pełni skomputeryzowanych procesach. Mogą występować jako dodatki towarzyszące dokumentom tradycyjnym, wydrukowanym na papierze albo stanowić samodzielny dokument. W przypadku gdy dokument digitalny nie posiada swojego odpowiednika w wersji drukowanej, proces jego zabezpieczenia i długoterminowej archiwizacji nabiera szczególnego znaczenia.

Dokumenty zdigitalizowane, sporadycznie nazywane „dygitalizatami", są efektem procesów digitalizacji zbiorów analogowych, mających na celu m.in. zabezpieczenie przed zniszczeniem oraz poprawienie dostępu do rzadkich i wartościowych obiektów analogowych. Jakość zabezpieczenia dokumentów zdigitalizowanych

zależy od zastosowanej techniki ucyfrowienia. Fakt, że w wielu krajach dyskusja na temat standaryzacji i unifikacji stosowanych metod wciąż pozostaje otwarta, prowadzi do stosowania niejednolitych praktyk.

W wymienionych podziałach wskazuje się na mnogość typów dokumentów cyfrowych przechowywanych w instytucjach archiwizujących. Jednocześnie zaznacza się możliwość pojawienia się z czasem nowych rodzajów dokumentów, które będą stanowić dziedzictwo cyfrowe i powinny zostać zarchiwizowane. Tym samym trudno o definitywne ustalenia w odniesieniu do kategoryzacji zasobów cyfrowych oraz zakresu metod ich archiwizacji. Biorąc pod uwagę tempo rozwoju technologicznego, instytucje archiwizujące powinny raczej nastawić się na stałe obserwowanie zachodzących zmian i względnie szybkie dostosowywanie do nich swych zasobów merytorycznych i infrastrukturalnych [Bide, 2000, s. 5-6; Steenbakkers, 2000, s. 7-8].

1.4. Archiwum cyfrowe

W celu uniknięcia terminologicznej niejednoznaczności rozważania na temat organizacji, funkcjonowania i atrybutów archiwów cyfrowych warto rozpocząć od zwrócenia uwagi na wieloznaczność nazw stosowanych w odniesieniu do nich, szczególnie w piśmiennictwie polskim. Dla zapewnienia jasności wywodu konieczne jest rozróżnienie zakresu znaczeniowego nierzadko zamiennie stosowanych wyrażeń *biblioteka cyfrowa*, *archiwum cyfrowe* oraz *repozytorium cyfrowe*.

1.4.1. Rozróżnienie pojęć: biblioteka cyfrowa, archiwum cyfrowe, repozytorium cyfrowe

Z definicji encyklopedycznych wynika, że pod pojęciem *biblioteka cyfrowa* (*biblioteka elektroniczna*) należy rozumieć instytucję bądź organizację, w której przeważającą część zgromadzonych zasobów stanowią dokumenty cyfrowe (elektroniczne), udostępniane za pomocą urządzeń komputerowych na miejscu (w bibliotece) bądź poza biblioteką, poprzez sieci komputerowe [Reitz, 2004, s. 217; Strauch i Rehm, 2007, s. 140]. Pojęcie to bywa także wyjaśniane jako kolekcja dzieł w formacie cyfrowym dostępna użytkownikom lub grupom użytkowników [Feather i Sturges, 2003, s. 651]. Dostępne przeglądy definicyjne pozwalają wnioskować, że biblioteka cyfrowa to skomplikowany system gromadzenia, wyszukiwania, przechowywania i zarządzania danymi, informacjami, treściami [Janiak, 2012, s. 15-65]. Wśród zadań biblioteki cyfrowej wymieniane są: gromadzenie, opracowywanie i udostępnianie zbiorów cyfrowych, zgodnie z regułami stosowanymi

w bibliotekarstwie. Ewentualnie do zadań tych można również włączyć ochronę zbiorów, tj.: ochronę nośników elektronicznych i zapisanych na nich danych, tworzenie i przechowywanie kopii zapasowych. Ochroną zbiorów mogą jednak zajmować się wyspecjalizowane instytucje [Bojar, 2002, s. 34]. W piśmiennictwie przedmiotu, głównie polskojęzycznym, biblioteki cyfrowe bywają określane również jako *repozytoria cyfrowe*. Praktyka ta jednak powinna budzić wątpliwości, bowiem *repozytorium cyfrowe* jest – zgodnie ze źródłami leksykograficznymi – synonimem terminu *archiwum cyfrowe* [Reitz, 2004, s. 216]. Określenie *repozytorium* najczęściej stosowane jest w odniesieniu do repozytoriów instytucjonalnych, w których gromadzi się, przechowuje i na bieżąco udostępnia w otwartym modelu publikowania dorobek pracowników naukowo-dydaktycznych uczelni lub wyniki prac naukowo-badawczych prowadzonych w instytutach badawczo-rozwojowych. Zwraca się uwagę na problemy terminologiczne dotyczące tego pojęcia oraz relacje między repozytorium a biblioteką cyfrową. Brakuje wyraźnego podziału na biblioteki cyfrowe i repozytoria. Rozstrzygnięcia w tym zakresie wymagałyby dokładnej analizy wszystkich funkcji przez nie pełnionych wraz z określeniem przewagi jednych nad innymi [Sapa, 2013, s. 117-131].

Stosowanie terminu *repozytorium cyfrowe* w odniesieniu do bibliotek cyfrowych może wynikać z postrzegania bibliotek cyfrowych jako baz danych, magazynów czy systemów przechowywania dokumentów. Według różnych źródeł słowo *repozytorium* oznacza właśnie bazę danych, bazę dokumentów, system przechowywania informacji [Clavel-Merrin, 2000; Freedman, 2004, s. 684; Przyłuska, 2008]. W niektórych źródłach dodaje się jednak, iż *repozytorium* to system przechowywania cyfrowych dokumentów głównie w celu ich przyszłego, późniejszego udostępniania (ang. *stored for subsequent access*) [Clavel-Merrin, 2000, s. 14]. Realizuje się w nim funkcja typowa dla archiwów.

Pod pojęciem *archiwum cyfrowe* oraz *archiwum elektroniczne* należy rozumieć organizację ludzi oraz narzędzi (lub system złożony z osób oraz przyjętych rozwiązań organizacyjnych i technicznych) powołaną w celu zgromadzenia, przechowania oraz zapewnienia długoterminowego dostępu i użyteczności cyfrowego materiału [Clavel-Merrin, 2000, s. 6; Kriterienkatalog, 2006; Reitz, 2004, s. 216]. Działania archiwum koncentrują się na pracach związanych z przeprowadzeniem cyfrowych dokumentów przez kolejne etapy rozwoju technologicznego, przy użyciu najróżniejszych narzędzi i metod archiwizacji, m.in. migracji oraz emulacji. Docelowo archiwum ma zapewnić obecnym oraz przyszłym użytkownikom możliwość odczytu i interpretacji autentycznych, integralnych, wiarygodnych dokumentów cyfrowych [Neuroth i in., 2009]. W wypowiedziach na temat archiwów cyfrowych autorzy często odwołują się do standardu archiwizacji dokumentów cyfrowych OAIS [Januszko-Szakiel, 2005, s. 341-358; OAIS, 2002], w którym – oprócz wymienionych

cech – uwzględnia się dążenie archiwum cyfrowego do stałej obserwacji i zabezpieczenia zmieniających się potrzeb docelowej grupy użytkowników nazywanych niekiedy klientami bądź odbiorcami usług archiwum. W przywoływanej dokumentacji OAIS archiwum elektroniczne jest również określane synonimicznie terminem *repozytorium cyfrowe* [OAIS, 2002; Research Libraries Group i OCLC, 2002].

Reasumując, terminem *repozytorium cyfrowe* określane są przede wszystkim systemy publikowania prac naukowych w modelu open access, ale także archiwa oraz biblioteki cyfrowe. W celu eliminacji nieporozumień terminologicznych wymagane jest dookreślenie, czy zastosowanie terminu *repozytorium* odnosi się do cyfrowych repozytoriów archiwalnych, czy do cyfrowych repozytoriów instytucjonalnych i bibliotecznych. Wymienione repozytoria różnią się założeniami oraz celami, które realizują. Repozytoria archiwalne służą głównie długoterminowej i bezpiecznej ochronie użyteczności zasobów cyfrowych; ich funkcja związana z bieżącym udostępnianiem gromadzonych i archiwizowanych materiałów jest drugorzędna bądź marginalna. Natomiast działania cyfrowych repozytoriów instytucjonalnych i bibliotecznych są skoncentrowane przede wszystkim na organizacji bieżącego dostępu do dokumentów cyfrowych. Coraz częściej jednak przy cyfrowych bibliotekach i repozytoriach instytucjonalnych tworzony jest tzw. moduł archiwalny, którego celem jest właśnie długoterminowe zabezpieczenie użyteczności szczególnie ważnych materiałów [CRL, 2010].

Terminy *archiwum cyfrowe* i *repozytorium cyfrowe* będą odnosić się w niniejszej publikacji zawsze do systemów deponowania i długoterminowego archiwizowania zasobów cyfrowych.

1.4.2. Archiwa cyfrowe – typologia

Spektrum funkcjonujących oraz wciąż powstających archiwów cyfrowych jest bardzo szerokie. W jednym z opracowań przedmiotu proponuje się następującą ich typologię [Kriterienkatalog, 2006, s. 3]:

- cyfrowe archiwa bibliotek narodowych, realizujące zadania gromadzenia oraz zabezpieczenia cyfrowych produktów wydawniczych, naukowych zasobów sieciowych, także wyników projektów digitalizacyjnych; docelową grupą użytkowników jest ogół społeczeństwa;
- cyfrowe archiwa bibliotek uczelnianych gromadzące i zarządzające cyfrowymi publikacjami, głównie wydawnictw naukowych, ale dodatkowo kolekcjonujące media dla procesów zdalnego nauczania, cyfrowe wersje rozpraw naukowych, a także rozmaite opracowania pracowników dydaktycznych i naukowych w formie preprintów; docelową grupę użytkowników takich archiwów stanowią głównie studenci oraz pracownicy uczelni;

- cyfrowe archiwa centrów i instytutów badawczych działające w celu gromadzenia i zabezpieczenia danych powstających w wyniku ich badawczej działalności; klientami archiwów tego typu są zazwyczaj specjaliści dziedzinowi, którzy łączą swą wiedzę z danymi wynikowymi instytucji badawczych, interpretują je i ewentualnie dostarczają nowe wnioski i opracowania przedmiotu;
- cyfrowe archiwa sektora administracji, zarządzania, biznesu, powstające na mocy przepisów obligujących instytucje do stosownego zarządzania oraz przechowania przez określony czas elektronicznych dokumentów powstających i przydatnych w toku ich działalności; w zależności od typu dokumentu, docelową grupą użytkowników może być ogół społeczeństwa bądź pracownicy samej instytucji dla potrzeb wykonywanych przez nich zadań. Od archiwów instytucji pamięci, sektora nauki i kultury odróżnia je okres przechowywania zbiorów. Istotne jest nie zachowanie długoterminowe (jak najodleglejsze w przyszłości), lecz dostępność dokumentów podyktowana rozmaitymi przepisami prawnymi (niekiedy prawo nakazuje zniszczenie dokumentu, ogranicza bądź zabrania jego użytkowania);
- cyfrowe archiwa instytucji archiwalnych i muzealnych, w których przechowuje się i zarządza cyfrowymi obiektami archiwalnymi i muzealnymi oraz „dygitalizatami" obiektów analogowych. Mogą być użytkowane przez całe społeczeństwo, głównie jednak przeznaczone są dla osób zawodowo związanych ze światem kultury, sztuki i nauki;
- cyfrowe archiwa organizowane przez zewnętrznych usługodawców, przyjmujące zlecenia archiwizacji zasobów cyfrowych rozmaitych instytucji, zarówno tych z sektora biznesu oraz administracji, jak i nauki oraz kultury; odpowiedzialność za gromadzenie i przekazanie zasobów do archiwum ponoszą instytucje zlecające, natomiast usługodawcy przyjmują obowiązek zabezpieczenia ich dostępności i użyteczności w określonym czasie.

Każde z wymienionych typów archiwów (repozytoriów) cyfrowych można ogólnie określić jako zbiór dokumentów cyfrowych zgromadzonych i przechowywanych w określonym miejscu, przez określony czas i dla określonych celów. Dyskusyjne jednak mogłoby stać się stosowanie jednakowego określenia w odniesieniu do archiwów cyfrowych materiałów bibliotecznych i archiwalnych oraz – na przykład – archiwum cyfrowych dokumentów bankowych, podatkowych czy administracyjnych. Należy zwrócić uwagę na odmienność wyobrażeń o systemach archiwalnych instytucji pamięci narodowej i systemach sektora biznesu lub administracji. Zasadnicza różnica pomiędzy powyższymi systemami tkwi w założeniach dotyczących okresu przechowywania dokumentów. Archiwalne systemy biznesowe, w zależności od rodzaju przechowywanych dokumentów, realizowanych zadań oraz procedur prawnych, mają zapewnić dostępność i użyteczność deponowanych materiałów przez trzy do pięćdziesięciu lat [Konstankiewicz, 2005, s. 49-62; Konstankiewicz,

2006, s. 53-60; Sasin, 2004, passim], natomiast systemy archiwalne (depozytowe) instytucji pamięci powinny gwarantować utrzymanie użyteczności zbiorów przez okres stu i więcej lat [Borghoff, 2005]. Wiąże się to z innymi założeniami organizacyjnymi oraz rozwiązaniami technicznymi funkcjonowania systemów. Ponadto elektroniczne systemy archiwalne w ujęciu biznesowym operują dokumentami prymarnymi, tj. stworzonymi od początku jako dokumenty cyfrowe; natomiast systemy archiwalno-biblioteczne dotyczą dodatkowo dokumentów wtórnych, tj. cyfrowych surogatów dokumentów papierowych. Skutkuje to innymi oczekiwaniami i nakłada na system archiwalny dodatkowe powinności. Archiwum elektroniczne jako zbiór dokumentów cyfrowych w sensie archiwalno-bibliotecznym musi sprostać zasadniczym problemom, czyli: długotrwałemu przechowywaniu dokumentów cyfrowych, zarządzaniu wielką liczbą dokumentów cyfrowych oraz umożliwić ich sprawne indeksowanie, wyszukiwanie i udostępnianie użytkownikom [Radwański, 2005, s. 101].

Istotę funkcjonowania systemów archiwalnych (depozytowych), głównie tych organizowanych w bibliotekach i archiwach narodowych, przybliżają wypowiedzi specjalistów. Wynika z nich, że działanie systemów archiwalnych wymaga dwóch zasadniczych elementów składowych: serwera, na którym umieszczony jest informatyczny system realizujący zadania bieżącej obsługi archiwum (lub biblioteki) i użytkowników, oraz serwera depozytowego (stanowiącego jądro archiwum), którego zadaniem jest długoterminowa archiwizacja materiału cyfrowego [LoC, 2003]. Rozróżnienie to ma ogromne znaczenie z uwagi na brak zgodności co do tego, czy instytucje archiwizujące powinny posiadać dwa odrębne zasoby, z których jeden byłby kompletnym zbiorem dokumentów cyfrowych, natomiast drugi tworzyłby wyselekcjonowaną kolekcję dokumentów stanowiących dziedzictwo nauki i kultury, deponowanych długoterminowo z myślą – tylko i wyłącznie – o przyszłych użytkownikach (model rozłączny zasobów wewnątrz instytucji archiwizującej), czy też wszystkie instytucjonalne zasoby cyfrowe powinny być przechowywane długoterminowo w systemie depozytowym i stąd pobierane również przez system informatyczny w celu bieżącego udostępniania użytkownikom (model zasobów połączonych wewnątrz instytucji archiwizującej).

Argumentem przemawiającym za rozłącznością zasobów depozytowego oraz bieżącego jest ich odmienny charakter. Zasoby archiwizowane w systemie depozytowym powinny być – w myśl niektórych opinii – starannie dobraną, reprezentatywną kolekcją dóbr nauki i kultury, nikłe są bowiem możliwości zachowania dla przyszłych użytkowników wszystkich dokumentów cyfrowych. Wydaje się wysoce prawdopodobne, że nawet centralne biblioteki, archiwa i muzea narodowe, ustawowo zobowiązane do zachowania kompletnego dziedzictwa nauki i kultury, będą zmuszone do wyboru tych dokumentów, które z racji swej treści i formy zasługują na archiwizację długoterminową. Jednocześnie instytucje te muszą brać pod uwagę

fakt, iż obecni użytkownicy mają prawo do bieżącego korzystania z wszelkich dostępnych materiałów, bez względu na to, czy stanowią one dziedzictwo narodowe, czy nie. Stąd pomysł, aby serwery biblioteczne, archiwalne i muzealne gromadziły i udostępniały na bieżąco wszystkie zgromadzone dokumenty z myślą o obecnych użytkownikach, natomiast serwery depozytowe przejęły zadania długoterminowej archiwizacji kolekcji wyselekcjonowanej, utworzonej z dokumentów stanowiących dziedzictwo nauki i kultury, zabezpieczając potrzeby przyszłych użytkowników.

Kolejnym punktem polemiki dotyczącej organizacji i funkcjonowania systemów depozytowych w instytucjach archiwizujących jest częstotliwość użytkowania dokumentów. Niektórzy znawcy tematu twierdzą, że użytkowanie zasobów, zwłaszcza tych z systemu depozytowego, i to jak najczęstsze, jest bardzo pożądane. Wraz z realizacją procesów użytkowania wzrasta prawdopodobieństwo wykrycia ewentualnych utrudnień odczytu i prezentacji treści dokumentów cyfrowych, tym samym minimalizuje się ryzyko bezpowrotnej utraty dokumentu, np. z racji niezauważonych zmian technologicznych i wywołanych przez nie braków w otoczeniu sprzętowym lub programowym. Bieżące udostępnianie zbiorów cyfrowych jest – ich zdaniem – sposobem obserwacji zmian technologicznych, tym samym zwiększeniem szans na podjęcie odpowiednich działań we właściwym czasie. W bieżącym udostępnianiu archiwizowanych zbiorów widzą oni strategię długoterminowego utrzymania ich użyteczności. Takie podejście wywodzi się prawdopodobnie z doświadczeń uzyskanych podczas funkcjonowania komputerowych systemów zarządzających archiwaliami tradycyjnymi, w których moduł Konserwacja umożliwia przejście do pliku zawierającego wykaz jednostek archiwalnych, które powinny zostać poddane zabiegom konserwatorskim. Jednostki wymagające konserwacji są dostrzegane właśnie w procesie udostępniania archiwaliów. Wykrycie dokumentu wymagającego prac konserwatorskich wiąże się z poczynieniem w module Konserwacja adnotacji o jego stanie [Pest, 2007, s. 15-23].

Z kolei zwolennicy rozdzielnego modelu zasobów archiwum elektronicznego bronią przekonania, iż bieżące udostępnianie zbiorów z depozytu może okazać się zgubne w skutkach. W procesach ich częstego użytkowania i w wyniku ewentualnych niepożądanych działań użytkowników zbiory są narażone na ryzyko naruszenia ich integralności i autentyczności. Dodatkowo udostępnianie i użytkowanie dokumentów z depozytu może utrudniać prace charakterystyczne dla procesu długoterminowej archiwizacji. Dlatego też uważa się, że archiwalne zbiory depozytowe należałoby oddzielić od zbiorów użytku bieżącego, udostępniać je wyłącznie osobom upoważnionym do prac konserwatorskich, przewidzianych w strategii ich długoterminowej archiwizacji.

Chociaż w obu przedstawionych podejściach występują elementy racjonalnego postępowania ze zbiorami cyfrowymi, bardziej przekonujący wydaje się pogląd

o bieżącym użytkowaniu archiwizowanych zasobów. Przy obecnych możliwościach technicznych oraz starannie przemyślanej taktyce działania specjaliści są w stanie ochronić autentyczność i integralność materiału cyfrowego. Kolejnym argumentem przemawiającym przeciwko rozdzielności zasobów jest stały wzrost liczby nowopowstających dokumentów i idącej za tym ich łącznej objętości. Podwajanie nawet najmniejszych objętościowo zasobów wiąże się z dodatkowymi kosztami ich utrzymania oraz zwiększonymi wymaganiami technicznymi stawianymi przed systemami opartymi na zdublowanych zbiorach.

Abstrahując od wymienionych wcześniej kwestii, prawdopodobne wydaje się, że zarówno model rozłączenia zasobów, jak i model zasobów połączonych przyniósłby pożądany efekt końcowy. Oba modele archiwizacji są w stanie zapewnić dostępność i użyteczność cyfrowego materiału w długim czasie. Jednak aby miały szansę sprawdzić się w praktyce, musiałyby zostać wbudowane w sprecyzowany, kompleksowy program długoterminowej archiwizacji zbiorów cyfrowych. Pomimo wysiłków wielu instytucji na świecie, takie programy należą wciąż do rzadkości, a jeśli nawet powstają, nie są popularyzowane i rekomendowane, ponieważ zwykle są rozwiązaniami tymczasowymi, modyfikowanymi, dostosowywanymi do zmieniających się potrzeb i okoliczności, a przede wszystkim są to rozwiązania niesprawdzone. Ich efektywność można zweryfikować i ocenić jedynie w dłuższym czasie i w obliczu zachodzących zmian technologicznych.

Z dyskusji prowadzonej w piśmiennictwie przedmiotu wynika, że istnieje duże zapotrzebowanie na wszelkie hipotetyczne scenariusze organizacji i działania systemów archiwizacji zasobów cyfrowych. Należy jednak koncentrować się na tworzeniu rozwiązań kompleksowych, wielozadaniowych, elastycznych, umożliwiających ewentualną reorganizację wstępnych założeń. Trzeba również uwzględniać istniejące standardy.

1.4.3. Open Archival Information System (OAIS) – standard w zakresie archiwizacji zasobów cyfrowych

Przedmiotem rozważań o modelu referencyjnym OAIS są jego główne aspekty, a mianowicie struktura informacji deponowanych w OAIS, otoczenie archiwum OAIS oraz schemat funkcjonowania archiwum OAIS.

Open Archival Information System – definicja

Open Archival Information System (OAIS) to referencyjny model organizacji i przebiegu procesu długoterminowej archiwizacji obiektów cyfrowych. OAIS jest definiowany jako organizacja składająca się z osób i infrastruktury, których

starania skoncentrowane są na długoterminowym przechowywaniu, zabezpieczaniu i udostępnianiu obiektów cyfrowych wyznaczonej grupie użytkowników (ang. *designated community*). *Open* oznacza, że proponowany standard organizacji archiwów elektronicznych został opracowany na forum otwartym, przy współudziale wielu specjalistów; nie implikuje jednak otwartości w sensie nieograniczonego dostępu do archiwum oraz do treści przechowywanych w nim dokumentów cyfrowych [OAIS, 2002].

Model referencyjny OAIS został stworzony przez Consultative Committee for Space Data Systems (CCSDS)[4] na potrzeby archiwizacji i wymiany danych elektronicznych zawierających informacje z badań przestrzeni kosmicznej. W maju 1999 r. została zaprezentowana pierwsza wersja modelu OAIS, a w lutym 2003 r., po licznych poprawkach, model OAIS został zaakceptowany przez International Organization for Standardization ISO jako norma postępowania w zakresie długoterminowej archiwizacji danych cyfrowych (ISO 14721:2003). Pomimo że model OAIS został stworzony głównie z myślą o archiwizacji jednego typu danych cyfrowych, jest on uznawany za uniwersalny model organizowania i funkcjonowania archiwów cyfrowych i stosowany do gromadzenia, przechowywania oraz udostępniania różnych typów dokumentów cyfrowych [Research Libraries Group i OCLC, 2002]. Wykorzystuje się go w wielu światowych bibliotekach, archiwach i muzeach, w których realizowane są projekty długoterminowej archiwizacji zbiorów cyfrowych.

Struktura informacji w modelu referencyjnym OAIS

W archiwach zgodnych z modelem OAIS istotne jest rozróżnienie pomiędzy danymi cyfrowymi (ang. *Data Object*) a obiektami informacyjnymi (ang. *Information Object*). Na obiekt informacyjny, oprócz danych cyfrowych, składają się także narzędzia konieczne do przetworzenia surowych danych do postaci zrozumiałej dla użytkownika. Narzędzia muszą pozwalać na przedstawienie treści w postaci dostosowanej do poziomu wiedzy i umiejętności użytkownika. W modelu OAIS wiedza i umiejętności są określone terminem *Knowledge Base*, a narzędzia używane do przetwarzania danych nazywane są *Representation Information* [OAIS, 2002].

Jako przykład autorzy modelu opisują sytuację, w której korzystanie z tekstów angielskojęzycznych uzależnione jest od znajomości języka angielskiego. Jeśli odbiorca tekstu nie posiada w swoich „zasobach" (ang. *Knowledge Base*) umie-

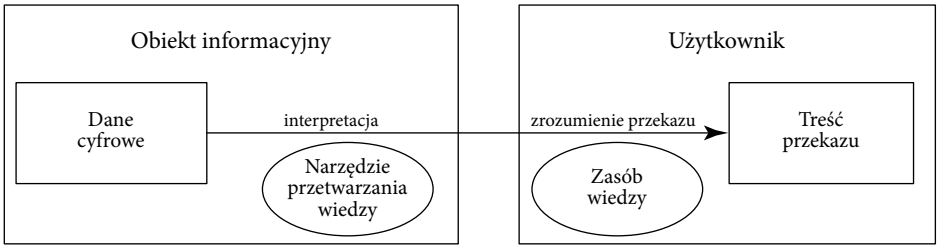
[4] Komitet CCSDS został powołany w 1982 r. Jest organizacją składającą się z przedstawicieli wielu światowych agencji badań przestrzeni kosmicznej i podlega bezpośrednio agencji NASA. *The Consultative Committee for Space Data System* [online]. Reston VA, USA. CCSDS/AIAA, 2010. Dostępny w WWW: http://www.ccsds.org/ [Dostęp: 10.07.2017].

jętności posługiwania się językiem angielskim, to wówczas, aby przetłumaczyć angielski tekst – niezbędne będzie użycie słownika i podręcznika gramatyki, które w tym przypadku stanowią narzędzia do przetwarzania danych (ang. *Representation Information*). Na tej samej zasadzie odbywa się korzystanie z dokumentów cyfrowych. Dane zapisane w postaci kodu zero-jedynkowego (ang. *Data Object*), aby były użyteczne, muszą zostać przedstawione w formie czytelnej dla użytkownika, za pomocą odpowiednich narzędzi w postaci sprzętu oraz oprogramowania (ang. *Representation Information*). Otrzymana w ten sposób treść przekazu powinna być zrozumiała dla użytkownika, przy założeniu określonego zasobu jego wiedzy i umiejętności (ang. *Knowledge Base*).

Przyjęcie w nazwie *Information Object* członu *Information* wynika zapewne z faktu, że – jak już wspomniano – model został stworzony dla celów archiwizacji i wymiany danych cyfrowych zawierających w swej treści określone informacje (w pierwotnym zastosowaniu pochodzące z badań przestrzeni kosmicznej). Z uwagi jednak, że OAIS został zaakceptowany jako standard archiwizacji wszelkich dokumentów cyfrowych – bez względu na ich zawartość informacyjną – można przyjąć, że każdy dokument cyfrowy, należący do zbiorów bibliotecznych, archiwalnych, a także muzealnych, wraz z narzędziami umożliwiającymi jego odczyt będzie określany mianem *Information Object.*

Proces przetwarzania danych cyfrowych do postaci zrozumiałej dla użytkownika przedstawiono na rysunku 2.

***Rysunek 2.* Proces interpretacji i prezentacji treści danych cyfrowych (na podstawie modelu referencyjnego OAIS)**



Źródło: oprac. własne na podstawie [OAIS, 2002]

W referencyjnym modelu OAIS przyjęto zasadę, że dla prawidłowego przebiegu procesu archiwizacji istotne jest dokładne zidentyfikowanie obiektu cyfrowego oraz powiązanych z nim narzędzi. Bez odpowiedniego sprzętu i oprogramowania interpretacja przechowywanych danych cyfrowych nie będzie w przyszłości możliwa. Dlatego też w archiwach działających na podstawie modelu OAIS oprócz danych cyfrowych (ang. *Data Object*) przedmiotem archiwizacji są także

narzędzia (ang. *Representation Information*) umożliwiające odczytywanie danych i udostępnienie ich użytkownikowi w postaci dla niego zrozumiałej.

Kolejnym kluczowym pojęciem w modelu referencyjnym OAIS jest pakiet informacyjny (ang. *Information Package*) [OAIS, 2002]. Składa się on z dwóch komponentów, tj. przechowywanej informacji (ang. *Content Information*) oraz opisu przechowywania (ang. *Preservation Description Information* [PDI]). *Content Information* zawiera dane cyfrowe wraz z narzędziami ich odczytu i prezentacji, natomiast *PDI* to – w myśl modelu OAIS – wszelkie informacje konieczne w procesie przechowywania. Zalicza się tu cztery typy informacji, określane jako [OAIS, 2002]:

- proweniencja (ang. *Provenance*), czyli informacja o pochodzeniu, określa źródło obiektu informacyjnego, wskazuje na podmiot odpowiedzialny za opiekę nad obiektem od momentu jego powstania oraz dostarcza wiedzę na temat historii obiektu;
- kontekst (ang. *Context*) opisuje związek obiektu informacyjnego z innymi obiektami, które nie należą do danego pakietu informacyjnego;
- identyfikatory (ang. *Reference*) to informacje zapewniające jednoznaczną identyfikację obiektu informacyjnego; ich zadaniem jest dostarczenie identyfikatora dokumentu cyfrowego, który odróżnia określony dokument od innych. W archiwach cyfrowych identyfikatory występują pod nazwą *Digital Object Identifier* (DOI) czy też *Persistent Identifier* (PI) [Schöning-Walter, 2008, s. 32-38]. Identyfikatory obiektów cyfrowych są odpowiednikami numerów ISBN oraz ISSN;
- mechanizmy ochrony autentyczności danych (ang. *Fixity*) zabezpieczają autentyczność i integralność obiektów informacyjnych przed jakimikolwiek nieudokumentowanymi zmianami.

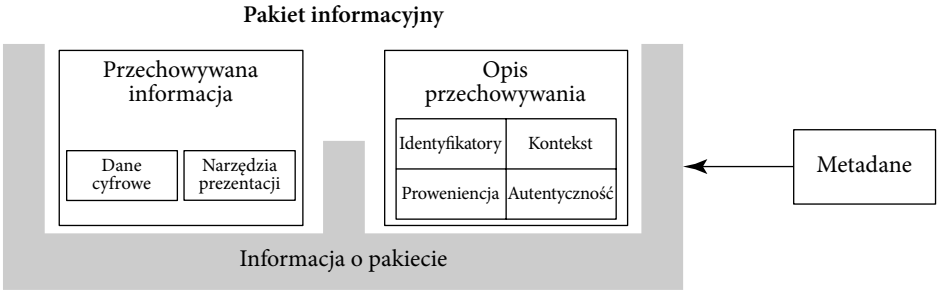
PDI jest więc zarówno pewnego rodzaju informatorem o pochodzeniu i historii obiektu informacyjnego, jego przynależności oraz powiązaniach z innymi obiektami w archiwum, jak i mechanizmem chroniącym jego integralność i autentyczność.

W celu powiązania obu komponentów pakietu informacyjnego model referencyjny OAIS przewiduje także element w postaci informacji o pakiecie (ang. *Packaging Information*). Jego zadaniem jest identyfikacja poszczególnych składników pakietu informacyjnego.

Elementem niezbędnym w archiwum cyfrowym są metadane przechowywanych obiektów (ang. *Information Packages*). W modelu referencyjnym OAIS określane są one terminem *Descriptive Information*. Metadane dostarczają informacje o zawartości pakietu informacyjnego oraz umożliwiają jego odnalezienie w archiwum.

Pakiet informacyjny wraz ze wszystkimi jego elementami składowymi należy traktować jako obiekt archiwizacji w archiwum cyfrowym OAIS (por. Rys. 3).

***Rysunek 3*. Pakiet Informacyjny (*Information Package*) jako obiekt archiwizacji w archiwum cyfrowym OAIS**

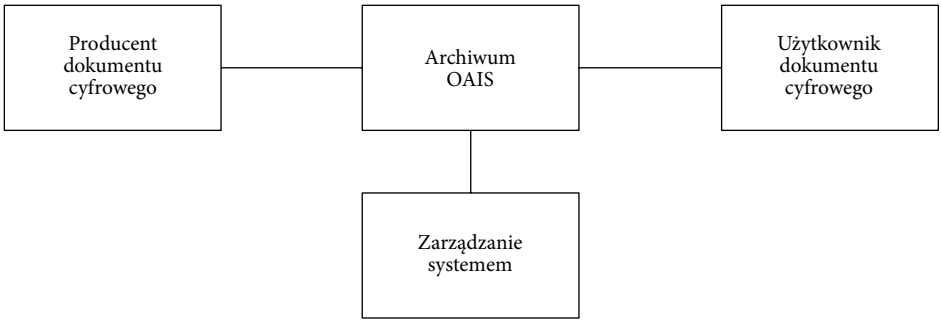


Źródło: oprac. własne na podstawie modelu referencyjnego OAIS [OAIS, 2002]

Otoczenie archiwum w modelu referencyjnym OAIS

Przedstawiony na rysunku 4 model referencyjny OAIS uwzględnia elementy otoczenia archiwum.

***Rysunek 4*. Otoczenie archiwum OAIS**



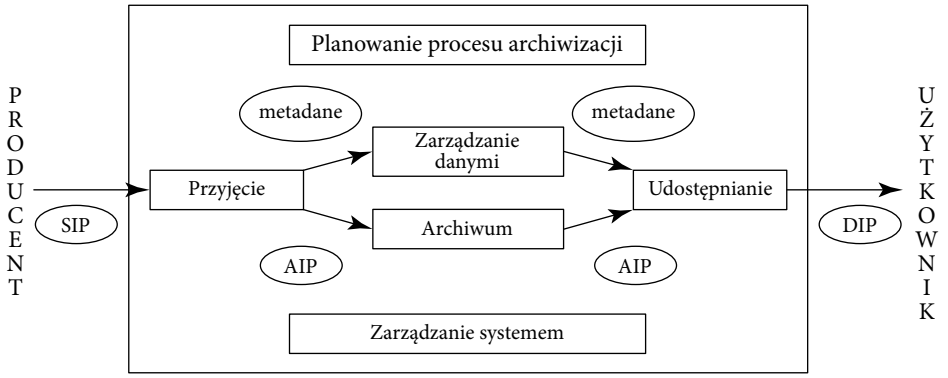
Źródło: oprac. własne na podstawie modelu referencyjnego OAIS [OAIS, 2002]

W punkcie centralnym modelu znajduje się archiwum ze swoimi zadaniami gromadzenia obiektów archiwizacji, ich długoterminowego zabezpieczania oraz udostępniania zainteresowanym kręgom użytkowników. Z archiwum ściśle powiązane są elementy określone jako *Producent*, czyli twórcy dokumentów cyfrowych (autorzy oraz wszelkie instytucje, w których one powstają), *Użytkownik*, czyli użytkownicy, z myślą o których organizuje się procesy archiwizacji zasobów cyfrowych, i wreszcie *Zarządzanie*, a więc jednostka zarządzająca, której działania koncentrują się wokół organizacji archiwum OAIS, organizacji procesów archiwizacji oraz czuwania nad ich prawidłowym przebiegiem.

Schemat funkcjonowania archiwum OAIS

Autorzy modelu OAIS dokonują rozróżnienia pomiędzy pakietami informacyjnymi, które są przechowywane w archiwum OAIS, następnie pakietami, które są dostarczane do archiwum przez producenta (twórcę) oraz pakietami wysyłanymi z archiwum w celu ich udostępnienia użytkownikowi. Pakiet informacyjny przesyłany od producenta do archiwum nazwany został w modelu OAIS zgłoszeniowym pakietem informacyjnym (ang. *Submission Information Packag* [SIP]). Jego forma oraz zawartość są wcześniej ustalone pomiędzy OAIS oraz producentem. Pakiet, który jest przechowywany w archiwum, oznaczono jako archiwizowany pakiet informacyjny (ang. *Archive Information Package* [AIP]). Natomiast dla pakietu udostępnianego użytkownikowi przyjęto określenie udostępnianego pakietu informacyjnego (ang. *Dissemination Information Package* [DIP]). Na zamówienie OAIS przekazuje użytkownikowi całość lub odpowiednią część pakietu informacyjnego AIP w formie pakietu DIP.

Rysunek 5. **Funkcjonowanie archiwum OAIS**



Źródło: oprac. własne na podstawie modelu referencyjnego OAIS [OAIS, 2002]

Rysunek 5 obrazuje istotę funkcjonowania archiwum zasobów cyfrowych zorganizowanego na podstawie modelu OAIS. Obejmuje on sześć jednostek funkcjonalnych oraz drogę obiektu archiwizacji od jego producenta do użytkownika. Jednostka *Przyjęcie* jest odpowiedzialna za przyjęcie zgłoszeniowego pakietu informacyjnego SIP od producenta oraz za przygotowanie go do umieszczenia i administrowania nim w archiwum. W zakresie jej zadań znajduje się m.in.: kontrola kompletności oraz autentyczności zgłoszeniowego pakietu informacyjnego, przekształcenie pakietu SIP w pakiet gotowy do archiwizacji oraz stworzenie do niego metadanych. Następnie archiwizowany pakiet informacyjny AIP przekazywany

jest do jednostki zajmującej się archiwizacją, tj. do *Archiwum*, a metadane odsyłane do jednostki *Zarządzanie danymi* odpowiedzialnej za zarządzanie zasobami archiwalnymi.

Kolejną, bardzo istotną jednostką funkcjonalną systemu OAIS, jest *Archiwum* odpowiedzialne za zapis, właściwe przechowywanie pakietów informacyjnych (AIP) oraz możliwość ich odczytu. *Archiwum* odpowiada za: długoterminowe przechowywanie i zapewnienie nienaruszalności pakietów AIP, okresowe przenoszenie danych na media nowszej generacji, migrację do aktualnie stosowanych formatów lub systemów, a w przypadku awarii systemu – za ich rekonstrukcję. Na żądanie *Archiwum* przekazuje określony pakiet AIP do jednostki *Udostępnianie*.

W archiwum cyfrowym niezbędna jest jednostka *Zarządzanie danymi*. Jej zadaniem jest utrzymywanie i udostępnianie szerokiego wachlarza informacji. Przykładami mogą być katalogi i inwentarze, na podstawie których można uzyskać określone zasoby z archiwum, a także statystyki dotyczące udostępniania zbiorów. Do zadań tej jednostki należy również zaliczyć kontrolę bezpieczeństwa danych oraz inne procedury narzucane przez OAIS.

Poprawne funkcjonowanie całego archiwum jest uzależnione od prac jednostki administrującej procesami w nim zachodzącymi. *Zarządzanie systemem* zajmuje się negocjowaniem warunków z producentami (na których podstawie dokumenty są transferowane do archiwum), czuwa nad kontrolą zgodności dostarczonych obiektów ze standardami archiwum oraz przejmuje odpowiedzialność za utrzymywanie sprawności sprzętu i oprogramowania w archiwum. Czyni także starania na rzecz rozwoju oraz nadzoru nad standardami potrzebnymi dla funkcjonowania archiwum.

Archiwa organizowane oraz działające na podstawie modelu OAIS starają się zapewnić na przyszłość stabilny dostęp do przechowywanych w nich różnorodnych dokumentów cyfrowych. W tym celu w ramach OAIS wyodrębniono jednostkę *Planowanie procesu archiwizacji*, która zajmuje się, m.in.: obserwowaniem rozwoju rynku sprzętu i oprogramowania, testowaniem nowych rozwiązań, kontrolowaniem, czy archiwizowane obiekty można uruchomić i odczytać. Odpowiedzialna jest też za wszelkie decyzje dotyczące strategii postępowania, m.in.: częstotliwości odświeżania danych, działań mających na celu dostosowanie rozwiązań do zmieniających się warunków sprzętowo-programowych (emulacja lub migracja) i udostępnienia treści dokumentów w zmienionych warunkach.

Proces udostępniania dokumentów cyfrowych w archiwach OAIS określony został terminem *Access*. W ramach udostępniania zasobów użytkownikom umożliwia się przeglądanie zawartości archiwum (poprzez katalogi online), określenie lokalizacji i dostępności konkretnych zbiorów. Na zamówienie użytkownika system tworzy i wysyła pakiety informacyjne typu DIP.

Na podstawie przytoczonych wcześniej schematów i opisów możliwe staje się prześledzenie drogi cyfrowego obiektu przez archiwum budowane zgodnie z założeniami OAIS. Producent, który chce przesłać do archiwum dokument cyfrowy w celu jego długoterminowego przechowania, powinien nadać dokumentowi właściwą – ustaloną wcześniej z archiwum – formę oraz dołączyć wszelkie dodatkowe informacje o dokumencie wraz z metadanymi. Jednak realizowane w zakresie archiwizacji projekty pokazują, że producenci nie zawsze podejmują się generowania metadanych, co stanowi dodatkowe zadanie dla pracowników archiwum. Dokument wraz z metadanymi przesyłany jest w postaci zgłoszeniowego pakietu informacyjnego SIP do działu przyjęcia, gdzie zostaje „rozpakowany" oraz sprawdzony pod względem kompletności i poprawności wszelkich niezbędnych informacji. W systemie archiwalnym każdy dokument przyjmuje postać AIP, jest zapisywany na serwerze depozytowym i przechowywany w sposób umożliwiający jego długotrwałą, stabilną użyteczność. Wygenerowane metadane archiwizowanych dokumentów są odsyłane do działu zajmującego się ich zarządzaniem. Archiwizowany obiekt AIP może być przekształcony w formę DIP, udostępnianą użytkownikowi na życzenie, w postaci umożliwiającej jego zrozumienie.

Model referencyjny OAIS wyjaśnia podstawowe terminy oraz identyfikuje kluczowe procesy z zakresu trwałej archiwizacji dokumentów cyfrowych. Przyczynia się także do wzrostu wiedzy na temat projektowania, budowania i funkcjonowania archiwów nastawionych na długoterminową ochronę przechowywanych obiektów cyfrowych oraz ich udostępniania w przyszłości. Model wskazuje również na niebagatelną rolę sprzętu i oprogramowania w utrzymaniu użyteczności cyfrowych obiektów informacyjnych. Uwzględnia potrzebę współpracy z instytucjami niepowiązanymi bezpośrednio z archiwum, jednak współodpowiedzialnymi za efektywność procesów archiwizacji. Dokładnie określa zakres spoczywającej na archiwum odpowiedzialności za ochronę pakietów informacyjnych. Jednakże przede wszystkim model OAIS – przez uznanie go za standard ISO – dostarcza zunifikowanego i powszechnie stosowanego nazewnictwa ułatwiającego międzynarodową dyskusję, wymianę pomysłów oraz doświadczeń w środowiskach zainteresowanych problematyką długoterminowej archiwizacji zasobów cyfrowych.

Autorzy modelu zapewniają, że OAIS może być stosowany przy organizacji wszelkich cyfrowych archiwów, ze specjalnym przeznaczeniem dla archiwów odpowiedzialnych za długoterminowe przechowywanie dokumentów [OAIS, 2002]. Przykładami potwierdzającymi tę tezę są realizowane z powodzeniem światowe projekty archiwizacji, wykorzystujące model referencyjny OAIS. Jedną z największych i pierwszych inicjatyw w tym obszarze przedsięwzięła narodowa biblioteka Holandii Koninklijke Bibliotheek (KB). W kooperacji z firmą IBM Koninklijke

Biblioteheek opracowała i zaimplementowała prototyp archiwum elektronicznego *Digital Informations and Archiving System* (DIAS), które w założeniu powinno zapewnić dostępność publikacji cyfrowych w przyszłości [Amse, 2003; Werf-Davelaar, 1999]. System DIAS stanowi jądro systemu depozytowego dla zasobów cyfrowych zgromadzonych w Koninklijke Bibliotheek – eDepot [Long-term, b.d.; Oltmans i Lemmen, 2006, s. 61-67; Ras, 2009]. System DIAS znalazł również zastosowanie w niemieckim systemie depozytowym – „Kopal-Archivsystem" [Kopal, 2007; Kopal, b.d.].

Na marginesie rozważań o modelu referencyjnym OAIS warto zaznaczyć, że jest on stale rozbudowywany. W 2004 r. został uzupełniony o model Producer-Archive Interface – Methodology Abstract Model (PAIMAS), znany jako norma ISO 20652:2006. Model PAIMAS stworzono w celu jednoznacznego uregulowania kwestii współpracy twórców (wydawców, producentów) dokumentów cyfrowych z systemami archiwalnymi. Zidentyfikowano, zdefiniowano i unormowano poszczególne czynności: od nawiązania kontaktu pomiędzy twórcą a systemem depozytowym, po akceptację i przyjęcie obiektu archiwalnego do depozytu. PAIMAS precyzuje procesy zachodzące w jednostkach *Zarządzanie systemem* oraz *Przyjęcie* OAIS [ISO, 2006b; PAIMAS, 2003]. Dopełnienie PAIMAS stanowi standard *Space data and information transfer systems – Producer-Archive Interface Specification* (PAIS) (ISO 20104:2015). Opracowanie modelów PAIMAS i PAIS oraz przyznanie im statusu normy ISO skłania do wniosku, że z czasem prawdopodobnie wszystkie procesy zachodzące zarówno wewnątrz archiwum OAIS, jak i w jego otoczeniu zostaną ustandaryzowane.

Szczegółowa analiza elementów składowych modelu OAIS, ich funkcji oraz możliwości praktycznego zastosowania, jest zawarta w poszczególnych seriach raportów CCSDS oraz książce *Advanced Digital Preservation* [CCSDS, b.d.; Giaretta, 2011]. Na rok 2017 zaplanowano przegląd i weryfikację modelu OAIS. Od jakiegoś czasu trwają konsultacje i gromadzenie sugestii zmian. W dyskusji aktywnie uczestniczą organizacje zajmujące się ochroną cyfrowych zasobów nauki i kultury [DPC, 2016]. Wydarzenia relacjonują m.in. Digital Preservation Coalition oraz Nestor.

1.4.4. Przykładowe wdrożenia systemów depozytowych dla dokumentów cyfrowych

Dotychczasowa wiedza o postępowaniu z materiałem cyfrowym w instytucjach pamięci wywodzi się m.in. z pierwszych projektów archiwizacyjnych realizowanych w krajach europejskich. W książce przybliżono wybrane inicjatywy dotyczące ochrony zasobów cyfrowych.

Deposit System for Electronic Publications (DSEP) – system depozytowy dokumentów cyfrowych w Holandii

Jednym z przykładowych wdrożeń systemu depozytowego w instytucji pamięci jest holenderski system DSEP. Dyskusje na temat utworzenia systemu depozytowego dla publikacji elektronicznych zgromadzonych w centralnej bibliotece Holandii Koninklijke Bibliotheek podjęto w 1995 r. W ramach projektu Depot for the Dutch Electronic Publications – Innovative Scientific Informationstructure (DNEP-IWI), realizowanego w latach 1996-1998, prowadzono prace badawcze polegające głównie na teoretycznych studiach literatury przedmiotu oraz zdefiniowaniu procesów, które zachodzą zarówno wewnątrz, jak i w otoczeniu archiwów cyfrowych, i które powinny zostać uwzględnione w projektowaniu systemu depozytowego. Był to projekt o charakterze przygotowawczym do podjęcia właściwych prac nad zaprojektowaniem i stworzeniem systemu depozytowego dla holenderskiego zasobu cyfrowego [e-Depot, b.d.; History: the KB, b.d.; Werf-Davelaar, 1999].

W latach 1998-2000 Koninklijke Bibliotheek kierowała projektem „Networked European Deposit Library – NEDLIB", którego uczestnicy – tj.: przedstawiciele ośmiu europejskich bibliotek narodowych, jednego archiwum, dwóch firm informatycznych oraz trzech wydawnictw naukowych – postawili sobie za cel szukanie rozwiązań będących w stanie zapewnić długoterminowy dostęp do elektronicznych zbiorów bibliotecznych, archiwalnych oraz muzealnych [Steenbakkers, 2000, s. 21-22; Werf-Davelaar, 1999]. W rezultacie projektu powstały, m.in., założenia systemu depozytowego dla publikacji elektronicznych DSEP *Deposit System for Electronic Publications*, opracowane na podstawie modelu referencyjnego OAIS. Z uwagi na fakt, że pierwotna wersja modelu OAIS z 1999 r. nie zakładała długoterminowej archiwizacji danych cyfrowych, a jedynie ich przechowywanie dla potrzeb bieżącego użytku, w czasie realizacji projektu NEDLIB konieczne stało się wprowadzenie jednostki funkcjonalnej odpowiedzialnej za długoterminowe utrzymanie dostępności i użyteczności zarchiwizowanego materiału cyfrowego. W konsekwencji powstała jednostka *Planowanie procesu archiwizacji* (ang. *Preservation Planning*), wbudowana następnie do wersji modelu OAIS – zaakceptowanej i obowiązującej obecnie jako standard ISO w obszarze archiwizacji dokumentów cyfrowych [OAIS, 2002].

Założeniem projektu NEDLIB było opracowanie takiego systemu, który umożliwiałby realizację zadań związanych z długoterminową archiwizacją zasobów cyfrowych w ścisłej współpracy z funkcjonującym już systemem bibliotecznym, archiwalnym lub muzealnym. System depozytowy miałby stanowić dopełnienie procesów gromadzenia, opracowania, przechowywania, utrzymywania w należytym stanie oraz udostępniania zbiorów cyfrowych [Borghoff i in., 2003, s. 32]. Bez

względu na to, w jakiej instytucji archiwizującej miałby zostać zaimplementowany system DSEP, ma on wykorzystywać infrastrukturę funkcjonującego w niej systemu i stanowić jego integralną część (por. Rys. 6).

***Rysunek 6*. Model systemu depozytowego dla publikacji elektronicznych DSEP**

Źródło: oprac. własne na podstawie [Steenbakkers, 2000, s. 21-22]

Selekcja (ang. *Selection*) – pierwszy wymieniony w modelu moduł reprezentuje wszelkie czynności związane z wyborem spośród dostępnych publikacji elektronicznych odprowadzanych do biblioteki, tych, które powinny znaleźć się w systemie depozytowym, w celu ich długoterminowej archiwizacji. W procesie selekcji decydującą rolę pełnią ludzie opracowujący strategię włączania lub wykluczania publikacji elektronicznej z kolekcji depozytowej. Strategia doboru publikacji jest kwestią indywidualną biblioteki i zwykle bazuje na narodowej polityce

gromadzenia. Czynnikami decydującymi mogą być również umowy z wydawcami dostarczającymi publikacje oraz wytyczne informatyków sprawujących opiekę techniczną nad archiwum elektronicznym [Werf, 2000, s. 7-8]. Chodzi o określenie parametrów, którymi powinny charakteryzować się dokumenty cyfrowe, aby mogły zostać wprowadzone do depozytu i podlegać procesom archiwizacji. Jednym z nich jest likwidacja przez wydawców mechanizmów zabezpieczających publikację przed jej nielegalnym kopiowaniem oraz tzw. „time locków", czyli blokad czasowych, regulujących okres dostępności publikacji [Amse, 2003]. Twórcy systemu depozytowego utrzymują, że umożliwia on przechowanie dokumentów elektronicznych bez względu na format, w jakim zostały zapisane oraz nośnik, na którym występują. Ich zdaniem system depozytowy powinien, według założeń, zachować też zdolność przyjmowania w przyszłości do depozytu publikacji elektronicznych w aktualnie nieznanych formatach [Steenbakkers, 2000, s. 21-22; Werf, 2000, s. 7-8]. Świadczyłoby to o otwartości i elastyczności architektury systemu depozytowego, co w przypadku heterogeniczności dokumentów cyfrowych ma niebywale istotne znaczenie.

Ważnym czynnikiem wpływającym na procesy selekcji jest kondycja finansowa instytucji archiwizującej [Werf, 2000, s. 7-8], bowiem z jej budżetu muszą zostać wyasygnowane środki na zaprojektowanie i zbudowanie systemu depozytowego o odpowiedniej pojemności, w dalszej zaś kolejności na jego implementację, testowanie, stałe utrzymywanie oraz (w razie potrzeby) rozbudowę. Za idealną można uznać sytuację, w której biblioteki narodowe mogłyby pozwolić sobie na wdrożenie i utrzymanie systemu depozytowego mieszczącego wszystkie opublikowane w kraju dokumenty cyfrowe, tak by chociaż ta jedna instytucja w kraju zachowała pełny dorobek jego nauki i kultury. Pomimo że ani w literaturze przedmiotu, ani w innych istniejących źródłach nie wspomina się o nakładach finansowych koniecznych dla uruchomienia takiego projektu, na jaki pozwoliła sobie holenderska Koninklijke Bibliotheek, pewnym jest, że instytucje biblioteczne dążące do utworzenia i utrzymywania systemu depozytowego dla publikacji elektronicznych muszą liczyć się z poważnymi obciążeniami finansowymi.

Kolejny moduł systemu to *Gromadzenie* (ang. *Acquisition*). Nazwą tą określa się w praktyce bibliotecznej wszelkie czynności związane z nabywaniem bądź otrzymywaniem zbiorów. W piśmiennictwie przedmiotu zaznacza się, że proces ten w przypadku publikacji elektronicznych nie różni się niczym szczególnym od procesu gromadzenia dokumentów tradycyjnych. Nieco bardziej skomplikowane może okazać się jedynie postępowanie związane z nabywaniem praw użytkowania niektórych form publikacji [Liegmann, 2001, s. 106-109]. Przykładem są czasopisma elektroniczne, których wydawcy godzą się wprawdzie – w ramach udzielonych licencji – na ich umieszczanie i przechowywanie w systemie depozy-

towym biblioteki, jednak czynią przy tym pewne obostrzenia, dotyczące choćby udostępniania treści dokumentów lub podejmowania czynności typowych dla procesów archiwizacji jak migracja, która czasem okazuje się niezbędna dla utrzymania użyteczności publikacji elektronicznych. Mogą jednak prowadzić do zmian w dokumentach, naruszając tym samym ich autentyczność.

W ramach procesu gromadzenia publikacji elektronicznych następuje też wymiana informacji bibliograficznych pomiędzy wydawnictwami a depozytem bibliotecznym oraz dostarczanie przez wydawców metadanych archiwizowanych dokumentów oraz wszelkich informacji, na których podstawie metadane mogą zostać wygenerowane [Werf, 2000, s. 7-8].

Z uwagi na fakt, iż biblioteczne systemy depozytowe powstają głównie przy bibliotekach narodowych, wszelkie archiwizowane tam publikacje elektroniczne powinny być przekazywane bezpłatnie, w ramach przysługujących im egzemplarzy obowiązkowych, ewentualnie na bazie licencji lub innych rodzajów umów zawieranych pomiędzy biblioteką a wydawcami [Werf, 2000, s. 7-8]. Z doświadczeń współpracy bibliotek narodowych z firmami wydawniczymi wynika, że tylko nieliczni wydawcy są świadomi potrzeby współpracy w tym zakresie. Problem polega na tym, że ustawy o bibliotekach narodowych nie regulują kwestii odprowadzania egzemplarza obowiązkowego wszelkich ukazujących się publikacji elektronicznych – mowa tu głównie o publikacjach sieciowych. Pomimo starań i ponawianych prób zawierania umów z wydawcami, tylko nieliczni zobowiązują się odprowadzać opublikowane przez nich dokumenty cyfrowe do biblioteki w celu ich długoterminowej archiwizacji. Nasuwa się wniosek, że niezbędne jest nowelizowanie istniejących ustaw o bibliotekach narodowych, gdyż tylko poprzez nowe regulacje prawne biblioteki narodowe zyskają szansę na realizację zadań, do których zostały powołane, tj.: gromadzenia, przechowywania oraz udostępniania kompletnej narodowej kolekcji dorobku nauki i kultury.

Trzeci moduł systemu – nazwany *Capture*, spotykany też pod nazwą *delivery* lub *harvest* – odpowiedzialny jest za przyjęcie publikacji elektronicznej do depozytu. W praktyce wygląda to tak, że publikacje elektroniczne offline dostarczane są do instytucji archiwizującej za pomocą tradycyjnych form transportu, natomiast publikacje online są transferowane przez Internet. W szczególnych przypadkach biblioteki cyfrowe dopuszczają dostarczanie publikacji sieciowych na elektronicznych mediach przenośnych, by następnie umieścić je na serwerze i udostępniać w sieci. Sposób dostarczenia publikacji do archiwum jest kwestią indywidualnych umów pomiędzy wydawnictwami a instytucjami archiwizującymi.

W cytowanych już źródłach, opisujących system DSEP, zwraca się uwagę na trzy następujące po sobie kroki postępowania w ramach procesu wprowadzania publikacji elektronicznej do depozytu:

- ustalenie źródła pochodzenia przysłanej do biblioteki publikacji elektronicznej (ang. *authentication*). Celem tego zabiegu jest sprawdzenie, czy publikacja pochodzi od wydawcy umieszczonego w rejestrze wydawców współpracujących z biblioteką;
- działania, które mają zapewnić, że wprowadzana do depozytu publikacja jest kompletna i wolna od wirusów oraz spełnia inne wymogi instytucji archiwizującej (ang. *quality scan*);
- dostarczenie do depozytu publikacji elektronicznej w postaci pakietu informacyjnego typu *Submission Information Package* (SIP), (ang. *transfer to deposit*).

Biblioteki zwykle nie mają wpływu na to, w jakich formatach zapisu danych publikowane są dokumenty elektroniczne. Przyjmują wszystkie możliwe publikacje, troszcząc się następnie o ich konwersję do formatów odpowiadających standardom DSEP i zapewniających ich przechowanie na przyszłość [Borghoff i in., 2003, s. 34]. Nie tylko konwertowanie treści publikacji elektronicznych (w trakcie włączania ich do kolekcji depozytowej) z nietypowych do standardowych formatów zapisu danych, ale i – niejednokrotnie konieczne w procesie archiwizacji – migrowanie do nowszych formatów, budzą pewne zastrzeżenia, bowiem wiążą się ze zmianami oryginalnej formy publikacji. Kłóci się to z głównym założeniem archiwizacji odnośnie do zachowania autentyczności publikacji elektronicznych. Celem procesu długoterminowej archiwizacji jest przechowanie dla przyszłych pokoleń użytkowników kompletnego zbioru autentycznych i integralnych publikacji elektronicznych [Attributes, 2001], co oznacza, że należy unikać wszelkich działań, w których wyniku pierwotna treść i forma publikacji ulegną zmianom. Dla specjalistów projektujących systemy depozytowe i odpowiedzialnych za utrzymanie ich zawartości to trudne zadanie, któremu jednak trzeba wyjść naprzeciw. Niestety nie jest to jedyna skomplikowana kwestia związana z archiwizowaniem publikacji elektronicznych i dlatego, m.in. zostało ono określone mianem wyzwania dla instytucji archiwizujących oraz dla wydawnictw [Börsenverein, 1996].

W ramach czwartego modułu, tj. *Rejestracji* (ang. *Registration*), następuje rejestracja pakietu informacyjnego wprowadzanego do depozytu. Rejestracja może obejmować też czynności związane z wysyłaniem potwierdzeń przyjęcia publikacji do depozytu bądź jej odrzucenia [Werf, 2000, s. 7-8]. Wprowadzony do depozytu dokument cyfrowy podlega kolejnym procesom kontrolnym, określanym w modelu jako *Kontrola* (ang. *Verification*), których głównym celem jest ustalenie autentyczności oraz integralności dokumentu cyfrowego. Weryfikacja przebiega w czterech fazach [Nedlib, b.d.]:

- *Validate Package* – kontrola integralności pakietu informacyjnego. System dokonuje też sprawdzenia, czy pakiet odpowiada ustalonym standardom;

- *Check logical Integrity* – rozpakowanie pakietu informacyjnego w celu ustalenia obecności wszystkich potrzebnych komponentów, tj.: danych reprezentujących treść publikacji, metadanych oraz narzędzi niezbędnych do odczytania i interpretacji danych cyfrowych;
- *Establish authenticity* – oznakowanie autentyczności kopii publikacji przyjętej do depozytu w celu archiwizacji;
- *Installation and testing* – publikacja wraz z oprogramowaniem wspomagającym jest wczytywana do terminalu w celu poddania jej testom na właściwe działanie.

Pakiety informacyjne, które pomyślnie przeszły proces weryfikacji, są poddawane dalszym procesom w obrębie systemu depozytowego. Jednak specjaliści dopuszczają też sytuację, w której proces weryfikacji kończy się odesłaniem publikacji do wydawcy. Sporządzają wówczas stosowny raport o wykrytych błędach, uzasadniający odrzucenie publikacji przez system, dodatkowo o zaistniałym fakcie informują dział *Gromadzenia* (ang. *Acquisition*) [Werf, 2000, s. 7-8].

Weryfikacja należy do jednego z trudniejszych i bardziej skomplikowanych procesów. Wynika to z faktu, że obiekty cyfrowe mogą zostać podrobione lub zmienione w dużo łatwiejszy sposób niż analogowe. Problemy dostarcza też kontrola integralności publikacji sieciowych. Tendencja łączenia za pomocą hiperłączy treści różnych publikacji znacznie utrudnia określenie „granic" poszczególnych publikacji. Wciąż bez odpowiedzi pozostaje pytanie, jak w takiej sytuacji zautomatyzować proces kontroli integralności publikacji [Liegmann, 2001, s. 106-109]. W tym celu niewątpliwie konieczne są mechanizmy radzące sobie z odróżnieniem zawartości publikacji, tzn. rozpoznające linki wewnętrzne, należące do podlegającego archiwizacji obiektu, oraz zewnętrzne, stosowane na zasadzie odsyłaczy do treści innych publikacji sieciowych. Dostarczenie pomysłu i stworzenie takich mechanizmów to kolejna, otwarta wciąż kwestia w temacie archiwizacji publikacji elektronicznych.

Moduł *Opis* (ang. *Description*) reprezentuje etap prac bibliotecznych związanych z katalogowaniem zbiorów. Celem procesu katalogowania jest wprowadzenie opisów bibliograficznych publikacji elektronicznych wraz z metadanymi do systemu katalogowego zbiorów bibliotecznych (OPAC), by mogły być przez ten system wyszukiwane. Katalogowanie odbywa się zwykle z uwzględnieniem narodowych norm katalogowania. Zakłada się, że opisy bibliograficzne publikacji elektronicznych archiwizowanych długoterminowo w depozytach będą ujmowane w bibliografii narodowej [Werf, 2000, s. 7-8].

W procesie katalogowania generowane są też „techniczne" metadane, które stanowią ważną kategorię danych o publikacji i są niezbędne dla właściwego przebiegu prac w zakresie długoterminowego utrzymania użyteczności publikacji elektronicznych [Liegmann, 2001, s. 106-109].

Po skatalogowaniu publikacji elektronicznych następuje ich przyjęcie do depozytowego systemu przechowywania danych, reprezentowanego przez moduł *Przechowywanie* (ang. *Storage*). W założeniu depozytowy system przechowywania publikacji elektronicznych ma zapewniać regularną kontrolę możliwości ich odczytu i interpretacji; w przypadku zagrożenia utraty danych powinny zostać podjęte wszelkie działania na rzecz ich utrzymania, natomiast w przypadku utraty danych – poczynione próby ich odzyskania. Głównym zadaniem modułu *Przechowywanie* jest zapewnienie odczytu kodu zero-jedynkowego w jego oryginalnej formie, co odbywa się za pomocą:

- regularnego odświeżania nośnika zapisu danych cyfrowych (ang. *periodic refreshment*);
- okresowego sporządzania kopii zapasowych danych (ang. *continuous backup*) [Nedlib, b.d.].

Jednocześnie zwraca się uwagę, że dla utrzymania kodu zero-jedynkowego istotne jest zagwarantowanie bezpieczeństwa miejscu, w którym przechowywane są nośniki z jego zapisem. Konieczne są ustalenia regulujące kwestie dostępu do archiwum czy przeprowadzania prac na dokumentach cyfrowych oraz dotyczące zabezpieczenia archiwum przed ewentualnymi katastrofami [Werf, 2000, s. 7-8]. Jednym z elementów strategii długoterminowej archiwizacji ma być tworzenie systemów depozytowych o pojemności, która umożliwiałaby przechowanie oprócz narodowej kolekcji cyfrowej, wraz z jej kopią zapasową, także kopii zapasowej kolekcji instytucji partnerskiej. Jest to sposób na zabezpieczenie kopii zapasowej archiwum w miejscu terytorialnie oddalonym od archiwum głównego. Cel ten jednak może zostać osiągnięty tylko i wyłącznie przy ścisłej współpracy bibliotek na rzecz wypracowania zunifikowanych – w skali międzynarodowej – metod archiwizacji dokumentów cyfrowych, w szczególności przyjęcia i stosowania jednolitych parametrów dla budowanych systemów depozytowych.

Kolejny moduł – *Archiwizacja* (ang. *Preservation*) – odzwierciedla wszelkie czynności, wymagane dla długoterminowej ochrony kolekcji depozytowej. Nie można jednak jasno zdefiniować obszaru jego zadań. Innymi słowy, nie ma sprecyzowanego planu działania czy konkretnej recepty stosowanej w celu zarchiwizowania publikacji elektronicznych. Ich długoterminowa ochrona wymaga użycia różnych – w zależności od konkretnej sytuacji – metod archiwizacji. Wspominano już, że warunkiem koniecznym dla przedsięwzięcia jakiejkolwiek strategii archiwizacji jest obserwowanie zmian zachodzących w technologii, co zostało określone terminem *Technology Watch*. W konsekwencji obserwacji postępu technologicznego podejmowane są – zależnie od potrzeby – następujące działania na publikacjach elektronicznych: zmiana formatu dokumentu, emulacja jego systemowego otoczenia, odświeżenie medium, zmiana rodzaju medium, aktualizacja

„technicznych" metadanych, a także kontrola jego integralności oraz autentyczności. Trzeba podkreślić, że niektóre z wymienionych działań mogą prowadzić do zmian w formie i treści publikacji elektronicznych, a tym samym do utraty ich autentyczności. Zachodzi zatem konieczność dokładnego dokumentowania wszelkich prac podejmowanych w procesie archiwizacji oraz ich efektów, bowiem tylko poprzez dokładny opis każdej przeprowadzanej na obiekcie operacji (z uwzględnieniem porównania wersji oryginalnej z tą, którą uzyskuje się w rezultacie przeprowadzanych operacji) możliwe będzie w przyszłości odtworzenie jego historii, śledzenie wszelkich zmian, porównanie wersji oryginalnej z zachowaną [Liegmann, 2001, s. 106-109].

W raportach NEDLIB [Werf, 2000, s. 7-8] procesowi *Archiwizacja* poświęca się szczególnie wiele uwagi. Zaznacza się w nich, że spełnia kluczową rolę w systemie depozytowym i pieczołowicie opisuje jego części składowe. Na proces *Archiwizacja* składają się tzw. subprocesy, tj. *Planowanie archiwizacji* (ang. *Preservation Planning*) oraz *Działania archiwizacyjne* (ang. *Preservation Activities*). Pierwszy z subprocesów odpowiada za identyfikację wszelkich problemów i zadań, które wiążą się z długoterminową archiwizacją publikacji elektronicznych oraz za dostarczanie pomysłów i narzędzi do ich rozwiązywania. W jego ramach prowadzi się badania i opracowuje strategię ochrony danych cyfrowych w następujących obszarach:

- *Development Preservation Standards & Strategies* – rozwój standardów i strategii umożliwiających archiwum aktywne uczestnictwo w procesach publikowania elektronicznego, reagowanie na rozwój technologiczny i dostosowywanie polityki archiwizacji do zmieniających się form publikacji elektronicznych;
- *Development Packaging Designs* – projektowanie i dopasowywanie do zmieniających się okoliczności, standardów i formatów pakietów informacyjnych typu SIP, AIP oraz DIP;
- *Reference Platforms Defining* – identyfikacja platformy programowo-sprzętowej, konieczna do uruchomienia i udostępnienia publikacji elektronicznej. W procesie tym następuje przyporządkowanie konkretnego sprzętu i oprogramowania do każdego typu publikacji elektronicznych przechowywanych w archiwum;
- *Technology Monitoring* znany także jako *Technology Watch* – obserwacja zmian i rozwoju branży IT, a w szczególności tych jej gałęzi, które są powiązane z przechowywaniem danych cyfrowych oraz formatami zapisu dokumentów cyfrowych.

Subproces *Preservation Activities* odpowiedzialny jest za koordynację wszelkich czynności i zadań podejmowanych w ramach procesu archiwizacji publikacji elektronicznych oraz kontrolę ich zgodności ze standardami DSEP. W zakresie *Preservation Activities* dokonuje się następujących działań:

- tworzenia i implementacji nowej platformy programowo-sprzętowej w systemie depozytowym DSEP (ang. *Create New Reference Platform*) – działania te mają miejsce wówczas, gdy zachodzi potrzeba przeniesienia kolekcji depozytowej do nowego otoczenia sprzętowo-programowego;
- przekształcania przechowywanych w archiwum pakietów typu AIP do nowszej postaci w przypadku zmiany projektu archiwalnego pakietu informacyjnego lub aktualizowanie ich zawartości w związku z przeprowadzaniem procesów emulacji bądź migracji (ang. *Archival Information Update*).

Dziewiąty moduł – *Dostarczanie* (ang. *Delivery*) – reprezentuje działania na rzecz przygotowania kopii archiwizowanej publikacji elektronicznej do jej udostępnienia zainteresowanym użytkownikom. Proces ten umożliwia wyszukiwanie publikacji przechowywanych w depozycie, w razie potrzeby sporządzenie kopii określonej publikacji, oraz jej przesłanie do działu udostępniania. W opinii specjalistów publikacja ma zostać tak przygotowana, aby użytkownik mógł otrzymać, w zależności od zapotrzebowania, jej całość lub odpowiedni fragment, w celach: przeglądnięcia, wydruku lub zapisu na nośnik. System przewiduje też opcję zlecenia wydruku publikacji na żądanie (ang. *print on demand*) [Werf, 2000, s. 7-8]. Dostarczenie publikacji użytkownikowi powinno odbywać się w możliwie standardowej formie, tj. takiej, która stawia najmniej specyficzne wymagania dotyczące warunków użytkowania. Pakiet informacyjny *Dissemination Information Package* (DIP) powinien zawierać wszelkie elementy konieczne do użytkowania publikacji.

W modelu zawarty jest też typowy dla systemów bibliotecznych moduł *Udostępnianie* (ang. *Access*), reprezentujący te elementy infrastruktury systemu bibliotecznego, które odnoszą się do otoczenia użytkownika końcowego oraz gwarantują mu dostęp do bibliotecznych zbiorów. *Access* obejmuje m.in.: dostępność narzędzi wyszukiwawczych, identyfikację użytkownika, zarządzanie prawami użytkowników, określanie profilu użytkownika [Werf, 2000, s. 7-8]. System depozytowy publikacji elektronicznych korzysta z katalogu systemu bibliotecznego, dlatego użytkownik nie jest zmuszony do przeszukiwania dwóch oddzielnych katalogów. Do zalet należy możliwość zamówienia dokumentu depozytowego poprzez istniejący i znany już użytkownikom system biblioteczny [Liegmann, 2001, s. 106-109].

Ostatnim modułem jest *Monitorowanie* (ang. *Monitoring*) – związane z kontrolą jakości pracy zarówno całego systemu, jak i efektywności poszczególnych zachodzących w nim procesów [Nedlib, b.d.].

Zaprezentowany graficznie i opisany wcześniej model, skomponowany z modułów systemu biblioteki cyfrowej i systemu depozytowego publikacji elektronicznych, pokazuje, które zadania przejmuje system depozytowy DSEP oraz w jaki sposób współtworzy wraz z funkcjonującym już systemem biblioteki cyfrowej całość. Wzajemnie uzupełniające się moduły obu systemów mają zapewnić nie

tylko zgromadzenie, opracowanie i bieżące udostępnianie publikacji elektronicznych, lecz również ich przechowanie dla przyszłych pokoleń użytkowników.

Modele OAIS oraz DSEP są od czasu zakończenia i opublikowania rezultatów projektu NEDLIB, tj. od 2000 r., postrzegane jako istotny krok naprzód w dziedzinie długoterminowej archiwizacji zasobów cyfrowych. Pomimo dość dużego stopnia szczegółowości, są to opisy tylko modeli funkcjonalnych, natomiast nie dostarczają one gotowych struktur dla implementacji w praktyce [Borghoff i in., 2003, s. 36; Werf, 2000, s. 7-8].

Należy jednak uwzględnić fakt, że po zakończeniu projektu NEDLIB w Koninklijke Bibliotheek trwały nieprzerwanie intensywne prace nad „wbudowaniem" założeń modelu OAIS oraz DSEP w zdolny do implementacji system depozytowy. W grudniu 2002 r. w Koninklijke Bibliotheek został wdrożony i uruchomiony system depozytowy eDepot [History: the KB, b.d.].

Kopal-Archivsystem – depozytowy system dla niemieckiego zasobu cyfrowego

Do problematyki długoterminowej archiwizacji zasobów cyfrowych w Niemczech odniesiono się po raz pierwszy, podobnie jak w Holandii, w 1995 r. Wówczas na jednym z posiedzeń Deutsche Forschungsgemeinschaft (DFG) uznano długoterminową ochronę zasobów cyfrowych za jeden z obszarów działalności i rozwoju niemieckich bibliotek cyfrowych. Jednocześnie podjęto decyzję o współpracy na rzecz połączenia istniejących zasobów cyfrowych i utworzenia jednej kolekcji niemieckiego zasobu cyfrowego, chronionej na podstawie zunifikowanego programu archiwizacji, opracowanego i akceptowanego przez wszystkie środowiska zainteresowane długotrwałym zabezpieczeniem zgromadzonych materiałów cyfrowych [Jehn i Schrimpf, 2009]. W ramach współpracy wyłoniona została grupa osób, które podjęły inicjatywę zaplanowania i realizacji pierwszego projektu dedykowanego stricte zagadnieniom długoterminowej archiwizacji. Projekt Digital Library Konzepte, finansowany ze środków Bundesministerium für Bildung und Forschung (BMBF), realizowany był w 2003 r. i trwał sześć miesięcy. Najbardziej zaangażowaną instytucją w tym projekcie była Deutsche Nationalbibliothek. W ramach projektu w siedzibie DNB we Frankfurcie nad Menem odbyły się warsztaty, w których do udziału zaproszono głównie bibliotekarzy, bibliotekoznawców i informatologów, ale również archiwistów, muzealników oraz wydawców. Celem warsztatów było podjęcie dyskusji na temat długoterminowej archiwizacji niemieckich zasobów cyfrowych [Nestor, 2003; Workshop, 2003]. W wyniku projektu Digital Library Konzepte powstała dokumentacja, w której długoterminową archiwizację dokumentów cyfrowych zaakceptowano jako nowe zadanie w dotychczasowej działalności instytucji. Zdefiniowano najważniejsze

obszary działań oraz konkretne czynności, które powinny zostać podjęte w celu opracowania programu długoterminowej archiwizacji zasobów cyfrowych dla niemieckich instytucji pamięci. Zadeklarowano również, że program będzie opracowywany we współpracy wszelkich zainteresowanych środowisk pod przewodnictwem Deutsche Nationalbibliothek.

Kontynuacją Digital Library Konzepte był projekt Nestor, finansowany również z budżetu niemieckiego ministerstwa BMBF, realizowany od 2003 do 2009 r. Podstawowym założeniem projektu było stworzenie centrum kompetencyjnego do spraw długoterminowej archiwizacji dokumentów cyfrowych. W jego ramach zrzeszono specjalistów z bibliotek, archiwów, muzeów, wydawnictw i miejsc wydawniczych, instytutów badawczych, uczelni wyższych, urzędów, instytucji medialnych, centrów informatycznych i innych instytucji, w celu opracowania wspólnego programu działań, podziału kompetencji i zadań. Założono, że wyzwaniu długoterminowej archiwizacji uda się wyjść naprzeciw konsolidując zasoby wiedzy, doświadczeń i środków w skali całego kraju oraz przy uwzględnieniu doświadczeń bardziej zaawansowanych organizacji i instytucji z innych krajów. Efektem projektu Nestor jest działające i obecnie popularne już w świecie centrum kompetencyjne Nestor – Kompetenznetzwerk für digitale Langzeitarchivierung in Deutschland. Nestor jest również platformą, która służy zainteresowanym środowiskom informacjami, wiedzą i doświadczeniem w zakresie długoterminowej archiwizacji. Do jej zadań należy organizowanie rozmaitych form kształcenia i dokształcania na wszystkich stopniach – od podstawowego, ogólnego zakresu wiedzy do poziomu kompetencji potrzebnych przy wykonywaniu czynności specjalistycznych, związanych z trwałą ochroną zasobów cyfrowych. W ramach projektu Nestor prowadzone są także prace nad identyfikacją i analizą norm, standardów i rekomendowanych rozwiązań w zakresie długoterminowej ochrony cyfrowych zasobów archiwalnych. W wyniku tych prac wszelkie niemieckie inicjatywy archiwizacyjne mają być projektowane i realizowane z uwzględnieniem najlepszych praktyk i wzorców [Nestor, 2003; Nestor, 2009].

W początkowej fazie realizacji projektu Nestor zapadła decyzja o konieczności zaprojektowania i budowy archiwalnego systemu depozytowego zdolnego do przechowania i zagwarantowania użyteczności niemieckich zasobów cyfrowych w długim czasie. Projekt i budowa niemieckiego systemu depozytowego stała się przedmiotem oddzielnego projektu Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen – Kopal [Altenhöner i in., 2008; Januszko-Szakiel, 2009a].

Projekt Kopal był realizowany od lipca 2004 do czerwca 2007 r. pod kierownictwem Deutsche Nationalbibliothek (DNB), przy finansowym wsparciu z budżetu Bundesministerium für Bildung und Forschung (BMBF). W ramach tego projektu współpracę podjęły: centralna biblioteka Niemiec – Deutsche National-

bibliothek, miejska i uniwersytecka biblioteka miasta Getyngi – Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB), niemiecki oddział firmy IBM (IBM Deutschland GmbH) oraz firma informatyczna z Getyngi specjalizująca się w przetwarzaniu danych zawierających informacje naukowe – Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG). Celem projektu było stworzenie, implementacja i praktyczne sprawdzenie działania systemu długoterminowej archiwizacji niemieckich zasobów cyfrowych. Założono, że będzie to system bezpiecznej, stabilnej archiwizacji heterogenicznych zasobów cyfrowych, przydatny i użytkowany przez rozmaite instytucje oraz organizacje przechowujące i udostępniające zasoby cyfrowe; system zdolny do integracji z funkcjonującymi już w instytucjach archiwizujących systemami informatycznymi. Planowano stworzyć system skalowalny i zachowujący swą przydatność pomimo zmian technologicznych. Już we wstępnej fazie prac projektowych nad systemem założono permanentny rozwój i dostosowywanie do zmieniających się okoliczności. Podstawowym założeniem było dostosowanie niemieckiego systemu depozytowego do rekomendowanych norm i standardów międzynarodowych. W ogólnym ujęciu projekt Kopal miał dostarczyć podstawy dla organizacyjnego oraz technicznego rozwiązania problemu długoterminowej archiwizacji cyfrowych zasobów [Kopal, 2007; Kopal, b.d.].

Zespół niemiecki, opracowując system depozytowy dla rodzimych zasobów cyfrowych, postanowił wzorować się na doświadczeniach holenderskich, pochodzących przede wszystkim z projektu NEDLIB realizowanego w Koninklijke Bibliotheek w latach 1998-2000 [NEDLIB publications, b.d.; Steenbakkers, 2000]. Jak już wspomniano, w ramach projektu NEDLIB oraz kontynuowanej po zakończeniu projektu współpracy bibliotekarzy KB z informatykami firmy IBM opracowany został system archiwizacji cyfrowych zasobów Holandii e-Depot. Techniczne rozwiązanie, które jest jądrem systemu e-Depot, nazywa się DIAS (Digital Information and Archiving System) i jest autorskim produktem firmy IBM, opracowanym na podstawie modelu referencyjnego OAIS specjalnie dla potrzeb archiwizacyjnych KB [e-Depot, b.d.].

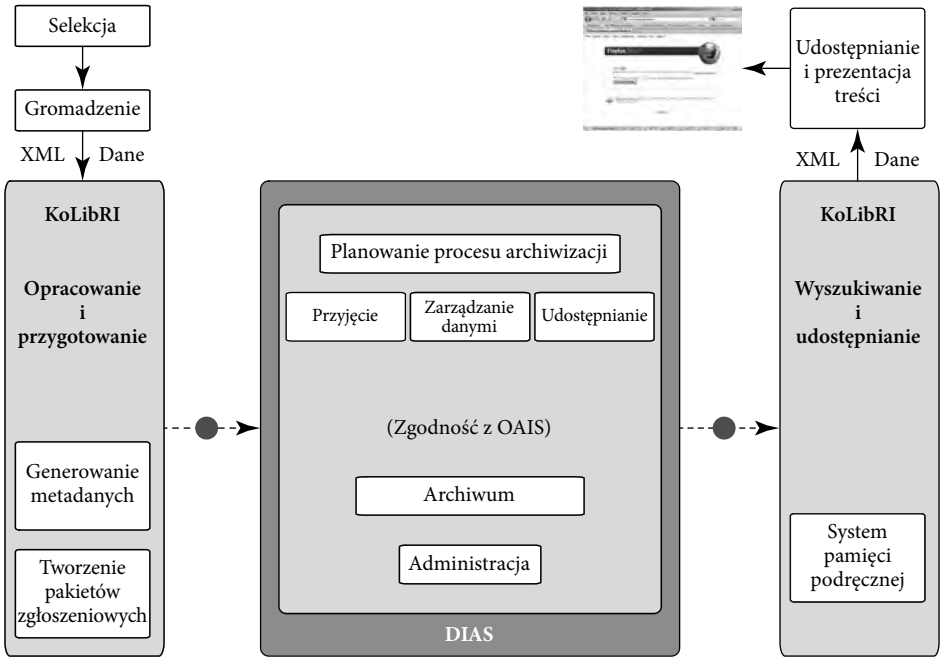
W Niemczech, po dokładnej analizie i ocenie stopnia przydatności, rozwiązanie DIAS zostało przystosowane i wdrożone jako DIAS-Core do systemu depozytowego Kopal Solution, nazywanego też Kopal-Archivsystem. Główne części składowe Kopal Solution to właśnie DIAS-Core oraz kopal Tools [Kopal, 2007; Kopal, b.d.].

DIAS-Core, podobnie jak w przypadku holenderskiego systemu depozytowego eDepot, stanowi główny element systemu Kopal Solution i daje techniczne podstawy dla długoterminowego utrzymania użyteczności archiwizowanych zasobów cyfrowych. Jest udostępniany przez firmę IBM na podstawie umowy

licencyjnej. Natomiast Kopal Tools, określane jako Kopal Library for Retrival and Ingest – KoLibRI to oprogramowanie typu open source stworzone wspólnie przez partnerów projektu Kopal, zgodne z założeniami i współpracujące z oprogramowaniem DIAS-Core. Oprogramowanie KoLibRI może być modyfikowane w celu dostosowania do indywidualnych potrzeb systemów poszczególnych instytucji. Jest odpowiedzialne za takie czynności w obrębie systemu depozytowego jak: przygotowanie pakietu archiwalnego do umieszczenia i administrowania nim w archiwum, stworzenie metadanych, przekazanie pakietu do archiwum, następnie wyszukanie i udostępnienie użytkownikom publikacji w czytelnej formie.

Współdziałanie rozwiązania DIAS z oprogramowaniem KoLibRI w Kopal-Archivsystem zostało zobrazowane na rysunku 7.

Rysunek 7. **Kopal-Archivsystem**

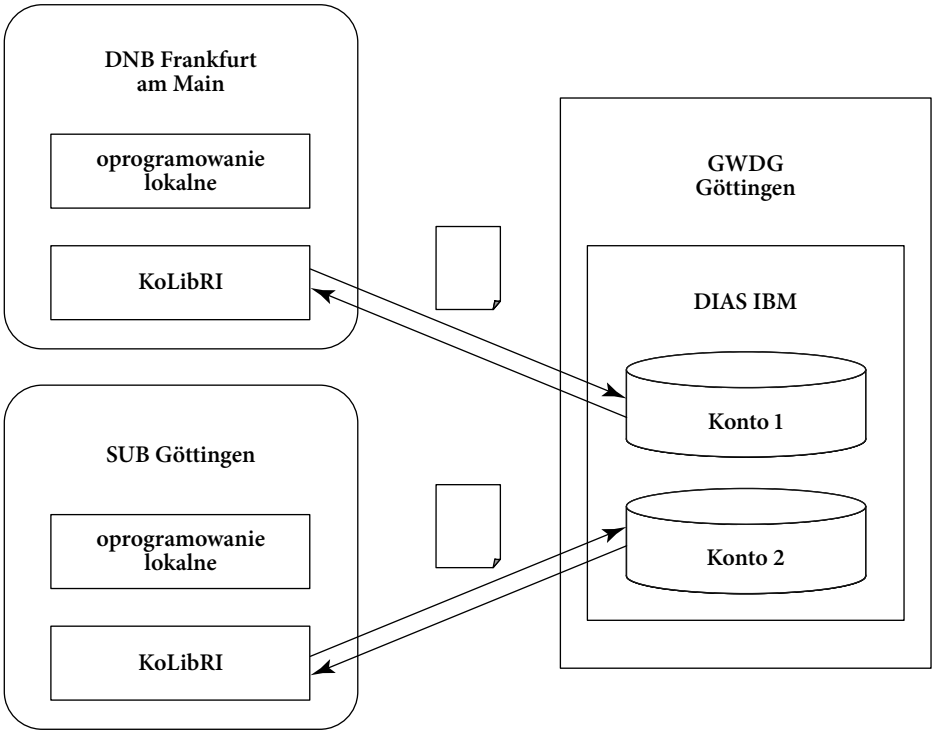


Źródło: oprac. własne na podstawie [Altenhöner, 2007]

W projekcie Kopal przyjęto, że niemieckie zasoby cyfrowe zostaną zgromadzone i przechowywane w jednym centralnym archiwum. Zdecydowano, że jądro archiwum, czyli DIAS-Core wraz ze wszystkimi zasobami będzie posadowione w Getyndze i zarządzane przez jednego z partnerów projektu – firmę informatyczną. Za pośrednictwem oprogramowania KoLibRI instytucje oddające w depozyt

swoje zasoby cyfrowe (DNB oraz SUB Göttingen) mogą je na bieżąco wyszukiwać, pobierać i udostępniać użytkownikom (por. Rys. 8).

***Rysunek 8*. Organizacja niemieckiego systemu depozytowego opracowanego w ramach projektu Kopal**



Źródło: oprac. własne na podstawie [Steinke, 2007]

Pierwsze niemieckie dokumenty cyfrowe przekształcono do postaci pakietów archiwalnych i umieszczono w systemie Kopal Solution w lipcu 2006 r. Były to: kolekcja elektronicznych rozpraw doktorskich, czasopisma sieciowe z repozytorium SpringerLink, monografie, treści publikacji z płyt CD-ROM, dokumenty multimedialne, kolekcje digitalizatów, dokumenty audio; łącznie 600 TB danych cyfrowych. Do archiwum zostały przekazane dokumenty w różnych formatach zapisu danych cyfrowych (GIFF, HTML, JPEG, PDF, TIFF, XML, UTF8, WAFE i inne).

W systemach depozytowych wykorzystujących rozwiązanie DIAS nie ma ostatecznie zdefiniowanej strategii długoterminowej archiwizacji. Stosowana taktyka polega na umiejscowieniu archiwum (jądra systemu z danymi) u doświadczonego partnera bądź renomowanego usługodawcy, zmianie pierwotnego medium zapisu

danych i ochronie treści publikacji cyfrowych w systemach depozytowych, archiwizacji równoległej w terytorialnie odległych miejscach, tworzeniu kopii zapasowych, sum kontrolnych oraz szczegółowych metadanych technicznych. Ważnym elementem tej taktyki jest również zarządzanie procesem migracji (ang. *migration management*).

W ramach tworzenia strategii archiwizacji DIAS opracowywana jest koncepcja *Preservation Planning*, czyli rozpoznawania obiektów cyfrowych, które pod wpływem zmian technologicznych nie będą w przyszłości użyteczne i planowania odpowiednich czynności konserwatorskich. Niezbędnym warunkiem dla *Preservation Planning* jest tzw. *Bitstream Preservation*, czyli bezpieczne i bezstratne utrzymywanie substancji obiektów cyfrowych (kodu binarnego).

Z systemu depozytowego stworzonego w ramach projektu Kopal korzystają głównie DNB oraz SUB Göttingen. Została jednak podjęta akcja promocyjna w celu rozszerzenia grupy użytkowników, a tym samym kooperantów testujących i przyczyniających się do rozwoju zarówno rozwiązania DIAS, jak i sprzężonego z nim oprogramowania KoLibRI. Współpraca ma umożliwić szybszy rozwój rozwiązania, zwiększenie siły przebicia w procesach standaryzacyjnych, obniżenie kosztów prac badawczych przez ich podział na więcej korzystających instytucji partnerskich [Kopal, 2007]. Rozwiązanie wzbudziło powszechne zainteresowanie wśród niemieckich instytucji dziedzictwa kulturowego, jednak złożoność i koszt systemu odstrasza wielu potencjalnych użytkowników.

LuKII – LOCKSS und KOPAL Infrastruktur und Interoperabilität

LuKII to projekt realizowany w kooperacji Uniwersytetu Humboldta w Berlinie oraz Deutsche Nationalbibliothek. Łączy rozwiązania proponowane przez usługi LOCKSS [Filas i Wiorogórska, 2010, s. 33-43; LOCKSS Program, b.d.] oraz Kopal [DFG, 2014; Hein i in., 2011, s. 51-53]. Celem projektu jest połączenie mocnych stron obu rozwiązań w postaci usługi LuKII. Opracowywane w Niemczech rozwiązanie ma bazować na korzystnym finansowo rozproszonym modelu trwałej ochrony integralności substancji (tj. kodu zero-jedynkowego) archiwizowanych obiektów (co jest domeną LOCKSS), ze zdolnością utrzymania w długim czasie użyteczności archiwizowanych obiektów poprzez metodę migracji, co z kolei jest atutem systemu Kopal. Do projektu przyłączyły się również inne niemieckie instytucje, udostępniając do podejmowanych testów zasoby cyfrowe (głównie kolekcje repozytoriów instytucjonalnych). W ramach projektu zaplanowano opracowanie rozwiązania o nazwie CLOCKSS (Controlled LOCKSS). Na tej podstawie została powołana usługa (moduł) kontrolowanego dostępu do systemu deponowania zasobów cyfrowych, proponowanego w ramach LOCKSS, w którym są składowane

i trwale archiwizowane naukowe zasoby bibliotek i wydawnictw całego świata. CLOCKSS (Controlled LOCKSS) to wspólne przedsięwzięcie, typu non profit, wydawców akademickich i bibliotek naukowych, których celem jest zbudowanie trwałego, rozproszonego geograficznie, ciemnego archiwum (ang. *dark archiv*), zapewniającego długotrwałe przetrwanie internetowych publikacji naukowych, głównie czasopism. Kolekcje cyfrowe niedostępne już u wydawcy stanowią tzw. zawartość wyzwoloną (triggered content) i dzięki usłudze CLOCKSS są dostępne bezpłatnie i trwale, jako dzieła osierocone na licencji Creative Commons [CLOCKSS Archive, 2017].

1.4.5. Wiarygodność archiwów cyfrowych

W środowisku instytucji pamięci mocno akcentowane są starania, aby wszelkie funkcjonujące i planowane archiwa cyfrowe, zwłaszcza te przechowujące cyfrowe dziedzictwo nauki i kultury, zasłużyły na miano instytucji wiarygodnych, autentycznych, stabilnych i niezawodnych (ang. *trusted digital repository, trustworthy digital repositories,* niem. *vertrauenswürdiges digitales Langzeitarchiv*) [Center for Research Libraries i OCLC, 2007; Kriterienkatalog, 2006; Research Libraries Group i OCLC, 2002]. Wiarygodne archiwum cyfrowe to takie, które gwarantuje dostępność przechowywanych i zarządzanych w nim dokumentów obecnie i w odległej przyszłości, przyjmuje odpowiedzialność za przeprowadzanie stosownych prac konserwatorskich w imieniu swoich deponentów oraz na rzecz potrzeb obecnych i przyszłych użytkowników. Projekt takiego archiwum powinien uwzględnić powszechnie przyjęte umowy i standardy w zakresie zapewnienia ciągłości zarządzania bezpieczeństwem i dostępności deponowanego w nim materiału cyfrowego. Ważne jest wypracowanie metod oceniania systemu pod względem spełniania oczekiwań użytkowników dotyczących jego wiarygodności. W celu odróżniania wiarygodnych archiwów od tych, które nie gwarantują niezawodności w zakresie utrzymania dostępu do zdeponowanego materiału cyfrowego, niezbędne jest zidentyfikowanie atrybutów systemu, które są w stanie przekonać użytkowników o jego zdolności do długoterminowego i stabilnego zarządzania dokumentami cyfrowymi. Zdaniem specjalistów działania oraz wydajność i efektywność systemów archiwizacji powinny podlegać okresowym pomiarom i kontroli, a zatem być sprawdzalne i mierzalne [Attributes, 2001; Research Libraries Group i OCLC, 2002]. Dodatkowo archiwum cyfrowe powinno posiadać sprecyzowaną politykę sporządzania i kontroli kopii zapasowych archiwizowanych obiektów, dysponować narzędziami wykrywania i odzysku utraconych lub uszkodzonych dokumentów, dbać o kontakty z wydawcami, zapewniać jawność informacji o tym, co i w jaki sposób jest archiwizowane. W opracowaniach przedmiotu zwraca się też uwagę

na równowagę kosztów i korzyści – innymi słowy, należy starannie rozważyć, czy cyfrowy obiekt jest wart kosztów ponoszonych w procesie jego archiwizacji. Atrybuty archiwum cyfrowego zakwalifikowano do czterech grup: odpowiedzialność administracyjna, wykonalność organizacyjna, równowaga ekonomiczna, podstawy proceduralne [Attributes, 2001; Research Libraries Group i OCLC, 2002].

Wieloletnie wysiłki instytucji pamięci wielu krajów świata, w szczególności prace dwóch grup roboczych (1) RLG-NARA Digital Repository Certification Task Force oraz (2) Nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland przyniosły propozycje katalogów kryteriów identyfikacyjnych dla wiarygodnych, stabilnych, długoterminowych archiwów cyfrowych, na których podstawie możliwe stało się opracowanie modelu audytu i certyfikacji archiwów cyfrowych [Center for Research Libraries i OCLC, 2007; Januszko-Szakiel, 2009b, s. 325-347, Januszko-Szakiel i in., 2016; Kriterienkatalog, 2006]. Zaproponowane katalogi grup RLG-NARA oraz Nestor mają charakter vademecum. Zawierają wskazówki dla prac koncepcyjnych, planowania oraz organizacji, a w późniejszej fazie także kontroli funkcjonowania wiarygodnych długoterminowych archiwów cyfrowych. Są pomyślane jako podręcznik dla wszelkich instytucji, które podejmują zadania archiwizacji zasobów cyfrowych, oraz dla komercyjnych usługodawców w zakresie archiwistyki cyfrowej. W pracach nad charakterystyką wiarygodnych archiwów cyfrowych za punkt wyjściowy obie grupy przyjęły referencyjny model OAIS. Posługują się standaryzowanym nazewnictwem procesów zachodzących w archiwach cyfrowych, akceptują normatywny sposób organizacji i funkcjonowania obiektów cyfrowych w archiwach. Katalogi kryteriów zostały uznane w 2012 r. przez International Organization for Standardization za normę ewaluacji i certyfikacji długoterminowych wiarygodnych kolekcji cyfrowych (ISO 16363, 2012).

W obu katalogach, oprócz kryteriów oceny wiarygodności archiwów cyfrowych, wymieniono cztery zasady ich wprowadzania do archiwów [Center for Research Libraries i OCLC, 2007; Kriterienkatalog, 2006]:

- Dokumentacja: wszelkie pomysły, plany, wdrożenia, zastosowane normy jakościowe i normy bezpieczeństwa dotyczące archiwum cyfrowego, powinny być skrupulatnie dokumentowane i poddawane ocenie wewnętrznej oraz zewnętrznej. Cenne mogą okazać się wszelkie spostrzeżenia i wskazówki, zarówno ze strony pracowników instytucji inicjującej działania, jak i opinii publicznej. Możliwie wczesna ocena jest środkiem zapobiegawczym przed popełnieniem ewentualnych błędów. Ponadto staranne, konsekwentne dokumentowanie wszystkich obszarów i etapów prac umożliwia kompleksową ocenę spójności i poprawności pracy archiwum.
- Przejrzystość: zarówno działania związane z tworzeniem archiwum cyfrowego, jak i jego późniejsze funkcjonowanie powinny być transparentne wewnątrz oraz

na zewnątrz organizacji. W celu osiągnięcia przejrzystości zewnętrznej zaleca się podawanie do ogólnej wiadomości stanu zaawansowania prac. Permanentne publikowanie tworzonych raportów i dokumentacji, tym samym informowanie deponentów oraz użytkowników o wszystkim, co dotyczy archiwum, umożliwia im samodzielną ocenę stopnia jego wiarygodności. Twórcom[5] dodatkowo pozwala przekonać się czy archiwum jest w stanie zapewnić ich produktom stosowną ochronę. Przejrzystość wewnętrzna natomiast jest konieczna dla osób czynnie zaangażowanych w prace nad tworzeniem i funkcjonowaniem archiwum. Poprzez bieżące wzajemne informowanie o wykonywanych i planowanych czynnościach we wszystkich obszarach działań, menedżerowie oraz wykonawcy zadań mają możliwość stwierdzenia, czy archiwum osiąga zakładany stan rozwoju i jakości. Ponadto umożliwia osobom skupionym nad poszczególnymi fragmentami prac i zadań archiwum, objąć je i zrozumieć ich znaczenie w działaniu jednego systemu. W przypadku fragmentów dokumentacji, zawierających informacje tajne, kieruje się je tylko do wiadomości osób upoważnionych. Mówi się wówczas o przejrzystości zawężonej.

- Adekwatność: przy tworzeniu archiwów cyfrowych zaleca się przyjęcie zasady adekwatności, w myśl której mało realne, a nawet niemożliwe jest stworzenie rozwiązań absolutnie standardowych. Wszelkie proponowane metody, sposoby bądź formy postępowania na rzecz ochrony zasobów cyfrowych wymagają każdorazowego oszacowania ich przydatności w realizacji zadań określonej instytucji archiwizującej. Rozwiązania z powodzeniem sprawdzające się w jednym archiwum mogą tylko częściowo, albo w ogóle nie znaleźć zastosowania w innym. Należy uwzględnić obowiązujące przepisy prawne, zasoby personalne oraz finansowe, także istniejące już struktury organizacyjne oraz stopień zaawansowania prac na rzecz opracowania instytucjonalnej, lokalnej albo narodowej strategii długoterminowej archiwizacji zasobów cyfrowych.
- Ewaluacja: w przypadku oceny wiarygodności archiwum cyfrowego – postrzeganej szczególnie jako jego zdolność do ochrony długoterminowej – nie wymieniono dotychczas atrybutów dających się obiektywnie i jednoznacznie oszacować, zmierzyć. Wciąż trwają prace nad utworzeniem ostatecznego wykazu wskaźników reprezentujących wiarygodność archiwum. Wszelkie propozycje wskaźników powinny zostać podane do powszechnej wiadomości, nie

[5] Pod pojęciem *twórcy* rozumie się wszystkie osoby (autora, informatyka, wydawcę, dostawcę), które mają swój udział w opublikowaniu i dostarczeniu do archiwum dokumentu cyfrowego w jego ostatecznej treści i formie. W modelu OAIS pojawia się jeszcze pojęcie *producent.* Określa się nim dostawcę, najczęściej wydawcę publikacji elektronicznych, który zgłasza i przesyła dokument do archiwum.

tylko, by uzyskać komentarze i opinie o nich, ale również, by były elementem działań na rzecz zapewnienia przejrzystości archiwum.

Wiarygodność archiwów cyfrowych – zdaniem przedstawicieli grupy Nestor – nie powinna być postrzegana jako pojęcie absolutne, lecz odnoszące się każdorazowo do indywidualnych założeń, zadań i celów poszczególnych archiwów. Każde archiwum powinno opublikować swoje cele i – w myśl zasady adekwatności – wybrać spośród istniejących rozwiązań te, które umożliwią ich realizację. Proces ewaluacji wiarygodności archiwum powinien polegać na obserwacji transparentnych poczynań archiwum i opiniowaniu przez obserwatorów (głównie deponentów i użytkowników), w jaki sposób radzi sobie ono z realizacją wytyczonych celów [Kriterienkatalog, 2006]. Grupa Nestor zaznacza, że wiarygodność archiwów cyfrowych jest ściśle powiązana ze stabilnością techniczną zapewnianą przez specjalistów branży IT [Kriterienkatalog, 2006].

Kryteria identyfikacji wiarygodnego długoterminowego archiwum cyfrowego dają się zgrupować w trzech klasach [Center for Research Libraries i OCLC, 2007; Kriterienkatalog, 2006]:

A – ramy organizacyjne

Archiwum cyfrowe działa w organizacyjnych ramach wynikających ze zdefiniowanych celów archiwum, uwarunkowań prawnych, a także zasobów kadrowych i finansowych.

1. Definiowanie celu działalności archiwum cyfrowego

Archiwum cyfrowe określa swoje założenia, obowiązki, zadania do wykonania oraz zasady, którymi kieruje się w ich wykonywaniu. Cele archiwum cyfrowego są transparentne, publikowane w formie tzw. policy[6].

1.1. Opracowanie kryteriów gromadzenia obiektów cyfrowych

Archiwum cyfrowe możliwie jednoznacznie wskazuje cechy obiektów cyfrowych podlegających długoterminowej ochronie. Niekiedy przyjęcie do kolekcji archiwalnej obiektu cyfrowego może być podyktowane odrębnymi przepisami

[6] Zagadnienie *Preservation Policy* jest szerzej omawiane w rozdziale drugim książki. Najogólniej pod tym pojęciem należy rozumieć zbiór dokumentów (o charakterze ustaw, postanowień, umów, rozporządzeń, wytycznych) regulujących procesy długoterminowej archiwizacji zasobów cyfrowych. W odróżnieniu od strategii archiwizacji, która określa sposób ochrony materiału cyfrowego, *preservation policy* wskazuje, m.in.: co, gdzie, dlaczego i jak długo powinno podlegać ochronie. *Preservation Policy* jest niezbędną podstawą strategii archiwizacji. Zob.: [Neuroth i in., 2009].

prawnymi, którym instytucja archiwizująca powinna podporządkować się w swej działalności. Oprócz procedur selekcji i oceny, archiwum cyfrowe określa zasady przekazania obiektu cyfrowego do archiwum.

1.2. Przyjęcie odpowiedzialności za długoterminową ochronę obiektów cyfrowych

Archiwum cyfrowe oświadcza, iż przyjmuje odpowiedzialność za długoterminowe zabezpieczenie dostępności oraz użyteczności zasobów cyfrowych, zgromadzonych na podstawie ustaleń wynikających z punktu 1.1.

1.3. Definiowanie grupy użytkowników docelowych archiwum cyfrowego

Archiwum cyfrowe określa grupę/y użytkowników, dla której/ych potrzeb działa, rozpoznaje specyfikę ich oczekiwań i na tej podstawie dobiera odpowiednie narzędzia i metody pracy. Ponadto archiwum przyjmuje obowiązek stałego monitorowania wymogów użytkowników i dostosowywania do nich swych usług.

2. Tworzenie możliwości użytkowania archiwalnych zasobów cyfrowych

Archiwum cyfrowe przyjmuje za podstawowe zadanie stworzenie swym obecnym i przyszłym klientom możliwości użytkowania, czyli odczytu i interpretacji treści reprezentowanych przez chronione obiekty archiwalne. Zakres użytkowania może być różny, w zależności od ewentualnych obostrzeń prawnych lub niepowodzenia w zachowaniu pewnych atrybutów oryginału. W przypadkach tego typu niepowodzeń czy też poważniejszych zagrożeń utraty zasobów archiwalnych wykorzystywane są zasoby tzw. ciemnych archiwów (nieoficjalnych, drugiego obiegu), których założeniem jest ochrona zasobów, jednak bez możliwości ich tradycyjnego użytkowania. Są pomyślane jako forma zastępstwa dla archiwów cyfrowych w sytuacjach szczególnie kryzysowych.

2.1. Organizacja dostępu użytkowników do archiwalnych zasobów cyfrowych

Archiwum cyfrowe gwarantuje uprawnionym użytkownikom dostęp do obiektów cyfrowych, stwarzając przy tym stosowne narzędzia ich wyszukiwania. Ustala jasne zasady organizacyjne korzystania z zasobów oraz informuje o ewentualnych kosztach, np.: wydruku, zapisu na nośniku, wysłania pocztą elektroniczną.

2.2. Zapewnienie użytkownikom możliwości interpretacji treści cyfrowych obiektów

Archiwum cyfrowe podejmuje działania na rzecz zagwarantowania użytkownikom możliwości odczytu i interpretacji dokumentów cyfrowych – zarówno ich treści, jak i metadanych. W tym celu wymagany jest szereg zabiegów natury technicznej, m.in. okresowa kontrola odczytu i interpretacji obiektów. Archiwa

stosują również tzw. formularz zwrotny, dzięki któremu użytkownicy mogą zgłaszać ewentualne trudności odczytu i interpretacji.

3. Respektowanie przepisów prawnych i umownych

Archiwa cyfrowe działają na bazie określonych regulacji ustawowych oraz umownych. Dotyczą one w szczególności sposobu pozyskiwania zasobów archiwalnych, ich ochrony oraz udostępniania. Archiwa starają się uwzględniać zarówno interesy twórców publikacji, jak i potrzeby użytkowników.

3.1. Prawne uregulowanie współpracy archiwum cyfrowego z twórcami publikacji

W celu działania planowego oraz zgodnego z prawem archiwa cyfrowe zawierają formalne porozumienia z wydawcami publikacji, w których precyzowane są warunki procesów: przekazanie publikacji do archiwum, archiwizacja, użytkowanie. Pewne obowiązki oraz zadania zarówno archiwów, jak i wydawców mogą wynikać z obowiązujących aktów prawnych; dodatkowe porozumienia i umowy określają sposób realizacji tychże zadań.

3.2. Respektowanie przepisów prawnych dotyczących procesów długoterminowej ochrony obiektów cyfrowych

Archiwa cyfrowe regulują stosownymi zapisami procesy związane z archiwizacją obiektów cyfrowych, np. prawo dostępu w celu przeprowadzania określonych prac konserwatorskich na obiektach. Ponadto przestrzegania wymagają zapisy ustawy o prawie autorskim związane z ewentualnymi zmianami treści i formy dokumentu.

3.3. Respektowanie przepisów prawnych dotyczących procesów użytkowania zasobów cyfrowych

Archiwa cyfrowe dbają o to, aby użytkowanie deponowanych zasobów cyfrowych odbywało się z poszanowaniem przepisów prawnych. Przestrzegania wymagają przede wszystkim ustawy o prawie autorskim oraz o ochronie danych, ale również przepisy regulujące okres przechowywania dokumentów w archiwach. Wszelkie ograniczenia i bariery uniemożliwiające użytkowanie zasobów powinny być dokumentowane wraz z uzasadnieniem ich podstaw.

4. Dostosowanie formy organizacyjnej archiwum cyfrowego do realizowanych w nim celów

W zależności od założeń archiwa cyfrowe mogą działać i zapewniać ochronę krótko-, średnio-, lub długoterminową zdeponowanych zasobów cyfrowych. Wydajność oraz trwałość archiwum cyfrowego to parametry podlegające ocenie deponentów oraz użytkowników i wpływające na jego wiarygodność. Podstawami tej oceny są:

4.1. Zabezpieczenie finansowania działalności archiwum cyfrowego

Archiwum cyfrowe zapewnia swych klientów o finansowym zabezpieczeniu swojej działalności. Zarówno archiwa państwowe, jak i prywatne – w szczególności te długoterminowe – powinny wskazać prawną podstawę oraz źródła finansowania. Archiwa państwowe są zazwyczaj finansowane z budżetu państwa i to państwo jest gwarantem ich finansowej stabilności. Natomiast archiwa prywatne świadczą usługi odpłatnie, ich finansowa kondycja jest wynikiem powodzenia na rynku, wewnętrznej gospodarki finansowej, finansowego planowania. Aktualna sytuacja i polityka finansowa archiwum cyfrowego przekłada się na ocenę jego wiarygodności.

4.2. Dyspozycyjność personelu o odpowiednich kwalifikacjach

Archiwum cyfrowe zatrudnia kadrę z odpowiednimi kwalifikacjami. Dla zapewnienia długoterminowej skuteczności konieczny jest stały rozwój pracowników – dostosowywanie kwalifikacji do zmieniających się okoliczności funkcjonowania archiwum, tak aby wszystkie czynności związane z funkcjonowaniem archiwum były wykonywane zgodnie z założeniami ilościowymi, jakościowymi oraz terminowymi. W zasadzie wszystkie archiwa cyfrowe, w szczególności jednak te z założeniem funkcjonowania długoterminowego, muszą w swych planach organizacyjnych i finansowych uwzględnić procesy dokształcania kadry. Zapewnione powinny być czas i pieniądze na udział personelu w specjalistycznych kursach, szkoleniach, krajowych oraz międzynarodowych konferencjach. Należy także uwzględnić potrzebę dostępu do fachowej literatury. To właśnie odbyte przez pracowników kursy, zdobyte umiejętności wpływają na ocenę wiarygodności archiwum.

4.3. Powoływanie stosownych struktur organizacyjnych

Struktura organizacyjna archiwum cyfrowego powinna być ściśle dostosowana do jego założeń, realizowanych celów, zadań. Procesom zachodzącym w archiwach należy przyporządkować adekwatne zasoby personalne oraz materialne, umożliwiające realizację założonych celów.

4.4. Sporządzanie planów długoterminowych

Archiwa cyfrowe sporządzają plany działania, w których uwzględniane są wszelkie zadania do wykonania obecnie i w przyszłości, wraz z określeniem terminów. Dla zapewnienia ich długoterminowego działania prowadzą też tzw. zapobiegawcze planowanie strategiczne, polegające na stałej obserwacji pewnych zjawisk, przewidywaniu ewentualnych zmian i wytyczaniu w związku z nimi nowych zadań. Monitorują głównie zmiany technologiczne (w modelu OAIS

określane jako Monitor Technology) oraz zmiany oczekiwań i potrzeb użytkowników (za OAIS Monitor Designated Community). Mogą zmieniać się również podstawy prawne oraz finansowe działania archiwów cyfrowych. Elementem planowania jest zabezpieczenie potrzebnych zasobów.

4.5. Kontynuacja ochrony zasobów archiwalnych w sytuacjach kryzysowych

Archiwa cyfrowe, w szczególności te długoterminowe, opracowują strategię postępowania i zapewnienia ciągłości ochrony zasobów w sytuacjach kryzysowych. Przy założeniu ewentualnej potrzeby przekazania zasobów archiwalnych do instytucji partnerskiej bądź następczej, archiwum cyfrowe odpowiednio wcześniej planuje proces przekazania swoich obowiązków, definiuje jego warunki i przygotowuje potrzebną infrastrukturę. Konieczny jest przede wszystkim staranny dobór instytucji partnerskiej oraz zawiązanie umowy o partnerstwie, na mocy której w razie konieczności instytucja ta obejmie zagrożone zasoby stosowną ochroną. W takich sytuacjach szczególne znaczenie ma staranna dokumentacja dotycząca wszystkich zasobów kolekcji wraz z metadanymi. Dokumentacja stanowi dla archiwum przejmującego obowiązki ochrony podstawowe źródło wiedzy o ilościowym i jakościowym stanie zasobów oraz przyjętej strategii ich archiwizacji.

5. Zarządzanie jakością w archiwum cyfrowym

Oddział zarządzania jakością kontroluje realizację wszystkich procesów i zadań składających się na osiągnięcie celów archiwum cyfrowego. Oddział obejmuje kontrolą procesy zachodzące we wszystkich obszarach działalności archiwum.

5.1. Podział zadań i obowiązków w ramach realizowanych procesów

Oddział zarządzania jakością dba o przyporządkowanie wszystkim procesom i zadaniom realizowanym w archiwum cyfrowym odpowiednich zasobów kadrowych i materialnych. Szczególnie starannie definiuje odpowiedzialność za realizację procesów i zadań współzależnych (przy wzajemnym oddziaływaniu wielu osób bądź zespołów na efekt końcowy). Równie istotna jest odpowiedzialność za procesy zewnętrzne, realizowane poza archiwum, jednak wpływające na przebieg procesów wewnętrznych (np. tworzenie i dostarczanie obiektów cyfrowych do archiwum).

5.2. Zarządzanie dokumentacją archiwum cyfrowego

Oddział zarządzania jakością dba o sprawne działanie systemu zarządzania dokumentacją, dotyczącą wszystkich elementów składowych archiwum. Sprawuje kontrolę nad przestrzeganiem reguł dotyczących kompletności, poprawności, aktualności, zrozumiałości oraz dostępności dokumentacji. Dokumentacja archiwum powstaje wg precyzyjnych wytycznych.

5.3. Reagowanie archiwum cyfrowego na zmiany

Oddział zarządzania jakością nadzoruje procesy monitoringu zmian głównie natury technicznej (np. standardy formatów zapisu i nośników danych cyfrowych), ale również organizacyjnej (np. sposób finansowania działań archiwum, przekazanie odpowiedzialności instytucji partnerskiej lub następczej), a także natury społecznej (np. postaw i oczekiwań użytkowników archiwum). Opóźniona reakcja na zmiany może wywołać poważne utrudnienia w realizacji celów archiwum, dlatego system zarządzania jakością dba, aby zmiany możliwie wcześnie rozpoznać, przewidzieć ich wpływ na realizację zadań archiwum, następnie zaplanować, wprowadzić i skontrolować właściwe działania aktualizacyjne.

B – schemat postępowania z obiektami cyfrowymi

Wszelkie zabiegi na obiektach cyfrowych – głównie natury technicznej – odnoszą się do zachowania autentyczności, integralności, wiarygodności oraz dostępności zarówno obiektów, jak i metadanych.

6. Zabezpieczenie integralności obiektów cyfrowych

W celu zapewnienia integralności obiektów cyfrowych (rozumianej głównie jako kompletność obiektu cyfrowego) oraz wykluczenia wszelkich niezamierzonych modyfikacji na nim, archiwum podejmuje właściwe działania natury organizacyjnej oraz technicznej. Odpowiednio wczesna reakcja na przewidywalne zmiany umożliwia rozpoznanie oraz korektę nieprawidłowości.

6.1. Zabezpieczenie integralności obiektów cyfrowych „na wejściu" do archiwum cyfrowego (za OAIS: Ingest)

Archiwum cyfrowe ustala z twórcami, głównie wydawcami oraz dostawcami, jakimi cechami muszą charakteryzować się obiekty cyfrowe, aby archiwum przejęło odpowiedzialność za dalszą ochronę ich integralności. Archiwum określa również techniczne wymagania dostarczenia publikacji. „Na wejściu" do archiwum obiekt cyfrowy poddawany jest przede wszystkim kontroli integralności; sprawdzane są także inne parametry jakościowe.

6.2. Zabezpieczenie integralności obiektów cyfrowych w procesie archiwizacji (za OAIS: Archival Storage)

Archiwum cyfrowe chroni integralność obiektów cyfrowych poprzez rozmaite zabiegi, przede wszystkim ustala jakość mediów stosowanych do zapisu danych cyfrowych (wybiera nośniki certyfikowane i spełniające określone normy jakościowe).

Archiwum ustala możliwie jednoznaczną politykę dostępu do obiektów cyfrowych przez pracowników archiwum (np. administratora systemu – w celu przeprowadzania prac konserwatorskich).

Archiwum kieruje się zrozumiałymi zasadami określania stopnia fizycznej redundancji. Precyzyjnie określa właściwą lokację archiwizowanych obiektów cyfrowych oraz przynależnych podsystemów.

6.3. Zabezpieczenie integralności obiektów cyfrowych w procesie użytkowania (za OAIS: Access)

Archiwum cyfrowe definiuje jasne zasady użytkowania obiektów cyfrowych. Chroni obiekty, ich metadane, a także inne elementy systemu przed jakimkolwiek działaniem nieupoważnionych użytkowników. Uprawnionym użytkownikom daje możliwość skontrolowania integralności obiektów cyfrowych.

Archiwum wyznacza granice swojej odpowiedzialności za integralność obiektów w procesie ich udostępnienia użytkownikom.

7. Zabezpieczenie autentyczności obiektów cyfrowych

Archiwum cyfrowe musi ochronić autentyczność obiektu cyfrowego, pojmowaną jako możliwość potwierdzenia autorstwa oraz prawdziwości treści w nim zawartych. Obiekt cyfrowy jest autentyczny wówczas, gdy przedstawia dokładnie to, co autor zamierzał za jego pośrednictwem przedstawić. Archiwum cyfrowe zabezpiecza autentyczność obiektów cyfrowych na etapie przyjęcia, przechowywania oraz udostępniania. Starannie dokumentuje przypadki, w których stwierdzono wątpliwość autentyczności obiektu oraz takie, w których autentyczność ewidentnie nie potwierdza się.

7.1. Zabezpieczenie autentyczności obiektów cyfrowych „na wejściu" do archiwum

Archiwum cyfrowe wymaga od firm wydawniczych oraz dostawczych, z którymi współpracuje, formalnego potwierdzenia rejestracji swojej działalności (przez autoryzowaną instytucję). „Na wejściu" archiwum cyfrowe wymaga od twórców potwierdzenia autentyczności obiektu, np. na podstawie metadanych dotyczących pochodzenia obiektu. Obiekty autentyczne mogą być oznaczane cyfrową sygnaturą.

7.2. Zabezpieczenie autentyczności obiektów cyfrowych w procesie archiwizacji

Archiwum cyfrowe tworzy pełny wykaz starannie opisanych przypadków manipulacji, w których wyniku doszło do zmian bądź usunięcia zarówno samego obiektu, jak i metadanych.

7.3. Zabezpieczenie autentyczności obiektów cyfrowych w procesie użytkowania

Archiwum cyfrowe powinno potwierdzić swoją autentyczność przed użytkownikami; dysponować i w razie potrzeby oddawać do wglądu dokumenty, z których wynika, że archiwum prowadzi zarejestrowaną, autoryzowaną działalność. Archiwum cyfrowe w procesie udostępniania stosuje sygnatury cyfrowe, toteż ważne jest udokumentowanie ich pochodzenia i zasad stosowania.

W celu ewentualnego oszacowania przez użytkowników autentyczności obiektów archiwum udostępnia metadane, w których zawarty jest opis pochodzenia obiektu oraz dokumentacja wszelkich zmian powstałych w wyniku procesu archiwizacji. Użytkownik może również zapoznać się z wykazem obiektów cyfrowych, do których archiwum ma wątpliwości bądź nie potwierdza autentyczności.

8. Długoterminowe planowanie technicznych procesów archiwizacji

Archiwum cyfrowe opracowuje długoterminowe plany, w których zawarte są wszelkie obecne i przyszłe zadania oraz terminy ich wykonania. Szczególne znaczenie ma strategiczne planowanie długoterminowe stricte dotyczące zadań natury technicznej (patrz p. 4.4), np.: zmiana nośników, konwersja do aktualnych formatów, przegląd integralności, autentyczności, kontrola dostępności, odczytu i prezentacji danych. Zadania techniczne odnoszą się zarówno do obiektów cyfrowych, jak i metadanych.

9. Określenie procedur gromadzenia obiektów cyfrowych

Archiwum cyfrowe opracowuje procedury dotyczące gromadzenia obiektów cyfrowych. W tym celu ustala wytyczne selekcji i oceny oraz dostarczenia obiektów do archiwum. Dopuszcza się zarówno manualny, jak i zautomatyzowany tryb dostarczenia obiektów do archiwum.

9.1. Opracowanie specyfikacji dotyczącej obiektów cyfrowych przekazywanych do archiwum (za OAIS: Submission Information Packages, SIPs)

Archiwum cyfrowe ustala z twórcami (głównie wydawcami i dostawcami), które parametry są konieczne, aby obiekt cyfrowy został przekazany do archiwum. Dzięki tym ustaleniom możliwa jest automatyzacja procesu dostarczania obiektów do archiwum oraz implementacja tzw. workflow, czyli sekwencji procedur przyjęcia i wdrożenia obiektu do zasobu archiwalnego. Specyfikacja jest podstawą kontroli jakości obiektów cyfrowych przekazywanych do archiwum.

9.2. Identyfikacja szczególnie znaczących i wartych zachowania cech obiektów cyfrowych

W niektórych przypadkach archiwum cyfrowe podejmuje decyzję o tym, które z cech obiektów cyfrowych zasługują na szczególną ochronę. Czyniąc to, powinno się

brać pod uwagę konkretne cele archiwum cyfrowego bądź misję, którą musi pełnić, możliwości techniczne oraz ponoszone nakłady, a wreszcie potrzeby użytkowników.

Niekiedy, w celu zachowania możliwie wielu cech obiektów cyfrowych, zachodzi konieczność ochrony jednego obiektu w kilku wariantach.

9.3. Przejęcie kontroli technicznej nad obiektami cyfrowymi

Zdarza się, że do archiwum cyfrowego trafiają obiekty wyposażone w mechanizm ograniczający ich użytkowanie (z racji obostrzeń prawnych lub komercyjnych interesów twórców). Archiwum dba jednak, aby przed włączeniem obiektów do zasobów archiwalnych usunąć wszelkie elementy ich wyposażenia, które mogłyby w jakikolwiek sposób blokować, ograniczać lub utrudniać realizację procesów ich długoterminowej ochrony.

10. Definiowanie i przestrzeganie procedur archiwizacji obiektów cyfrowych

Istota archiwów cyfrowych tkwi w realizacji procesów archiwizacyjnych. Najważniejsze z nich to: zdefiniowanie obiektu archiwalnego, jego zapis oraz wykonywanie zabiegów konserwatorskich.

10.1. Definiowanie obiektów archiwalnych (za OAIS: Archival Information Packages, AIPs)

Na obiekt archiwalny składają się dane reprezentujące zawartość (treść) dokumentu, zapisane w określonym formacie, oraz metadane, istotne dla procesów długoterminowej archiwizacji tegoż dokumentu, wpisane w zdefiniowaną strukturę.

Definiowanie obiektów archiwalnych obejmuje identyfikację zastosowanych do obiektów struktur, formatów i dostępnych metadanych (patrz p. 12). Skuteczność procesu długoterminowej archiwizacji zależy w dużej mierze od zastosowanych formatów. Archiwa cyfrowe zalecają tzw. uniwersalne otwarte formaty (UOF).

10.2. Transformacja gromadzonych obiektów cyfrowych do postaci obiektów archiwalnych

Archiwum cyfrowe przekształca gromadzone obiekty cyfrowe do postaci obiektów archiwalnych oraz dołącza do nich metadane zawierające informacje istotne dla realizacji procesów ich długoterminowej ochrony.

10.3. Zabezpieczenie zapisu i odczytu obiektów archiwalnych

Archiwum cyfrowe z pomocą narzędzi, którymi dysponuje, zapewnia odczyt obiektów archiwalnych, rozumiany jako możliwość odczytu mediów cyfrowych oraz zapisanych w nich kodów zero-jedynkowych.

10.4. Stosowanie strategii długoterminowej ochrony obiektów archiwalnych

Z prowadzonych przez archiwum cyfrowe planów działania (patrz p. 8) wiadomo, jakim zabiegom i w jakim czasie powinny zostać poddane obiekty archiwalne. Dla każdego obiektu archiwum wyznacza termin kontroli, w wyniku której podejmowane są decyzje o poddaniu go stosownym zabiegom konserwatorskim.

11. Ustalenie wytycznych użytkowania zasobów archiwalnych

Archiwum cyfrowe udostępnia zasoby archiwalne na podstawie zdefiniowanych zasad użytkowania. Ustalone są zasady przeszukiwania zasobów, dostępu do nich oraz zakres użytkowania.

11.1. Definiowanie obiektów użytkowych (za OAIS: Dissemination Information Packages, DIPs)

W zależności od zgłaszanych potrzeb użytkowników oraz parametrów obiektów archiwalnych archiwum cyfrowe definiuje obiekty użytkowe. Wyznacza otoczenie dla procesów wyszukiwania oraz użytkowania obiektów. W zależności od kontekstu użytkowania archiwum może udostępnić obiekt cyfrowy w rozmaitych postaciach użytkowych. Należy jednak mieć świadomość, że użytkowanie obiektu cyfrowego nie oznacza dostępu do chronionego obiektu archiwalnego, lecz do jego kopii lub derywatu wraz z wszelkimi informacjami potrzebnymi w procesie użytkowania obiektu.

Dopuszcza się również wymianę obiektów pomiędzy archiwami. Konieczna jest wówczas transformacja obiektu do standaryzowanego formatu eksportowego.

11.2. Transformacja obiektów archiwalnych do postaci obiektów użytkowych

Obiekty użytkowe powstają – przy zastosowaniu określonych metod – na bazie obiektów archiwalnych. Mogą być przechowywane w archiwum i udostępniane w przypadku nowej kwerendy, z uwzględnieniem ewentualnego dostosowania obiektu do nowego kontekstu użytkowania. Mogą również mieć charakter jednorazowy, co oznacza, że za każdym razem, gdy zgłaszane jest zapotrzebowanie na obiekt użytkowy, będzie on na bieżąco tworzony z obiektu archiwalnego.

12. Zarządzanie danymi

Oddział zarządzania danymi wspiera wszystkie istotne procesy zachodzące w archiwum cyfrowym – od gromadzenia i przyjęcia obiektów, poprzez ich archiwizację, aż do udostępniania. Chroni także ich integralność oraz autentyczność na wszystkich etapach ich przetwarzania i funkcjonowania w archiwum. W realizacji swych zadań oddział zarządzania danymi bazuje na starannie utworzonych i zapisanych metadanych zawierających:

- dane identyfikacyjne obiektów cyfrowych, stanowiące podstawę zarządzania nimi oraz ich relacjami[7];
- opis formy, treści i struktury obiektów – istotnych z punktu widzenia procesów ich wyszukiwania i użytkowania;
- opis techniczny obiektów – ważny dla zapewnienia możliwości odczytu, prezentacji i interpretacji obiektu, ochrony jego integralności oraz planowania i przeprowadzania zabiegów konserwatorskich;
- dokumentację wszelkich zauważonych zmian w obiektach – konieczną z punktu widzenia ochrony autentyczności obiektów;
- ewidencję wszelkich dokumentów o charakterze prawnym (ustaw, zarządzeń, umów, porozumień), których przestrzeganie jest konieczne w toku organizacji i realizacji działalności archiwum.

W zależności od konkretnych potrzeb archiwa mogą stosować rozmaite schematy metadanych. Sensowne jest jednak – z racji ewentualnej współpracy i potrzeby wymiany metadanych z instytucjami partnerskimi – zastosowanie formatu najbardziej rozpowszechnionego. Archiwum ustala reguły wypełniania pól metadanych (np. znormalizowaną terminologią). Możliwe jest również zastosowanie specjalnych narzędzi do automatycznego generowania (ekstrahowania) metadanych.

12.1. Trwałe identyfikowanie obiektów archiwalnych oraz ich relacji

Archiwum cyfrowe stosuje wewnętrzny system identyfikacyjny w celu efektywnego zarządzania obiektami archiwalnymi oraz ich relacjami, a także w celu jednoznacznego przyporządkowania danych treściowych do metadanych. Zastosowanie standaryzowanych identyfikatorów trwałych (unikalnych; ang. *Persistent Identifier*) zapewnia autentyczność i niezawodność w procesach bibliograficznych poszukiwań oraz cytowania treści obiektów archiwalnych.

Elektroniczne publikacje są sygnowane następującymi identyfikatorami:

- ISBN (International Standard Book Number) – międzynarodowy standard identyfikacyjny dla wydawnictw zwartych;
- ISSN (International Standard Serial Number) – międzynarodowy standard identyfikacyjny dla periodyków;
- URN (Uniform Resource Names) – międzynarodowy internetowy standard identyfikacyjny dla obiektów sieciowych;
- HDL (Handle System) – identyfikator, przypisywany na stałe do obiektów cyfrowych, niezależnie od ich fizycznego umiejscowienia i powiązany z danymi

[7] Pojęcie relacji obiektu archiwalnego to związek obiektu z innymi elementami do niego przynależnymi, np. kilka części składających się na jedną publikację elektroniczną lub kilka różnych wersji tej samej publikacji. Por: [Kriterienkatalog, 2006].

ze specjalnej bazy, na podstawie których możliwe jest uzyskanie podstawowych informacji o obiekcie;
- DOI (Digital Object Identifier) – identyfikator obiektów cyfrowych, stosowany m.in. w branży wydawniczej do oznaczania elektronicznych wersji publikacji naukowych. Techniczną podstawą dla DOI jest HDL;
- SRef (Scientific Reference linking system) – baza danych przechowująca hiperłącza do publikacji wraz z unikatowymi identyfikatorami publikacji; umożliwia wydawcom stosowanie odsyłaczy, które będą poprawnie funkcjonowały, pomimo zmian adresu publikacji, do której prowadzi odsyłacz.

12.2. Tworzenie metadanych opisujących treść i formę oraz umożliwiających identyfikację obiektów cyfrowych

Archiwum cyfrowe tworzy metadane, które (stosownie do jego potrzeb) opisują treść i formę obiektu oraz umożliwiają jego identyfikację. Zakres, struktura oraz treść opisowych metadanych zależne są od celów archiwum, potrzeb użytkowników oraz od samych obiektów cyfrowych. Formalny i treściowy opis obiektów umożliwia ich wyszukiwanie. Obecnie w instytucjach bibliotecznych oraz archiwalnych stosuje się rozmaite formaty metadanych. W bibliotekach najpopularniejszy jest Dublin Core (DC).

12.3. Tworzenie metadanych opisujących strukturę obiektów cyfrowych

Archiwum cyfrowe tworzy metadane strukturalne, dzięki którym możliwe jest przedstawienie kompleksowej struktury obiektu cyfrowego, a następnie jego zrekonstruowanie i użytkowanie jako całości. Przykładowo zdigitalizowana wersja drukowanej książki składa się m.in. z dwustu pojedynczych plików graficznych – w metadanych strukturalnych zapisuje się ich odpowiednie przyporządkowanie do właściwych miejsc na stronach książki. Podobną rolę pełnią metadane strukturalne w przypadku archiwizacji stron internetowych. Strony WWW składają się zazwyczaj z większej liczby stron HTML oraz plików graficznych (np. w formacie JPEG), powiązanych ze sobą linkami. W metadanych strukturalnych znajduje się dokładny opis tychże powiązań.

12.4. Tworzenie metadanych rejestrujących zmiany w obiektach cyfrowych

Archiwum cyfrowe dokumentuje w metadanych wszelkie zmiany, które zachodzą w obiektach cyfrowych, zarówno w wyniku zamierzonych, jak i niepożądanych działań. Dokumentowanie zmian jest zabiegiem koniecznym, nie tylko dla udowodnienia autentyczności obiektu archiwalnego, ale i do realizacji technicznych prac na obiektach. Metadane dokumentujące zmiany pełnią szczególnie istotną rolę w tych archiwach, które zdecydowały się na migrację jako strategię

długoterminowej archiwizacji. Migracja wiąże się z regularnymi zmianami obiektów cyfrowych. W tego typu metadanych rejestruje się też zmiany wynikające z procesów transformacji obiektów cyfrowych do postaci obiektów przekazywanych do archiwum, postaci obiektów archiwalnych oraz postaci obiektów użytkowych.

12.5. Tworzenie metadanych opisujących techniczne parametry obiektów cyfrowych

Archiwa cyfrowe sporządzają tzw. metadane techniczne dla obiektów cyfrowych, w których szczegółowo opisują techniczne parametry samych obiektów oraz wszystkich plików przynależących do poszczególnych obiektów kompleksowych. Dzięki tym metadanym możliwe jest zabezpieczenie integralności obiektów oraz zarządzanie pracami konserwatorskimi na obiektach.

12.6. Tworzenie metadanych opisujących zasady użytkowania obiektów cyfrowych

Archiwum cyfrowe tworzy metadane obiektów cyfrowych, w których starannie opisuje zasady użytkowania obiektów archiwalnych. Szczególnie pieczołowicie tworzy się takie metadane dla obiektów, których użytkowanie jest z jakichś powodów ograniczone. Metadane „użytkowe" służą zatem powiadomieniu użytkowników o prawach i warunkach użytkowania obiektów cyfrowych, zarządzaniu procesem użytkowania (np. kontrolą dostępu do obiektu) oraz zarządzaniu prawami autorskimi do obiektów.

12.7. Przyporządkowanie metadanych do obiektów cyfrowych

Archiwum cyfrowe powinno opracować system jednoznacznego oraz stabilnego wiązania metadanych zarówno z obiektami cyfrowymi w całości, jak i z poszczególnymi częściami składającymi się na ten obiekt. Archiwa mogą to osiągnąć, np. poprzez zastosowanie trwałych identyfikatorów do obiektów cyfrowych oraz ich poszczególnych części albo poprzez przechowywanie obiektu cyfrowego wraz z przynależącymi do niego częściami i metadanymi w jednej, dokładnie zdefiniowanej strukturze (za OAIS: SIP, AIP, DIP) oraz tym samym miejscu (tzw. hermetyzowanie, kapsułowanie).

C – infrastruktura i bezpieczeństwo

W archiwum cyfrowym istotne znaczenie ma techniczna sprawność oraz bezpieczeństwo systemu. W tym celu archiwum wdraża odpowiednią infrastrukturę z zakresu technologii informacyjnych (IT).

13. Zapewnienie stosownej infrastruktury z zakresu IT

Infrastruktura IT jest konieczna dla realizacji zadań związanych z zapewnieniem wszystkim obiektom archiwalnym bezpieczeństwa i stabilności w technicz-

nym znaczeniu. Archiwum implementuje infrastrukturę dostosowaną do indywidualnych, często bardzo specyficznych potrzeb.

13.1. Zastosowanie infrastruktury IT w pracach na obiektach cyfrowych

Archiwum cyfrowe implementuje infrastrukturę umożliwiającą realizację technicznych zadań związanych z ochroną obiektów cyfrowych na wszystkich etapach ich funkcjonowania w archiwum (w procesach przyjęcia, wdrożenia do kolekcji archiwalnej i użytkowania) oraz wspierającą procesy zarządzania danymi.

13.2. Zastosowanie infrastruktury IT w realizacji założeń bezpieczeństwa systemu

Archiwum cyfrowe przy wyborze i implementacji infrastruktury IT uwzględnia wymagania dotyczące bezpieczeństwa całości systemu. Możliwe są, np. techniczne zabezpieczenia w postaci haseł albo biometrycznych barier dostępu.

14. Zastosowanie infrastruktury dla ochrony archiwum cyfrowego i jego kolekcji

W archiwum cyfrowym potrzebna jest infrastruktura chroniąca obiekty archiwalne przed zagrożeniami wewnętrznymi oraz zewnętrznymi. Zagrożenia wewnętrzne najczęściej wynikają z błędów systemowych; należą do nich m.in.: niedomagania sprzętu albo uszkodzenia nośników danych cyfrowych. Natomiast zagrożenia zewnętrzne są utożsamiane przede wszystkim z naturalnymi zagrożeniami – głównie pożarem, powodzią, trzęsieniem ziemi itp., ale także z niepożądaną działalnością człowieka. Archiwum cyfrowe powinno przede wszystkim radzić sobie z odpieraniem zagrożeń obiektów archiwalnych, ale także chronić wszelkie współtworzące system zasoby materialne i ludzkie. Ponadto archiwum cyfrowe preferuje przetrzymywanie kopii bezpieczeństwa kolekcji archiwalnej w miejscu fizycznie oddalonym od głównej siedziby archiwum.

1.4.6. Audyt wiarygodnych archiwów cyfrowych

Grupa RLG-NARA zaznacza, że samo wyszczególnienie kryteriów dla audytu i certyfikacji wiarygodnych archiwów cyfrowych to za mało. Pełna użyteczność tych procesów będzie zależeć od narzędzi, które umożliwią wgląd i obiektywną (wewnętrzną oraz zewnętrzną) kontrolę procesów zachodzących w archiwum. Standaryzowana metoda kontroli i certyfikacji wiarygodności archiwów cyfrowych jest absolutnie niezbędnym elementem długoterminowej archiwizacji zasobów cyfrowych [RLG, 2005].

Obecnie narzędziem audytu wiarygodności archiwów cyfrowych są katalogi kryteriów zaproponowane przez grupy RLG-NARA i Nestor. Ocena wiarygodności archiwum cyfrowego powstaje na podstawie kontroli i adnotacji faktu występowania określonego kryterium w następujących fazach: koncepcyjnej, plano-

wania i dokumentacji, implementacji, ewaluacji. Uzupełnieniem obu katalogów kryteriów jest tabela (zaznacza się w niej występowanie poszczególnych kryteriów archiwum cyfrowego w tych czterech fazach). Uzupełniona została o rubrykę, w której odnotowuje się fakt opublikowania dokumentacji z procesu wdrożenia do archiwum określonego kryterium. Ten zabieg przyczynia się do budowania pozytywnego wizerunku i wiarygodności archiwum (patrz Tab. 1).

Warto zaznaczyć, że w tabeli występowanie kryteriów w poszczególnych fazach zapisane jest w formie czasu przeszłego dokonanego, zatem adnotacja świadczy o przejściu określonego kryterium przez kolejną z czterech, a właściwie pięciu faz.

Tabela 1. Kryteria oceny wiarygodności archiwów cyfrowych

		Zaplanowano	Udokumentowano	Zaimplementowano	Oceniono	Opublikowano
A	**RAMY ORGANIZACYJNE**					
1.	Definiowanie celu działalności archiwum cyfrowego					
1.1	Opracowanie kryteriów gromadzenia obiektów cyfrowych					
1.2	Przyjęcie odpowiedzialności za długoterminową archiwizację obiektów cyfrowych					
1.3	Definiowanie grup(y) użytkowników docelowych archiwum cyfrowego					
2.	Tworzenie możliwości użytkowania archiwalnych zasobów cyfrowych					
2.1	Organizacja dostępu użytkowników do archiwalnych zasobów cyfrowych					
2.2	Zapewnienie użytkownikom możliwości interpretacji treści zawartych w archiwalnych obiektach cyfrowych					
3.	Respektowanie przez archiwum cyfrowe przepisów prawnych i umownych					
3.1	Prawne uregulowanie współpracy archiwum z twórcami obiektów cyfrowych					
3.2	Respektowanie przepisów prawnych dotyczących procesu długoterminowej archiwizacji obiektów					
3.3	Respektowanie przepisów prawnych dotyczących procesu użytkowania archiwalnych zasobów					
4.	Dostosowanie formy organizacyjnej archiwum cyfrowego do realizowanych w nim celów					
4.1	Zabezpieczenie finansowania działalności archiwum cyfrowego					
4.2	Dyspozycyjność personelu o odpowiednich kwalifikacjach					
4.3	Powołanie stosownych struktur organizacyjnych					
4.4	Sporządzanie planów długoterminowych					
4.5	Kontynuacja ochrony zasobów w sytuacjach kryzysowych					
5.	Zarządzanie jakością w archiwum cyfrowym					
5.1	Podział zadań i obowiązków w ramach realizowanych procesów					
5.2	Sporządzanie i zarządzanie dokumentacją dotyczącą wszelkich elementów składowych archiwum cyfrowego					
5.3	Reagowanie archiwum cyfrowego na zmiany					

		Zaplanowano	Udokumentowano	Zaimplementowano	Oceniono	Opublikowano
B	**SCHEMAT POSTĘPOWANIA Z OBIEKTAMI CYFROWYMI**					
6.	Zabezpieczenie integralności obiektów cyfrowych					
6.1	Zabezpieczenie integralności obiektów cyfrowych „na wejściu" do archiwum					
6.2	Zabezpieczenie obiektów cyfrowych w procesie archiwizacji					
6.3	Zabezpieczenie obiektów cyfrowych w procesie użytkowania					
7.	Zabezpieczenie autentyczności obiektów cyfrowych					
7.1	Zabezpieczenie autentyczności obiektów cyfrowych „na wejściu" do archiwum					
7.2	Zabezpieczenie autentyczności obiektów cyfrowych w procesie archiwizacji					
7.3	Zabezpieczenie autentyczności obiektów cyfrowych w procesie użytkowania					
8.	Długoterminowe planowanie technicznych procesów archiwizacji zasobów archiwalnych					
9.	Określenie procedur gromadzenia obiektów cyfrowych					
9.1	Opracowanie specyfikacji obiektów cyfrowych przekazywanych do archiwum					
9.2	Identyfikacja szczególnie istotnych i wartych zachowania cech obiektów cyfrowych					
9.3	Przejęcie technicznej kontroli nad obiektami cyfrowymi					
10.	Definiowanie i przestrzeganie procedur archiwizacji obiektów cyfrowych					
10.1	Definiowanie obiektów archiwalnych					
10.2	Transformacja zgromadzonych obiektów cyfrowych do postaci obiektów archiwalnych					
10.3	Zapewnienie zapisu i czytelności obiektów archiwalnych					
10.4	Stosowanie strategii długoterminowej archiwizacji obiektów archiwalnych					
11.	Ustalenie wytycznych dla procesu użytkowania obiektów archiwalnych					
11.1	Definiowanie obiektów użytkowych					
11.2	Transformacja obiektów archiwalnych do postaci obiektów użytkowych					
12.	Zarządzanie danymi					
12.1	Trwałe identyfikowanie obiektów archiwalnych oraz ich relacji					
12.2	Tworzenie metadanych opisujących treść i formę oraz umożliwiających identyfikację obiektów cyfrowych					
12.3	Tworzenie metadanych opisujących strukturę obiektów cyfrowych					
12.4	Tworzenie metadanych rejestrujących zmiany zachodzące w obiektach cyfrowych					
12.5	Tworzenie metadanych opisujących techniczne parametry obiektów cyfrowych					
12.6	Tworzenie metadanych opisujących zasady użytkowania obiektów cyfrowych					
12.7	Przyporządkowanie metadanych do obiektów cyfrowych					
C	**INFRASTRUKTURA I BEZPIECZEŃSTWO**					
13.	Zapewnienie stosownej infrastruktury z zakresu IT					
13.1	Zastosowanie stosownej infrastruktury IT w procesach archiwizacji obiektów cyfrowych					
13.2	Zastosowanie infrastruktury IT w realizacji założeń bezpieczeństwa systemu archiwalnego					
14.	Zastosowanie infrastruktury w celu ochrony archiwum cyfrowego i jego zasobów					

Źródło: oprac. własne na podstawie [Center for Research Libraries i OCLC, 2007; Kriterienkatalog, 2006]

Zaproponowane katalogi są niewątpliwie znaczącym uzupełnieniem wiedzy o organizacji i funkcjonowaniu archiwów cyfrowych. Dostarczają systematyzacji dotychczasowych rozpoznań z tego zakresu oraz stanowią punkt wyjścia dla prac nad procesami oceny i certyfikacji archiwów cyfrowych. Nie są ani kompletne, ani szczegółowe, charakteryzuje je dość wysoki poziom abstrakcji wymienionych kryteriów, gdyż w zamiarze mają odnosić się do szerokiego spektrum archiwów cyfrowych. Wykazy szczegółowe i kompletne musiałyby dotyczyć wyłącznie konkretnych archiwów cyfrowych.

Autorzy katalogów podkreślają, że bardzo trudne, a wręcz niemożliwe jest spełnienie przez archiwa cyfrowe wszystkich kryteriów w stu procentach. Zakres wdrożenia i spełniania kryteriów jest zależny od indywidualnych założeń poszczególnych archiwów. Proces oceny wiarygodności należy przede wszystkim rozpocząć od rozpoznania celów archiwum oraz przyporządkowanych im zadań i na tej podstawie można dopiero szacować zasadność oraz zakres występowania kryteriów.

Pierwsze próby zastosowania normy ewaluacji i certyfikacji długoterminowych wiarygodnych kolekcji cyfrowych (ISO 16363: 2012) zostały poczynione. Audytem objęto archiwum CLOOCKS (Controlled LOCKSS). Od września 2013 do maja 2014 r. trwała rygorystyczna diagnoza wiarygodności instytucji i organizacji repozytorium zgodnie z deklarowanymi założeniami archiwum i oczekiwaniami użytkowników. W efekcie audytu poświadczono, że CLOCKSS jest wiarygodnym repozytorium zasobów cyfrowych. Członkowie panelu eksperckiego ds. certyfikacji The Center for Research Libraries (CRL) stwierdzili, że praktyki i usługi opisane w komunikacie CLOCKSS i opublikowanej dokumentacji jako zgodne z normą, faktycznie takie są.

W 2016 r. podjęto, prawdopodobnie pierwszą w naszym kraju, próbę ewaluacji polskich bibliotek cyfrowych w kontekście katalogów kryteriów wiarygodności kolekcji cyfrowych. Na przykładzie Małopolskiej Biblioteki Cyfrowej oraz Jagiellońskiej Biblioteki Cyfrowej starano się określić czy i w jakim stopniu polskie instytucje tworzące kolekcje materiałów cyfrowych są przygotowane do spełnienia wymogów instytucji wiarygodnych, gwarantujących osiągnięcie celu długotrwałego zabezpieczenia użyteczności przechowywanych zasobów cyfrowych. Wyniki diagnozy obu instytucji są odmienne z uwagi na znaczne różnice uwarunkowań ich organizacji i funkcjonowania, w znaczeniu czasu ich powoływania, dostępnej wówczas wiedzy i doświadczeń, dostępnych zasobów infrastrukturalnych, personalnych i finansowych. Należy też uwzględnić, że obie biblioteki cyfrowe były powoływane w warunkach nieświadomości ich inicjatorów odnośnie do istnienia opracowań, które mogły posłużyć jako vademecum w procesach ich organizacji i funkcjonowania. Jednak realizacja takiego przedsięwzięcia pokazuje, przed jakim spektrum celów, zadań i czynności stoją organizatorzy polskich kolekcji cyfrowych i jakie zagadnienia powinny zostać wpisane w plany ich rozwoju [Januszko-Szakiel i in., 2016, s. 189-224].

2. Techniczne, prawne i ekonomiczne zagadnienia długoterminowej archiwizacji zasobów cyfrowych

Organizacja trwałego i wiarygodnego systemu deponowania zasobów cyfrowych wymaga przede wszystkim uwzględnienia istniejących już norm i standardów oraz zaleceń, które wynikają z działań instytucji i organizacji od lat zaangażowanych w prace nad długoterminową archiwizacją dokumentów cyfrowych. Rozpoznania doświadczonych środowisk pokazują, że efektywność procesów archiwizacyjnych zależy od uwzględnienia szeregu zagadnień technicznych, prawnych oraz ekonomicznych.

2.1. Techniczne zagadnienia trwałej ochrony zasobów cyfrowych

Długoterminowa użyteczność zasobów cyfrowych wiąże się z systematycznym rozwiązywaniem wielu problemów natury technicznej. Najpoważniejsze dylematy, z którymi powinni liczyć się autorzy programów ochrony cyfrowych kolekcji, wynikają z niskiej trwałości nośników dokumentów cyfrowych. Nośniki tracą swe pierwotne właściwości po upływie kilku lub – rzadziej – kilkudziesięciu lat. Kolejny poważny problem to starzenie się środków dostępu, połączeń sprzętu i oprogramowania, które są wycofywane z rynku i zastępowane nowymi narzędziami, często niekompatybilnymi, toteż niezdolnymi do odtworzenia treści dokumentu w formie czytelnej dla użytkownika. Utrata dostępu do dokumentu może nastąpić też w wyniku zdarzeń losowych, takich jak: pożar, powódź, rozmaite awarie, zainfekowanie wirusem komputerowym, atak hakerów. Potrzebne są również rozwiązania dotyczące barier dostępu do zasobów cyfrowych, np. haseł, systemów szyfrowania. Obok zabezpieczeń technicznych niebagatelne znaczenie dla procesów ochrony mają prawidłowy zapis i opis dokumentów, umożliwiające identyfikację oraz odczyt. Utrata informacji kontekstowych (metadanych) grozi utratą czytelności i wiarygodności danych. Istotny jest także wybór formatów zapisu

treści dokumentów, bowiem zastosowanie niestandardowych formatów może wpłynąć negatywnie na efektywność długoterminowej użyteczności dokumentu.

W dalszej części książki omówione zostały najpowszechniejsze, uchodzące za standardowe, praktyki postępowania z zasobami cyfrowymi w celu zapewnienia ich dostępności, poufności, integralności i autentyczności. Zachowanie tych właściwości dokumentów cyfrowych stanowi poważną trudność. Wyzwaniem archiwistyki w świecie cyfrowym jest nie tylko zabezpieczenie nośników, ale przede wszystkim zachowanie informacji jako wartości, czyli istoty utworu [Daszewski, 2004].

2.1.1. Trwałość nośników danych cyfrowych

W opracowaniach z zakresu informatyki *nośnik danych cyfrowych* jest definiowany jako obiekt materialny, w którym lub na którym jest utrwalona informacja w postaci powierzchownych lub przestrzennych zmian właściwości obiektu, przy czym informacja jest reprezentowana za pomocą dwóch stanów mających odmienne, dające się rozróżnić właściwości mechaniczne, elektryczne, magnetyczne lub optyczne. Zgodnie z tą definicją nośnikami danych cyfrowych są m.in.: komputerowy dysk twardy, dysk optyczny (CD, DVD, Blue-ray, HD-DVD), karta magnetyczna, magnetyczna taśma cyfrowa i moduł pamięci półprzewodnikowej (ang. *flash memory*) [Bilski, 2008, s. 89]. Innymi słowy, *nośnik danych cyfrowych* bywa określany jako każde medium, które może przechować dane zapisane w postaci cyfrowej [Freedman, 2004, s. 522]. W terminologii informacji oraz języków i systemów informacyjno-wyszukiwawczych występuje pojęcie *nośnik elektroniczny*. Jest ono definiowane jako nośnik informacji, który pozwala na uzyskanie poprzez transformację utrwalonych na nim danych odpowiedniego układu elektronów będącego odwzorowaniem różnego rodzaju komunikatów, np.: tekstów różnych języków naturalnych w subkodzie akustycznym i subkodzie graficznym, innych komunikatów graficznych (jak obrazy, wykresy, animacje, filmy), komunikatów dźwiękowych (np. muzyki). Nośnikiem elektronicznym może być, zgodnie z tą definicją, taśma magnetyczna, pamięć operacyjna komputera, dysk twardy, dyskietka, płyta CD-ROM, płyta DVD, dysk MO (magnetooptyczny), moduł pamięci flash [Bojar, 2002, s. 171].

W literaturze przedmiotu dokonuje się różnorodnego podziału nośników danych cyfrowych. Rozróżnia się m.in. nośniki jednorazowego zapisu, czyli takie, na których można zapisać dane, ale ich usunięcie jest niemożliwe (np. dyski WORM i CD-R), oraz nośniki wielokrotnego zapisu, tj. takie, w przypadku których można wielokrotnie zapisać dane, usuwać je i ponownie zapisywać – dobrą egzemplifikacją są dyski i taśmy magnetyczne [Barczak i Sydoruk, 2002, s. 253-

260; Buczyński, 1999, s. 67-86; Ullrich, 2009] oraz dyski magnetooptyczne (MO) i dyski optyczne wielokrotnego zapisu (np. CD-RW, DVD-RW) [Buczyński, 1999, 67-86; Freedman, 2004, s. 522, 919]. Dysk i taśma magnetyczna oraz płyta optyczna są dodatkowo określane jako nośniki pamięci masowej [Freedman, 2004, s. 522], czyli służące do trwałego lub półtrwałego magazynowania dużych ilości danych cyfrowych [Freedman, 2004, s. 193-194, 571, 605, 819]. Ponadto zwraca się uwagę na podział nośników ze względu na ich funkcjonalność; wyróżnia się nośniki stałe (nierozerwalnie związane z urządzeniem odczytująco-zapisującym) i nośniki wymienne (służące do przenoszenia informacji między systemami) [Bilski, 2008, s. 95]. Nośniki różnią się także niezawodnością. Możliwość odczytu z nośnika stałego zależy od sprawności urządzenia oraz samego nośnika. Natomiast przy użyciu nośnika wymiennego sprawność urządzenia nie jest niezbędna do uzyskania dostępu do informacji; urządzenie można bowiem wymienić. Dodatkowo nośniki stałe rozwijają się stopniowo, ewolucyjnie, co oznacza, że poprawa parametrów w kolejnych modelach jest niewielka. Natomiast nośniki wymienne rozwijają się skokowo; kolejne generacje charakteryzują się znacznie lepszymi parametrami od poprzednich [Bilski, 2008, s. 95-96].

W celu rozpatrywania przydatności poszczególnych rodzajów nośników do długoterminowego przechowywania danych cyfrowych należy zdefiniować pojęcie *trwałość nośników danych*. Można ją określić jako zdolność do zachowywania wymaganych parametrów – w przypadku nośników danych cyfrowych wskaźnikiem najczęściej stosowanym do oceny trwałości jest liczba błędów odczytu przypadająca na konkretną ilość danych. Nietrwałość natomiast jest postrzegana zarówno jako rozpad nośników, jak i degeneracja sprzętu [Bilski, 2008, s. 427; Borawski, 2007, s. 199]. Powołując się na normy badania trwałości dysków optycznych ISO 18921 i ISO 18927, przyjęty koniec życia dysku optycznego wyznacza granica przekroczenia pięciu błędów w odczycie bitów przy 10 tys. bitów transmitowanych [Daszewski, 2006, s. 85-86].

Większość współczesnych nośników (np. płyty CD bądź karty pamięci flash) pozostaje daleko w tyle – pod względem trwałości – za drukowaną na papierze książką lub płytą winylową. Najtrwalszymi nośnikami danych są wszelkie skamieliny oraz wystrzelone w przestrzeń kosmiczną tzw. kapsuły czasu – ich trwałość oceniana jest jako „nieskończona". Do dziesięciu tysięcy lat mogą przetrwać gliniane tabliczki oraz umieszczone w ziemi „kapsuły czasu"; tysiąc lat to maksymalny „okres trwałości" książek oraz danych umieszczonych pod ziemią w ośrodkach przechowywania danych, (ang. *storage centers*). Zdecydowanie słabiej wypada większość użytkowanych współcześnie nośników cyfrowych [Cieślak, 2002]. Należy przyjąć, że dyskietki raczej nie mają szans służyć dłużej niż dziesięć lat, a trwałość dysków twardych oraz nośników ZIP nie przekracza dwóch dekad.

Nieco dłuższy okres trwałości jest szacowany dla taśm magnetycznych (min. 30 lat); lepiej pod tym względem wypada jedynie płyta optyczna, która teoretycznie powinna bezpiecznie magazynować dane nawet przez 100 lat, przy czym dotyczy to tylko najwyższej jakości płyt tłoczonych – trwałość zapisu przeciętnej jakości płyt zapisywanych samodzielnie zwykle nie przekracza 5 lat [Cieślak, 2002]. Należy zaznaczyć, że nośniki optyczne i półprzewodnikowe (głównie ze względu na swoją pojemność, łatwość zapisu/odczytu oraz cenę) wyparły z rynku zarówno dyskietki, jak i nośniki ZIP, natomiast zastosowanie taśm magnetycznych ograniczyły tylko do profesjonalnych magazynów danych [Freedman, 2004, s. 197].

Nośniki magnetyczne, szczególnie taśmy magnetyczne, znajdują dziś zastosowanie w systemach, w których większe znaczenie przykłada się do bezpieczeństwa i możliwej do przechowania ilości danych, niż do szybkości dostępu do nich. Jest to związane z budową kaset z taśmami magnetycznymi. Dostęp do obiektu zapisanego na taśmie wymaga kolejno umieszczenia taśmy w urządzeniu (czytniku), przewinięcia jej do punktu, w którym znajduje się odczytywany dokument, i dopiero wówczas odczytania danych. Proces ten może być dość długotrwały, zwłaszcza jeśli zsumuje się czasy fizycznego pobrania kasety z taśmą, jej umieszczenia w czytniku oraz czas przewijania taśmy do punktu zapisu interesującej informacji (długość pojedynczej taśmy wynosi obecnie 960 m). Najnowsza aktualnie generacja taśm magnetycznych (LTO 7) oferuje nieskompresowaną pojemność zapisu wynoszącą 6 TB, co przy zastosowaniu zautomatyzowanych bibliotek taśmowych (zdolnych do obsługi nawet tysięcy taśm) pozwala na tworzenie bardzo pojemnych magazynów danych [White Papers, 2017]. Szacuje się, że minimalny okres trwałości taśm magnetycznych wynosi ok. 30 lat, przy czym wymagane jest spełnienie szeregu warunków przechowywania i użytkowania, w tym m.in.: liczba operacji umieszczania kasety w czytniku ograniczona do 20 tys., ograniczenie liczby operacji zapisu/kasowania danych (dotyczy taśm wielokrotnego zapisu) oraz zapewnienie odpowiednich, stabilnych warunków mikroklimatycznych przechowywania i ochrony przed promieniowaniem elektromagnetycznym [MC Storage, 2017; Muszyński, 2006].

W odniesieniu do trwałości zapisu na nośnikach optycznych zaleca się uwzględnić fakt, że jakość płyt jest różna, zależna od ich budowy, jakości zarówno składników użytych do produkcji, jak i procesu wykonania oraz metody zapisu. Niezależnie od prognoz trwałości „odczytywalność" dysków optycznych należy sprawdzać w określonych odstępach czasu. Zaleca się też przechowywanie dysków w pozycji pionowej, gdyż chroni ono przed odkształceniami uniemożliwiającymi odczyt [Daszewski, 2006, s. 85-86, 91].

Płyty optyczne wypalane, w przeciwieństwie do tłoczonych, mają relatywnie krótki okres trwałości – od dwóch do pięciu lat w zależności od jakości płyty. Żywot takich płyt można jedynie nieznacznie przedłużyć, przechowując je w chłod-

nym i zaciemnionym miejscu. Problem leży w degradacji materiału – dyski optyczne powszechnie używane do wypalania, mają powierzchnię zapisu składającą się z warstwy barwnika organicznego, która może być modyfikowana przez wypalanie światłem lasera [Gozdek, 2013]. Proces degradacji może „przesuwać" (rozmywać) granice poszczególnych pól na powierzchni, aż staną się nieodczytywalne dla lasera. Większość tanich płyt sprzedawanych w sklepach dyskontowych bądź marketach ma okres trwałości nieprzekraczający dwóch lat. Niektóre, lepszej jakości płyty oferują dłuższy okres trwałości, ale nie przekracza on na ogół pięciu lat. Ponadto odróżnienie płyty wysokiej jakości od płyty niskiej jakości jest trudne, ponieważ niewielu dostawców podaje informacje na temat trwałości nośnika. Używanie do przechowywania płyt tłoczonych w miejsce samodzielnie zapisywanych nie jest efektywne ekonomicznie ze względu na bardzo wysoki koszt przygotowania matrycy (ang. *glass-master*). W związku z tym technikę tłoczenia stosuje się głównie do wielkoseryjnej replikacji, a nie produkcji pojedynczych egzemplarzy [Muszyński, 2006].

Do czynników niszczących dyski optyczne zalicza się zmiany temperatury oraz wilgotności, szczególnie o dużych wahaniach i w krótkim czasie. Skraplająca się wilgoć wybija jony z nośników i powoduje gromadzenie się ładunków elektrostatycznych, ułatwiając proces utleniania. Destrukcyjnie na materiały, z których zbudowane są płyty, działają formaldehyd, mgła solna, wibracje, środki czyszczące. Żywot płyt skraca powtarzanie zapisu na płytach wielokrotnego zapisu (RW). Trwałość i bezpieczny okres przechowywania danych na płycie zależy od rodzaju płyty, producenta, partii wyrobu i kontroli jakości. Płyty DVD są trwalsze od płyt CD – mimo że gęstość zapisu jest wyższa, to płyty DVD dają gwarancję dłuższego zachowania informacji. Według zaleceń (ISO, ANSI, NIST[1]) optymalne warunki przechowywania płyt to temperatura od 10 do 23°C i wilgotność względna (RH) 20-50% według ISO, od 15 do 20°C/RH 25-45% według ANSI, od 4 do 20°C/RH 20-50% według NIST. Ponadto należy zapewnić stabilność warunków: maksymalna zmienność +/- 0,6°C i RH +/- 3% na dobę. Warunki zalecane to temperatura 18°C/RH 30-40%, opakowania poliwęglanowe lub akrylowe, brak dostępu światła. Spełnienie tych zaleceń wymaga stosowania urządzeń regulujących warunki mikroklimatyczne w miejscu przechowywania nośników – w nieklimatyzowanym pomieszczeniu, bez technicznej stabilizacji poziomu wilgotności na przełomie zimy i wiosny, wilgotność względna (RH) może zmienić się w krótkim czasie z 30 do 90% [Daszewski, 2006, s. 85-86].

Kolejną grupą nośników danych cyfrowych są tzw. dyski twarde. Definiowane jako podstawowe nośniki pamięci masowej, stałe dyski, trwale zamocowane w napędzie. Mogą też występować w wymiennych kasetach, być wyjmowane

[1] Normy opracowane przez instytucje: The American National Standards Institute oraz The National Institute of Standards and Technology.

z komputera, stosowane jako nośnik kopii zapasowej lub do przenoszenia danych do innych komputerów [Freedman, 2004, s. 195-196]. Są to urządzenia, które łączą w jednej obudowie zarówno sam nośnik, jak i urządzenie zapisująco-odczytujące. Umieszczenie nośnika w hermetycznych warunkach próżniowych pozwala na eliminację zagrożenia mechanicznego i chemicznego uszkodzenia nośnika przez czynniki zewnętrzne, co znacząco wydłuża okres jego trwałości. Podobnie jak przedstawione wcześniej rodzaje nośników, dyski twarde również nie są pozbawione ograniczeń. Problem z tym rodzajem nośnika nie dotyczy jednak samej powierzchni zapisu, ale głównie trwałości elementów mechanicznych, a w szczególności tzw. łożyskowania dysku. Ze względu na wbudowanie mechaniki i nośnika w hermetyczną obudowę, uszkodzenie elementów mechanicznych powoduje uniemożliwienie dostępu do zgromadzonych danych – dostęp do nośnika jest możliwy jedynie przy zastosowaniu specjalistycznego wyposażenia służącego do odzyskiwania danych z uszkodzonych dysków twardych, które umożliwia m.in. przeprowadzanie operacji naprawy bądź wymiany zużytych elementów mechanicznych w bezpyłowych warunkach niskiej lub średniej próżni [Data Max, b.d.; HDD, 2014; Klein, 2017]. Profesjonalne systemy przechowywania danych zabezpiecza się przed awariami dysków twardych poprzez wykorzystanie tzw. macierzy dyskowych, czyli urządzeń składających się z co najmniej kilku dysków twardych, na których zapis wykonywany jest przez tzw. kontroler macierzy. W zależności od przyjętego standardu, możliwy jest zapis wszystkich danych równocześnie na dwóch dyskach (lub grupach dysków) w postaci kopii lustrzanej (RAID 1 – mirroring), jednak częściej stosowane są bardziej zaawansowane rozwiązania wykorzystujące równoległy z danymi zapis sum kontrolnych umożliwiających „odbudowę" macierzy w przypadku awarii pojedynczego dysku (RAID 5 i pochodne) [Baza wiedzy, 2016].

W związku z ograniczoną trwałością samodzielnie wypalanych płyt CD/DVD, do celów długoterminowej archiwizacji danych cyfrowych sugeruje się stosowanie taśm magnetycznych lub macierzy dyskowych [Muszyński, 2006].

Do podobnych wniosków prowadzi analiza rekomendacji organizacji International Association of Sound Audiovisual Archivers (IASA) dotyczących nośników dźwięku. Wynika z nich, że najpopularniejszy obecnie nośnik to płyta optyczna CD (CD-Recordable), jednak, niestety, niespełniająca stawianych przed nią wymagań, dlatego powinna być traktowana jedynie jako format przejściowy. W zaleceniach IASA zwraca się również uwagę na dwa dodatkowe problemy. Jednym z nich jest coraz krótszy żywot technologii, zatem najprawdopodobniej żaden z obecnych nośników dźwięku nie przetrwa równie długo, co płyta długogrająca bądź kaseta magnetofonowa. Dlatego też formułowany jest postulat uniezależnienia się od nośnika, prowadzący do koncepcji masowych systemów przechowywania danych w formie cyfrowej [Borawski, 2007, s. 200].

Rozpatrując jakość dysków optycznych warto uwzględnić fakt rozwoju rynku i pojawianie się produktów znacząco ulepszonych. W 2012 r. French National Laboratory of Metrology and Testing przebadało nowy nośnik DVD o nazwie GlasMasterDisc. Celem przeprowadzonych badań w warunkach przyspieszonego starzenia w ekstremalnych warunkach klimatycznych (90°C i 85% wilgotności względnej) była ocena potencjału różnych nośników DVD w obszarze trwałego przechowywania danych. Wyniki testu wskazują, że jakość danych zapisanych na GlasMasterDisc po 1000 godz. w skrajnych warunkach nie ulega znaczącemu pogorszeniu, podczas gdy inne dostępne w handlu nośniki DVD (włącznie z nośnikiem M-Disc, który posiada zapisywalną warstwę danych wykonaną z węgla szklistego w miejsce barwnika organicznego) deteriorują w czasie krótszym niż 250 godz. Na podstawie badania wysnuto wniosek o wysokiej przydatności nowego produktu w zadaniach trwałej archiwizacji [Perdereau, 2012].

Kolejny przełom mogą stanowić zapowiadane dyski szklane, kodujące treść w postaci nanostruktur zatopionych w szkle. Dane są przechowywane wewnątrz struktury dysków, w postaci „nanokratownic". Światło lasera skanuje je i zbiera informacje o ich rozmiarze, orientacji i pozycji w trzech osiach. Ze względu na aż 5 odczytywanych atrybutów każdego elementu określa się je jako dyski pięciowymiarowe lub prościej – 5D. Zastosowana metoda kodowania zapisu pozwala na uzyskanie bardzo dużej pojemności (360 TB dla struktury o rozmiarach płyty CD). Trwałość zapisu dysków 5D wynika z zastosowania szkła o wysokiej odporności. Według szacunków, zapis ma być bezpieczny nawet w temperaturze do 1000°C, a trwałość dysku ma wynieść ok 13,8 mld lat. Nowa technologia zapisu danych została opracowana przez naukowców z Uniwersytetu Swinburne [Płaza, 2016; Zhang, 2016].

Podsumowanie rozważań o nośnikach danych cyfrowych i ich trwałości może stanowić systematyzacja czynników mających wpływ na trwałość. Do wewnętrznych czynników należą: technologia wykonania, jakość materiałów użytych w procesie produkcji oraz jakość procesu produkcyjnego. Natomiast w obrębie czynników zewnętrznych wprowadzić można podział na czynniki systemowe i środowiskowe. Do systemowych zaliczone zostały: częstość wykonywania operacji dostępu do danych oraz jakość urządzeń odczytująco-zapisujących, a do środowiskowych – nadmierna temperatura i duża dynamika zmian temperatury, wysoka, względna wilgotność powietrza oraz kondensacja pary wodnej, różne formy promieniowania elektromagnetycznego (ultrafioletowe i widzialne, gamma, kosmiczne), pola magnetyczne i elektrostatyczne, zanieczyszczenia stałe i płynne, zanieczyszczenia unoszące się w powietrzu (opary wydzielane przez środki czyszczące, lakiery, farby, dym tytoniowy, spaliny) oraz oddziaływania mechaniczne [Bilski, 2008, s. 427-429].

2.1.2. Odświeżanie nośnika

W przypadku gdy strategia długoterminowej archiwizacji danych cyfrowych zakłada zachowanie danych na pierwotnie używanych nośnikach przenośnych, wówczas konieczne staje się planowe, cykliczne przeprowadzanie testów odczytu danych z nośników, w celu wczesnego wykrycia objawów degradacji nośnika [Liegmann, 2001, s. 100-105]. Podobnie jak w ocenie trwałości nośnika, najczęściej jako wskaźnik zachodzenia w nim negatywnych zmian wykorzystywana jest liczba błędów odczytu przypadająca na określoną ilość danych. Wystąpienie niewielkiej liczby błędów odczytu jest dopuszczalne i nie powoduje niekorzystnych zmian w obiekcie cyfrowym, ponieważ obok właściwych danych na nośniku zapisywane są także dodatkowe dane służące do korekcji występujących błędów. Mechanizm ten ma jednak ograniczoną skuteczność – jeżeli liczba błędów jest duża, wówczas mogą zostać przekroczone możliwości systemu korekcyjnego, a w konsekwencji powoduje to powstanie nieodwracalnych zmian w zapisanych danych cyfrowych. Zatem oprócz cyklicznego odczytu danych konieczne jest także ustalenie maksymalnej liczby wykrytych błędów, której przekroczenie będzie skutkowało podjęciem odpowiednich działań naprawczych.

Jeżeli system archiwalny zakłada, że nośniki pozostają w powszechnym użytku (co implikuje łatwy dostęp zarówno do nowych nośników, jak i urządzeń odczytujących), wówczas w przypadku stwierdzenia oznak degradacji nośnika konieczne staje się tzw. odświeżenie nośnika, a więc przekopiowanie danych na nowy nośnik tego samego typu. Tę samą operację należy przeprowadzić w przypadku osiągnięcia przez dany nośnik maksymalnego założonego wieku nośnika, rozumianego jako pesymistycznie pojmowany okres trwałości danego rodzaju nośnika. Natomiast gdy obserwacja zmian technologicznych wskazuje, iż zastosowany nośnik staje się przestarzały i wychodzi z powszechnego użytku, wtedy konieczne jest podjęcie działań mających na celu uniknięcie problemów związanych z dostępnością zarówno nowych nośników danego typu, jak i urządzeń służących do ich odczytu [Borghoff i in., 2003; Neuroth i in., 2009; Preserving, 1996].

2.1.3. Komputerowe muzea

Jedną z koncepcji długoterminowej archiwizacji zasobów cyfrowych jest tworzenie tzw. muzeów komputerowych, czyli ochrona treści dokumentów zapisanych na ich oryginalnych nośnikach wraz z platformą sprzętowo-programową potrzebną do ich odczytu [Huth, 2009]. Zachowanie oryginalnego oprogramowania i sprzętu bywa też określane mianem strategii zachowania technologii [Czermiński, 2002, s. 93; National Library of Australia, 2003, s. 140-141]. Chociaż takie rozwiązanie uchodzi za niepraktyczne i nie jest zalecane jako metoda długoterminowej archiwi-

zacji głównie z racji ogromnych przestrzeni magazynowych potrzebnych do składowania sprzętu, a także z powodu wysokich kosztów administracyjnych, ograniczeń dostępności dla użytkowników czy wreszcie problemów technicznych związanych ze starzeniem się, eksploatacją sprzętów oraz niską trwałością nośników [Fülle i Ott, 2006], to jednak w codziennej praktyce instytucje pamięci przechowują przestarzały sprzęt i prawdopodobnie pozostanie tak do czasu, kiedy będzie można wprowadzić strategie nowoczesne [Huth, 2009]. Sprzęt służący do odczytu danych cyfrowych ulega zniszczeniu, a naprawa i utrzymanie w stanie sprawności technicznej stają się coraz droższe i trudniejsze. Tworzenie replik sprzętu nie stanowi całościowego rozwiązania problemu, gdyż koszty takiej operacji są bez wątpienia wyższe, aniżeli próba utrzymania w ruchu oryginalnego urządzenia. Kolejnym mankamentem tej strategii jest problem z pozyskaniem nowych nośników danego typu. W związku z tym bardziej uzasadnioną formą zabezpieczenia jest przenoszenie danych na aktualnie używane nośniki [Digitale Erhaltungsstrategien, 2008].

2.1.4. Zmiana generacji nośnika

Kopiowanie danych cyfrowych z nośnika, który staje się przestarzały, na nośnik aktualnie i powszechnie stosowany określane jest mianem *zmiany generacji nośnika* [Borawski, 2007, s. 199; Liegmann, 2001, s. 100-105]. Stosowanie zmiany generacji nośnika powinno być uzupełnieniem metody odświeżania nośnika – w przypadku, gdy dany nośnik został zakwalifikowany jako potencjalnie zagrożony, wówczas należy sprawdzić, czy dany typ nośnika jest nadal powszechnie używany. Jeżeli tak jest, należy dokonać jedynie odświeżenia nośnika, natomiast w przypadku nośnika wychodzącego z użytku – należy zmienić generację nośnika na aktualną [Liegmann, 2001, s. 100-105].

Odrębny przypadek wymagający zmiany generacji stanowią nośniki o bardzo dużej trwałości, których stan nie kwalifikuje ich do podejmowania działań naprawczych, natomiast zmiany technologiczne powodują wypieranie z użytku takich nośników, co w konsekwencji może spowodować problemy z dostępnością urządzeń służących do odczytu zapisanych na nich danych. W momencie stwierdzenia, iż określony rodzaj nośnika staje się przestarzały, wymagane jest podjęcie działań mających na celu przeniesienie danych ze wszystkich nośników określonego typu na nośniki nowszej generacji, niezależenie od stanu technicznego poszczególnych egzemplarzy.

Stosowanie strategii polegającej na zmianie generacji nośnika pozwala na permanentne zachowanie dostępności substancji przechowywanego dokumentu cyfrowego, przy zachowaniu rozsądnego poziomu kosztów i korzystania z urządzeń, które w danym momencie są dostępne w powszechnym obrocie [Digitale Erhaltungsstrategien, 2008; Feenstra i IBM, 2000, s. 33; Liegmann, 2001, s. 100-105].

Interesującą alternatywą dla opisanej wcześniej cyklicznej zmiany generacji nośnika jest zastosowanie nośników danych o tzw. ekstremalnie długiej trwałości. Dotyczy to np. metalicznych albo ceramicznych dysków Rosetta firmy Norsam, na których dane analogowe bądź cyfrowe są przechowywane poprzez mikromechaniczne zmiany w strukturze powierzchni. Nośniki tego typu są zupełnie obojętne na wpływ otoczenia i dlatego bezterminowo trwałe, nie wymagają zabiegów konserwacyjnych. Z racji wysokich kosztów masowe zastosowanie tego rozwiązania nie jest jednak rozpatrywane [HD-Rosetta, b.d.].

2.1.5. Repozytoria danych cyfrowych

Opisywane powyżej metody odświeżania nośnika oraz zmiany jego generacji wydają się stanowić dobre rozwiązanie problemu zarówno stosunkowo niskiej trwałości nośników cyfrowych, jak i ich starzenia się, rozpatrywanego z punktu widzenia stosowanych rozwiązań technologicznych. Występowanie bardzo szybko zachodzących zmian technologicznych (związanych ze zmianami w dziedzinie stosowanych nośników danych cyfrowych) powoduje, że niemożliwym wydaje się długotrwałe zachowanie oryginalnego nośnika, implikujące przeprowadzanie kolejnych operacji zmiany generacji nośnika, zgodnie z zachodzącymi zmianami technologicznymi.

Rozpatrując problem z punktu widzenia zachowania formy dokumentu wraz z jego nośnikiem, zmiana generacji nośnika powoduje odejście od pierwotnie stosowanego rodzaju nośnika, a więc odejście w tym aspekcie od pierwotnej postaci zapisu dokumentu. Jednakże ze względu na specyfikę zapisu cyfrowego (która powoduje, iż dane zapisane w postaci cyfrowej na nośnikach muszą zostać odczytane, a następnie zinterpretowane i zaprezentowane odbiorcy przez odpowiednie urządzenie) rodzaj nośnika, na którym zostały zapisane dane cyfrowe, ma dla użytkownika drugorzędne znaczenie lub jest zupełnie nieistotny. Opierając się na tym spostrzeżeniu, zaproponowano oddzielenie substancji obiektów cyfrowych od fizycznego nośnika i umieszczenie tych obiektów w repozytoriach danych cyfrowych [Borawski, 2007, s. 200; Neuroth i in., 2009].

Repozytoria cyfrowe bazują na magazynach danych cyfrowych zarządzanych przez specjalizowany system wykonany w technologii bazodanowej. Natomiast każdy dokument cyfrowy wraz z niezbędnymi informacjami kontekstowymi, znajdujący się w systemie, traktowany jest jako pojedynczy obiekt. Repozytoria cyfrowe, w zależności od ich przeznaczenia oraz objętości przechowywanych danych, używają do składowania danych głównie macierze dyskowe oraz biblioteki taśmowe [Bilski, 2008, s. 443-461].

Zastosowanie takiego rozwiązania pozwala na eliminację zadań związanych z utrzymywaniem urządzeń obsługujących poszczególne rodzaje nośników, mo-

nitorowaniem stanu technicznego poszczególnych egzemplarzy nośników różnorodnych typów oraz podejmowaniem koniecznych działań naprawczych. Przechowywanie wszystkich dokumentów w jednym systemie repozytoryjnym wiąże się z długą listą wymagań dotyczących bezpieczeństwa systemu archiwalnego i zgromadzonych w nim danych. Zapewnienie bezpieczeństwa przechowywanych w repozytorium danych realizowane jest zgodnie z przyjętą przez daną organizację polityką bezpieczeństwa, określaną też jako system bezpieczeństwa. Bywa on definiowany jako zestaw praw, zasad i reguł opisanych w formie zaleceń i procedur określających, w jaki sposób dane powinny być zarządzane i zabezpieczane, dystrybuowane wewnątrz instytucji, pomiędzy jej jednostkami organizacyjnymi oraz jak udostępniane użytkownikom i partnerom zewnętrznym [Barczak i Sydoruk, 2002, s. 13]. Do realizacji polityki bezpieczeństwa stosuje się wiele różnorodnych środków ochrony, tj.: kontrolę dostępu do systemu informatycznego, ochronę kryptograficzną, podpis elektroniczny, sieciowe systemy zaporowe, tworzenie kopii zapasowych, ochronę przed oprogramowaniem destrukcyjnym, regulacje prawne dotyczące bezpieczeństwa danych [Januszewicz i Lewandowski, 2009, s. 8-10].

Wymagania związane z bezpieczeństwem systemu repozytoryjnego można podzielić na dwa obszary: ochrona danych przed skutkami deterioracji sprzętu i nośników używanych do ich przechowywania oraz zezwalanie na dostęp do zgromadzonych zbiorów zgodnie z uprawnieniami nadanymi poszczególnym użytkownikom.

System zabezpieczeń repozytorium cyfrowego powinien zapewnić także ochronę danych przed fizycznym dostępem do sprzętu osób nieupoważnionych oraz wpływem niekorzystnych warunków mikroklimatycznych, w tym klęsk żywiołowych.

Stosowane w repozytoriach cyfrowych macierze dyskowe oraz biblioteki taśmowe posiadają wbudowane funkcje diagnostyczne, które na bieżąco dostarczają informacji o stanie technicznym używanych nośników danych. Dzięki temu znacznie ułatwione są czynności konserwacyjne systemu, do których należy m.in. identyfikowanie urządzeń wykazujących objawy zużycia i wymiana ich na nowe.

W celu minimalizacji negatywnego wpływu warunków mikroklimatycznych panujących w pomieszczeniach, w których znajdują się nośniki danych, konieczne jest stosowanie urządzeń regulujących wilgotność i temperaturę. Niezbędne są dodatkowe rozwiązania pozwalające na szybkie i bezpieczne dla sprzętu opanowywanie ewentualnych pożarów. Ochrona przeciwpożarowa w pomieszczeniach zawierających sprzęt nieodporny na działanie wody najczęściej realizowana jest poprzez zastosowanie zestawu czujników reagujących na zadymienie oraz podwyższenie temperatury, które uruchamiają układ wpompowujący do pomieszczenia obojętny chemicznie gaz, działający dwutorowo – rozprężając się z butli, obniża

temperaturę w pomieszczeniu, oraz poprzez wyparcie z pomieszczenia tlenu powoduje zatrzymanie procesów spalania [Barczak i Sydoruk, 2002, s. 17-28].

Zastosowanie w praktyce takich zabezpieczeń nie eliminuje w stu procentach niebezpieczeństwa utraty danych, dlatego też równolegle stosowana jest odpowiednio opracowana strategia wykonywania kopii bezpieczeństwa przechowywanych danych. Poprzez kopię zapasową, powszechnie określaną zapożyczonym z języka angielskiego terminem *backup*, należy rozumieć kopię dodatkową programu, dysku lub danych, utworzoną w celu archiwizacji albo ochrony wartościowych plików przed utratą, na wypadek uszkodzenia lub zniszczenia bieżąco używanych wersji takich plików. Kopia zapasowa jest rodzajem „ubezpieczenia" [Woodcock, 2002, s. 45].

W celu zapewnienia wysokiego poziomu bezpieczeństwa zgromadzonych danych zaleca się utrzymywanie trzech niezależnych kopii zapasowych, tj.: kopii roboczej oraz jej drugiej i trzeciej kopii. Należałoby też przechowywać jedną z kopii w miejscu geograficznie oddalonym od repozytorium, co powinno pozwolić na jej przetrwanie nawet w przypadku całkowitego zniszczenia repozytorium (na skutek działań wojennych czy klęsk żywiołowych). Wykonywanie kopii zapasowych może odbywać się na dwa sposoby. Pierwszy – to cykliczne wykonywanie kopii wszystkich zgromadzonych w systemie danych. Otrzymywana jest wówczas tzw. pełna kopia zapasowa. W przypadku dużych systemów, a takimi bez wątpienia są systemy archiwalne, proces tworzenia pełnej kopii zapasowej stanowi duże obciążenie systemu informatycznego, jest także czaso- i kosztochłonny ze względu na konieczność wykonania kopii całego zbioru danych. Dlatego też tworzenie pełnej kopii zapasowej nie może być wykonywane zbyt często, aby nie powodować przeciążenia systemu. Wadą tej strategii jest ryzyko utraty danych zmodyfikowanych lub dodanych do systemu w okresie pomiędzy momentami wykonywania kopii. W celu minimalizacji tego ryzyka stosowany jest drugi ze sposobów, tj. strategia wykonywania tzw. kopii przyrostowych. Polega on na wykonywaniu dodatkowych kopii zapasowych obejmujących jedynie dane zmodyfikowane lub dodane do systemu od momentu wykonania ostatniej kopii zapasowej. Zastosowanie tej strategii pozwala na tworzenie kopii przyrostowych w krótszych odstępach czasu, ze względu na ich niewielką objętość w stosunku do pełnej kopii danych, co powoduje mniejsze obciążenie systemu. Metoda ta wymaga także okresowego tworzenia pełnych kopii danych, jednakże umożliwia wydłużenie okresów pomiędzy ich wykonywaniem przy jednoczesnym zachowaniu wysokiego poziomu bezpieczeństwa zgromadzonych danych. Wykonywanie przyrostowych kopii zapasowych wiąże się jednak z bardziej pracochłonną procedurą odtwarzania danych w przypadku awarii, ponieważ w celu odzyskania pełnego stanu danych konieczne jest przywrócenie danych z ostatniej pełnej kopii zapasowej, a następnie odtwarzanie danych ze wszystkich kolejno utworzonych kopii przyrostowych [Coy, 2006].

Kolejnym zagadnieniem związanym z bezpieczeństwem danych zgromadzonych w repozytorium jest zarządzanie dostępem do obiektów poprzez system uprawnień poszczególnych użytkowników. System zabezpieczeń musi zapewniać jednoznaczną identyfikację i weryfikację użytkownika repozytorium, a także umożliwić dostęp do zgromadzonych danych w zakresie przewidzianym przez wcześniej zdefiniowane uprawnienia danego użytkownika. W odniesieniu do użytkowników posiadających uprawnienia modyfikacji zgromadzonych zbiorów powinien zostać zapewniony odpowiedni poziom zabezpieczenia transmisji pomiędzy użytkownikiem a systemem, by zmniejszyć ryzyko nieautoryzowanej modyfikacji, a w skrajnych przypadkach – usunięcia danych. Ponadto każda operacja, której wynikiem jest modyfikacja obiektu w systemie, powinna być szczegółowo udokumentowana wraz ze wskazaniem danych użytkownika dokonującego modyfikacji [Januszewicz i Lewandowski, 2009, s. 51-68].

2.1.6. Migracja jako metoda długoterminowej archiwizacji danych cyfrowych

Jedną z proponowanych metod długoterminowego utrzymania użyteczności dokumentów cyfrowych jest *migracja*, definiowana jako konwersja dokumentu z oryginalnego formatu do nowszego, gdy format oryginalny staje się przestarzały i wychodzi z użycia [Borghoff i in., 2003, s. 37-55; Funk, 2008c]. Migracja bywa też określana jako „wyciągnięcie" danych cyfrowych z oryginalnego, starzejącego się formatu zapisu, w celu ich konwersji do formatu aktualnie stosowanego [Fülle i Ott, 2006, s. 12].

W procesie migracji elektroniczny obiekt powinien zostać tak zmodyfikowany przez działania zewnętrzne, aby mógł być używany w zmienionym otoczeniu systemowym bez utraty danych treściowych i strukturalnych. Mankamentem migracji danych do nowych warunków systemowych jest wykluczenie użycia dokumentu w „pierwotnych" warunkach systemowych. Konwersja danych może z dużym prawdopodobieństwem powodować powstanie przekłamań i odstępstw od oryginalnej wersji dokumentu. Przedmiotem odstępstw może być sposób prezentacji dokumentu (wygląd), interaktywne zachowanie, a nawet treść. Ponadto jeśli kolejnej konwersji dokonuje się na podstawie rezultatu poprzedniej konwersji, to ryzyko przekłamań wzrasta, zwłaszcza, że najczęściej nie ma już dostępu do pierwotnego, oryginalnego obiektu cyfrowego [Funk, 2008c].

Migracja do nowszego formatu zapisu powoduje więc powstanie dokumentu, który nie jest identyczny z dokumentem pierwotnym. I choć utrata danych niekoniecznie oznacza utratę bądź zmianę zawartości treściowej, sam plik jednak ulega zmianie [Fülle i Ott, 2006, s. 12; Smith, 2003, s. 109]. Ryzyko wystąpienia odstępstw

można zminimalizować, a zmiany poprawić, pod warunkiem dostępu do dokładnej dokumentacji dotyczącej konwertowanego dokumentu oraz dokładnej znajomości specyfiki zarówno bieżącego formatu, w którym zapisane są dane, jak i formatu, do którego dane te będą konwertowane [Fülle i Ott, 2006, s. 12].

Główną zaletą migracji jest stosunkowa łatwość udostępniania publikacji w formatach powszechnie stosowanych i ogólnie dostępnych. Wadę natomiast stanowią praco- i czasochłonność związane z regularnym konwertowaniem wszystkich danych. Im większe zasoby cyfrowe, tym proces bardziej kosztowny. Warte przemyślenia jest choćby częściowe zautomatyzowanie takiego procesu. Uwzględnienia wymaga również różnorodność formatów, w których publikacje występują. Każdy format wymaga odrębnego programu konwertującego (przygotowania konwertera) [Fülle i Ott, 2006, s. 12]. W celu zwiększenia efektywności i obniżenia kosztów migracji, celowe wydaje się ustalenie zamkniętej listy formatów, w których twórcy powinni przekazywać dokumenty przeznaczone do długotrwałej archiwizacji. Lista powinna być aktualizowana zgodnie z obserwowanymi zmianami technologicznymi zachodzącymi w tym zakresie.

Podsumowując, istota migracji polega na sukcesywnym przenoszeniu danych ze starszych, wychodzących z użycia formatów do formatów nowej generacji. Efektywność procesu migracji wymaga zarówno obserwacji zmian technologicznych (w zakresie formatów zapisu danych cyfrowych), jak i stosunkowo szybkiego reagowania na zaistniałe zmiany. W przypadku zbyt późnego podjęcia działań migracja do nowego formatu może okazać się problematyczna, głównie ze względu na trudności z odtworzeniem dokumentu w jego oryginalnym środowisku sprzętowo-programowym, co jest niezbędne w celu udokumentowania odstępstw między dokumentem otrzymanym w procesie migracji a wersją migrowaną.

Sposobem na uniknięcie przekłamań i odstępstw w kolejnych wersjach dokumentów cyfrowych od ich oryginałów, czyli ochronę dokumentów wraz z ich oryginalnym „look and feel", jest użytkowanie dokumentów w ich oryginalnym otoczeniu sprzętowo-programowym poprzez wyemulowanie ich oryginalnego otoczenia na urządzeniach aktualnie dostępnych [Oltmans i Kol, 2005].

2.1.7. Emulacja jako metoda długoterminowej archiwizacji dokumentów cyfrowych

W terminologii specjalistycznej pojęcie *emulacja* jest używane dla określenia procesu naśladowania, symulowania, a także imitowania zachowań określonego sprzętu i oprogramowania [Attributes, 2001; Schneider, 1997, s. 281]. W *Komputerowej Encyklopedii Microsoftu* [Woodcock, 2002, s. 191] termin *emulacja* został zdefiniowany jako proces imitowania przez komputer, urządzenie lub program

funkcji, które spełnia inny komputer, urządzenie lub program. Metoda emulacji polega na tworzeniu programów emulujących starsze platformy programowo-sprzętowe na platformach aktualnie użytkowanych [Borghoff i in., 2003, s. 59-69; Rothenberg, 2000, s. 1-4]. Zadaniem programów emulujących, nazywanych emulatorami, jest możliwie dokładne symulowanie architektury systemu, by różnica pomiędzy oryginalnym oraz naśladowanym systemem była niezauważalna. Emulator można zrealizować sprzętowo, programowo lub stosując obie te metody [Freedman, 2004, s. 218]. W przypadku dokumentów cyfrowych emulacja oznacza proces reprodukcji ich pierwotnego fenotypu [Borghoff i in., 2003, s. 59-83].

Emulacja jest traktowana jako jedyny sposób uniknięcia przekłamań i odstępstw w kolejnych wersjach dokumentów. Jedyną formą zapewnienia użytkownika o pełnej zgodności z oryginalną wersją, jest udostępnienie dokumentu w jego pierwotnej aplikacji, czyli wyemulowanie oryginalnego otoczenia programowego dokumentu.

Metoda emulacji stosowana jest głównie w przypadku dokumentów cyfrowych, których treść i program prezentujący są ze sobą nierozerwalnie powiązane. Często zdarza się, że aplikacja stworzona dla określonego systemu operacyjnego jest z nim tak powiązana, iż jej późniejsze przełożenie i zastosowanie w innych warunkach systemowych staje się niemożliwe. Zachodzi wówczas konieczność emulowania oryginalnego środowiska programowo-sprzętowego [Fülle i Ott, 2006, s. 12-14; Liegmann, 2001, s. 104]. Emulowanie można więc postrzegać jako migrację nie danych cyfrowych, lecz otoczenia ich odczytu.

Emulację opisuje się jako metodę polegającą na „zmuszaniu" przyszłych technologii do funkcjonowania identycznie jak oryginalne środowisko zachowanego obiektu, co ma pozwolić na prezentację oryginalnego obiektu w jego pierwotnej postaci na podstawie oryginalnego strumienia danych. Jednocześnie zwraca się uwagę na różnicę pomiędzy emulacją otoczenia sprzętowego a emulacją otoczenia programowego. Za bardziej odpowiednią uznawana jest emulacja sprzętu, co wynika z faktu, że specyfikacje sprzętu mogą okazać się łatwiejsze do zdefiniowania niż specyfikacje oprogramowania. Poza tym emulacja platformy sprzętowej może być bardzo elastyczna i umożliwić tym samym odtworzenie wielu systemów oraz prezentację różnych obiektów cyfrowych. Jako rozwiązanie alternatywne wymieniana jest emulacja pewnych aplikacji lub ich zachowań, jednak z uwagą, iż wadą takiego rozwiązania jest konieczność opracowania indywidualnego emulatora dla każdej aplikacji [Borghoff i in., 2003, s. 59-83; National Library of Australia, 2003, s. 149].

Wymienia się trzy alternatywne elementy mogące stanowić przedmiot emulacji. Są to: platforma sprzętowa (czyli komputer), platforma sprzętowa wraz z systemem operacyjnym oraz kompletne otoczenie odczytu dokumentu (tj. platforma sprzętowa wraz z systemem operacyjnym oraz programem umożliwiającym prezentację dokumentu) [Borghoff i in., 2003, s. 59-83].

O przedmiocie emulacji powinno zdecydować się na etapie planowania strategii długoterminowej archiwizacji. Jeśli emulowany ma być komputer, to zachodzi konieczność zachowania dokumentu elektronicznego wraz z odpowiednim programem go prezentującym oraz systemem operacyjnym; jeśli emulacji ma podlegać komputer i system operacyjny – archiwizuje się dokument wraz z programem go prezentującym. Wydawałoby się, że rozwiązaniem optymalnym jest odtworzenie w procesie emulacji kompletnego otoczenia odczytu dokumentu elektronicznego. Specjaliści tłumaczą jednak, że to złudne, ponieważ zwiększona ilość kombinacji „komputer – system operacyjny – program umożliwiający prezentację dokumentu" pociąga za sobą konieczność tworzenia dużej ilości emulatorów, tymczasem należy dążyć do minimalizacji złożoności narzędzi emulujących. Pomimo intensywnych badań i starań informatyków, nadal problemem jest precyzyjne, kompletne odtworzenie działania kompleksowych programów. Dużo łatwiej odtworzyć działanie sprzętu, toteż najbardziej zasadna zdaje się alternatywa pierwsza, czyli emulacja samego komputera i archiwizowanie, oprócz dokumentu, także stosownego systemu operacyjnego oraz programu umożliwiającego prezentację treści [Borghoff i in., 2003; Funk, 2008b; Liegmann, 2001].

Omawiając emulację, należałoby zwrócić uwagę na jej zalety i wady jako metody długoterminowej archiwizacji obiektów cyfrowych. Potencjalną zaletą emulacji jest fakt, iż jest znaną techniką informatyczną, która wypracowała istnienie emulatorów różnych platform i systemów: od najstarszych, tworzonych przez entuzjastów, do systemów nowoczesnych, tworzonych w celach komercyjnych, służących do testowania i uruchamiania oprogramowania na różnych platformach. Przy możliwie najszerszym zastosowaniu tej metody emulacja umożliwiałaby odtworzenie pełnej funkcjonalności wielu obiektów cyfrowych (w tym oprogramowania) na podstawie oryginalnego, niezmodyfikowanego strumienia danych w połączeniu z oryginalnym oprogramowaniem [Borghoff i in., 2003, s. 59-83; Funk, 2008b; National Library of Australia, 2003, s. 149]. Emulacja umożliwia zapewnienie autentyczności dokumentów cyfrowych, z racji zachowania w długim czasie niezmiennej struktury oryginalnego dokumentu oraz wiernego odtworzenia pierwotnego otoczenia sprzętowo-programowego. Nie ma również ograniczeń w zakresie typów dokumentów; specjaliści twierdzą, że przy zastosowaniu emulacji mogą zostać długoterminowo zarchiwizowane nawet dokumenty dynamiczne.

Do wad emulacji zalicza się fakt, że to metoda skomplikowana technicznie, wymagająca dużych nakładów pracy i specjalistycznej wiedzy, co wiąże się z wysokimi kosztami. Jako metoda długoterminowej ochrony obiektów cyfrowych emulacja wymaga wciąż wielu badań. Konieczne jest systematyczne prowadzenie eksperymentów, potwierdzających możliwość zastosowania tej metody do odczytu określonych typów publikacji elektronicznych. Jej zastosowanie w przyszłości

może być znacznie utrudnione, a w przypadku emulacji kompletnego środowiska odczytu i prezentacji dokumentu (sprzętu, systemu i oprogramowania) praktycznie niemożliwe z racji nieodpowiedniej dokumentacji współczesnego oprogramowania bądź stosowania niestandardowych formatów. Niemożliwa wydaje się także emulacja wszystkich funkcji systemu lub aplikacji z uwagi na wzrost złożoności systemów. Wraz ze zmianami przyszłych technologii i platform również emulatory będą wymagały konwersji lub wyemulowania ich własnych środowisk w nowych systemach, co oznacza nakładanie się wielu warstw emulatorów [Borghoff i in., 2003, s. 59-83; Funk, 2008b].

Długoterminowe użytkowanie sprzętu i oprogramowania wymaga regulacji prawnych. Symulowane przez emulatory systemy objęte są zwykle ochroną autorsko-prawną, zatem koniecznym jest pozyskanie licencji. Problematyczne może okazać się użytkowanie dokumentów cyfrowych w dalekiej przyszłości z powodu całkiem realnych zmian, które mogą nastąpić w zakresie standardowych elementów obsługi sprzętu komputerowego. Jest bowiem wielce prawdopodobne, że klawiatura i mysz wyjdą z użycia, prawdopodobnie zmienią się także formy komunikowania się człowieka z maszyną [Borghoff i in., 2003, s. 59-83; Funk, 2008b].

W wypowiedziach specjalistów zwraca się uwagę na konieczność tworzenia technicznych specyfikacji, zawierających szczegółowy opis wszystkich istotnych cech platform sprzętowych w celu ich odtworzenia na przyszłych platformach. Dodatkowo należy dążyć do tworzenia takich technicznych rozwiązań, które umożliwią hosting potrzebnych emulatorów na przyszłych platformach przy minimalnym nakładzie starań. Niezbędne jest tworzenie metadanych opisujących dokument cyfrowy i przyporządkowujących go do odpowiedniego oprogramowania oraz emulatorów umożliwiających odczyt w przyszłości. Bezwzględnie potrzebne jest zidentyfikowanie kryteriów autentyczności dokumentów i poprawności jej testowania dla poszczególnych typów dokumentów cyfrowych, w celu wprowadzenia mechanizmów oceny efektywności procesu długoterminowej archiwizacji bazującej na metodzie emulacji czy na innych metodach archiwizacji. Przyszłe badania nad potencjałem emulacji, jako metody długoterminowej ochrony użyteczności zasobów cyfrowych, nie obejdą się bez cyklicznego wykonywania testów i eksperymentów. Jeśli podczas serii eksperymentów nie wystąpią nieprzewidziane efekty, zasadnym będzie uznanie, że problem długoterminowej ochrony danych cyfrowych może być rozwiązany właśnie przy zastosowaniu metody emulacji [Giaretta, 2011; Rothenberg, 1995, s. 42-47; Suchodoletz, 2008, s. 64-144].

Konkludując, metoda emulacji, poprzez możliwość odtwarzania oryginalnego technicznego otoczenia obiektu cyfrowego przy użyciu aktualnej technologii, jest traktowana jako istotna strategia permanentnego dostępu do cyfrowego materiału. Jednak, wciąż jeszcze wymaga wielu badań i testów potwierdzających jej pełną

użyteczność w przypadku elektronicznych dokumentów bibliotecznych, archiwalnych czy muzealnych. Trudność polega na tym, że kierunek rozwoju i postęp w zakresie tworzonych narzędzi emulujących nie są podyktowane potrzebami instytucji pamięci, lecz uzależnione od potrzeb firm informatycznych. Badanie potrzeb bibliotek lub archiwów i opracowywanie rozwiązań wychodzących im naprzeciw to kosztowne inwestycje. Mogą sobie na nie pozwolić jedynie instytucje o świetnej kondycji finansowej.

2.1.8. Formaty zapisu dokumentów cyfrowych

Za jeden z bardziej złożonych problemów w systemach bibliotecznych i archiwalnych uznaje się normalizację zasobów. Instytucje pamięci mają do czynienia z mnogością formatów i parametrów dokumentów wytworzonych do różnych celów, co pociąga za sobą szereg specyficznych cech, przydatnych w jednych sytuacjach, a w innych niepożądanych i uciążliwych [Radwański, 2005, s. 102]. Potrzebna jest więc unifikacja formatów zapisu dokumentów cyfrowych. Starać się o to powinny przede wszystkim instytucje powołane do zadań ochrony zbiorów, ale również twórcy, którym – przynajmniej z założenia – powinno zależeć na długoterminowej użyteczności ich produktów. Optymalne, z punktu widzenia efektywności procesów archiwizacji, byłoby stosowanie przez twórców tzw. formatów otwartych (ang. *open file format*), do których dokumentacja jest powszechnie dostępna i nie podlega jakimkolwiek ograniczeniom, choćby natury licencyjnej [Fülle i Ott, 2006, s. 19]. Format otwarty cechuje się jawną, ogólnodostępną strukturą w odróżnieniu od formatu zamkniętego (ang. *proprietary format*), który zazwyczaj jest opatentowany lub chroniony restrykcyjnymi licencjami [Funk, 2008a].

Formaty otwarte – standardowe, kompatybilne z różnymi będącymi w zastosowaniu systemami – bywają też określane jako formaty wymiany oraz formaty długoterminowo stabilne [Fülle i Ott, 2006, s. 15-17]. Za optymalne uznawane są formaty, do których istnieje szczegółowa, ogólnodostępna dokumentacja, najlepiej w postaci normy ISO. Potrzebna jest jednak akceptacja i deklaracja stosowania tych formatów przez twórców publikacji i równocześnie odsyłanie standaryzowanych publikacji do instytucji archiwizujących. Publikacje w formatach standardowych zazwyczaj nie wymagają konwertowania, są przyjmowane i archiwizowane w oryginalnym środowisku.

Formaty zapisu cyfrowych dokumentów tekstowych

Plik tekstowy (ang. *plain text file*, ASCII file) zawiera jedynie tekst zapisany zgodnie z przyjętym standardem – zazwyczaj ASCII lub Unicode. Tekst jest zakodowany w wybranym standardzie wraz z podstawowymi komendami jego

sterowania (w kodzie ASCII są to znaki o kodach mniejszych od 32). Plik tekstowy w systemie operacyjnym ma zazwyczaj rozszerzenie TXT i charakteryzuje się małą objętością. W pierwszych znanych w świecie kolekcjach cyfrowych był stosowany do prezentacji dokumentów. Współcześnie jego zastosowanie do tych celów jest bardzo rzadkie [Freedmann, 2004, s. 597; Kolasa, 2012, s. 409-411].

RTF (Rich Text Format) – jest określany jako wzbogacony dokument tekstowy lub format tekstu wzbogaconego. Został opracowany w 1987 r. przez firmę Microsoft i jest formatem pliku służącym do międzyplatformowej wymiany informacji pomiędzy procesorami tekstów.

RTF posługuje się najczęściej zestawem znaków ANSI i zawiera informacje dotyczące formatowania tekstu na ekranie oraz wydruku. Od wersji 1.6 RTF obsługuje standard Unicode. Użycie dodatkowych kodów sterujących pozwala zapisywać dokumenty zawierające, obok treści, także informację o kroju czcionki, jej wielkości, kolorze oraz, np. tabele. Dzięki temu możliwe jest stosowanie RTF jako uniwersalnego formatu zapisu danych tekstowych w celu ich użytkowania na różnych platformach programowo-sprzętowych [Coy, 2006; Czajkowski, 2002, s. 545; Freedman, 2004, s. 704].

SGML (Standard Generalized Markup Language) – tłumaczony na język polski jako Standardowy Uogólniony Język Znaczników; ma status standardu ISO. SGML jest stosowany do opisu zasad interpretacji dokumentu, dlatego określa się go jako metajęzyk. Język SGML charakteryzuje się elastycznością i kompleksowością, dając tym samym użytkownikom szerokie możliwości, co stanowi jego zaletę. Jednakże bardzo wymagające i pracochłonne jest stworzenie oprogramowania, które umożliwia jego interpretację. Format SGML znajduje zastosowanie głównie w branży lotniczej, wojskowej i motoryzacyjnej. W dużo mniejszym zakresie posługują sie nim wydawcy i miejsca wydawnicze w procesach publikowania online. Faktem jednak jest, że jednym z pierwszych użytkowników SGML było Stowarzyszenie Amerykańskich Wydawców, stosując ten format dla celów ujednolicenia zasad tworzenia i wymiany elektronicznych dokumentów między autorami, redaktorami i wydawcami [Coy, 2006, s. 43-51; Czajkowski, 2002, s. 565; Freedman, 2004, s. 726].

HTML (Hypertext Markup Language) – jest opisywany jako Język Znaczników Hipertekstu i definiowany jako format dokumentów używanych w sieci WWW [Freedman, 2004, s. 282], HTML ma status języka prezentacji; pochodzi od języka SGML i stanowi jego konkretne, bardzo zawężone zastosowanie [Fülle i Ott, 2006, s. 19]. Zalicza się go do języków strukturalnych, tzn. opisujących strukturę dokumentów. Nadaje się raczej do prezentacji dokumentów, niż do ich trwałego archiwizowania [Coy, 2006, s. 44-45]. Ograniczenia formatu HTML wywołały potrzebę opracowania nowego, rozszerzonego formatu, nazwanego XML.

XML (EXtensible Markup Language) – otwarty standard opisu danych XML; znajduje zastosowanie przy definiowaniu elementów danych umieszczanych m.in. na stronach internetowych. Język XML posługuje się podobną strukturą znaczników jak HTML, przy czym ten drugi korzysta ze znaczników predefiniowanych, natomiast XML umożliwia projektantowi strony tworzenie znaczników własnych [Freedman, 2004, s. 973]. Uznano, że XML nadaje się doskonale do tzw. niezależnego od mediów zapisu treści i spełnia wszelkie kryteria formatu długoterminowo stabilnego [Fülle i Ott, 2006, s. 20].

Zaleta formatów XML oraz HTML tkwi w przystępności zapisu w nich dokumentów. Inaczej mówiąc, nie są potrzebne skomplikowane, złożone narzędzia, a wystarczy prosty edytor tekstu. Zarówno treść dokumentu, jak i znaczniki zakodowane są w postaci tekstu, co powoduje, że ich odczyt jest łatwy, analogiczny z procesami ich migracji i emulacji. Jednak relacja objętości opisu dokumentu w stosunku do treści dokumentu jest zdecydowanie mniej korzystna niż choćby w formatach PDF i TIFF [Borghoff i in., 2003, s. 110].

Formaty zapisu cyfrowych dokumentów graficznych

TIFF (Tagged Image File Format) – określany jest jako *znacznikowy format zapisu obrazu*. To komputerowy format plików graficznych, służący do zapisywania grafiki rastrowej. Uznawany za przenośny, najbardziej uniwersalny format, powszechnie stosowany w przemyśle drukarskim i poligraficznym. Podkreślana jest jego przydatność z racji takich cech jak: obsługa wszystkich rodzajów głębi i przestrzeni kolorów (RGB, CMYK), możliwość bezstratnej kompresji (algorytm LZW), możliwość przechowywania ścieżek i kanałów alfa oraz profili koloru i komentarzy tekstowych. Nade wszystko jest to standard międzyplatformowy [Kolasa, 2012, s. 406-407; Leśniewski, 2006]. TIFF jest precyzyjnie zdefiniowany normą ISO [ISO, 2001; ISO, 2004]. Powstało kilka różnych jego wersji, co niekiedy bywa problematyczne w kwestii zgodności pomiędzy programami [Freedman, 2004, s. 842-843].

Standard TIFF nadaje się do zapisu zdjęć archiwalnych, drukowania i archiwizowania fotografii cyfrowych. Obrazy zapisywane są z tzw. kompresją bezstratną, co oznacza, że podczas ich rejestracji nie dochodzi do usunięcia żadnych danych z pliku. Pozwala to zarówno na odtworzenie obrazu w jego pierwotnej postaci, jak i na jego dalszą obróbkę lub korektę [Kowalska, 2007, s. 33]. TIFF jest dobrym formatem do druku, gdyż potrafi zachować informacje o zarządzaniu kolorami, podziale kolorów, ścieżce (sposobie) przycinania motywów obrazu bez tła, sprawdza się również jako format archiwizacyjny z racji dobrej, bezstratnej jakości obrazów [Fülle i Ott, 2006, s. 21]. Dzięki rozpowszechnieniu oraz niezależności

sprzętowej TIFF odpowiada potrzebom procesu długoterminowej archiwizacji [Fülle i Ott, 2006, s. 21].

GIF (Grafics Interchange Format) – określany jako format wymiany grafiki służy do zapisu grafiki rastrowej [Freedman, 2004, s. 255]. To popularny format obsługujący kolory indeksowane. Duża skala jego występowania w Internecie wynika z obsługi przezroczystości w obrazach graficznych. Obrazy są zapisywane wyłącznie w formie prostokątnej lub kwadratowej, dlatego eksponowanie wybranego fragmentu (np. okrągłego przycisku) i ukrycie narożnych obszarów obrazu stanowi bardzo atrakcyjną funkcję tego formatu. Ponadto można go używać do zapisywania animacji. Jedną z wad jest konieczność redukcji liczby kolorów do ściśle określonej palety zawierającej nie więcej niż 256 barw [Czajkowski, 2002, s. 216; Leśniewski, 2002]. Z tego względu, zastosowanie GIF może powodować znaczne odstępstwa w stosunku do obrazu oryginalnego, co nie równoważy niewielkiego rozmiaru pliku wynikowego. Dlatego też nie jest zalecany do stosowania w procesach długoterminowej archiwizacji.

PNG (Portable Network Graphics) – jest formatem zapisu grafiki rastrowej o rozbudowanych możliwościach. Obsługuje 48-bitowy kolor, kanały alfa, korekcję gamma i kolorów oraz kompresję obrazu. Pozwala też stosować inną rozdzielczość przy drukowaniu, a inną przy wyświetlaniu obrazu na ekranie [Freedman, 2004, s. 607]. W 2004 r. format PNG został ustandaryzowany normą ISO. Rekomendowany jest przez W3C (World Wide Web Consortium) jako format archiwizacyjny; nie jest jednak tak popularny jak TIFF albo EPS. Za stosowaniem PNG jako formatu archiwizacyjnego przemawia fakt, że – w przeciwieństwie do TIFF – jest formatem otwartym. Przy zastosowaniu PNG do celów archiwizacji zaleca się stosowanie tylko kompresji bezstratnej, rezygnację z kolorów paletowych oraz zachowanie wraz z plikiem obrazu możliwie wiele dostępnych informacji technicznych na jego temat [Fülle i Ott, 2006, s. 22]. PNG, zgodnie z przewidywaniami, prawdopodobnie zastąpi format GIF, którego używanie łączy się z konsekwencjami natury prawnej. Metoda kompresji w formacie GIF została objęta patentem, co narzuca występowanie ograniczeń wynikających z warunków licencjonowania.

JPEG (Joint Photographic Experts Group) – to format standardowy służący do zapisu statycznych obrazów; bardzo popularny ze względu na wysoki współczynnik kompresji. Dzięki zastosowaniu dyskretnej transformaty kosinusowej uzyskuje się kompresję stratną ze współczynnikami kompresji dochodzącymi do 100:1, zważywszy, że utrata jakości przy współczynnikach od 10:1 do 20:1 jest niemal niezauważalna [Freedman, 2004, s. 346].

Format JPEG obsługuje kolory w trybie RGB, CMYK i odcieniach szarości, jednak nie oferuje ochrony przezroczystości ani kanałów alfa [Leśniewski, 2002].

Rozszerzeniem formatu JPEG jest JPEG++, umożliwiający wybieranie osobnych współczynników dla różnych obszarów ilustracji. Przykładowo tło może być skompresowane z wyższym współczynnikiem kompresji niż pierwszy plan ilustracji [Freedman, 2004, s. 346].

Obrazy kompresowane algorytmem JPEG są zapisywane w plikach w formacie JFIF (*JPEG File Interchange Format* – format wymiany plików JPEG). Stosowane rozszerzenia plików to .jpg oraz .jff. M-JPG oraz MPEG są odmianami tej kompresji dla cyfrowego wideo [Freedman, 2004, s. 346].

Z uwagi, że każdorazowa edycja obrazu wymaga zapisania pliku, to zaś wywołuje kompresję i w efekcie powoduje obniżenie jakości obrazu, format ten nie jest najlepszym rozwiązaniem do archiwizowania fotografii cyfrowych [Kowalska, 2007, s. 34]. Wyjątek stanowi format JPEG 2000 (opisany normą ISO) opracowany z uwzględnieniem kryteriów, którymi powinny charakteryzować się formaty archiwizacyjne. Jednak z racji wciąż niskiego stopnia akceptacji i rozpowszechnienia, format JPEG 2000 nadal nie posiada statusu formatu archiwizacyjnego [Fülle i Ott, 2006, s. 23].

SVG (Scalable Vector Graphics) – jest formatem zapisu grafiki wektorowej, służącym do włączania rysunków wektorowych do publikowanych w sieci WWW stron XML. Zgodnie z prognozami format SVG stał się międzynarodowym standardem zapisu grafiki wektorowej, standaryzowanym i wspieranym przez W3C, a także rekomendowanym jako format archiwizacyjny grafiki wektorowej [Freedman, 2004, s. 784; Fülle i Ott, 2006, s. 26].

Format SVG pozwala na używanie języków skryptowych, szablonów stylów (CSS), a także na rozszerzanie jego funkcjonalności poprzez dodawanie własnych właściwości zgodnie z założeniami XML [Fülle i Ott, 2006, s. 26].

FITS (Flexible Image Transport System) – to format pliku stosowany do przechowywania, przekazywania oraz przetwarzania obrazów zawierających informacje naukowe. Znajduje zastosowanie w branżach związanych z astronomią; korzysta z niego m.in. agencja NASA. Format FITS, zaprojektowany do przechowywania informacji naukowych, daje możliwość zapisania wielu dodatkowych danych wraz z oryginalnymi metadanymi. Główną zaletą formatu jest to, że metadane obrazów przechowywane są w nagłówku zapisanym standardowymi znakami ASCII, czytelnymi dla człowieka. Każdy plik w formacie FITS składa się z co najmniej jednego nagłówka zawierającego 80 znakowe bloki znaków ASCII kodujących pary *klucz/wartość*. Pary mogą dostarczać informacji o wymiarach obrazu, formacie danych, zawierać komentarze oraz dowolną inną informację, którą autor w nich zamieści [FITS, 2008].

PostScript to uniwersalny język opisu stron, używany powszechnie na wszystkich platformach. Niezależny od urządzenia wyjściowego, wyspecjalizowany w opisie strony/grafiki. Pierwsza wersja została ogłoszona w 1985 r. przez Adobe Systems.

PostScript jest wykorzystywany w profesjonalnych drukarniach i naświetlarniach. W 1990 r. powstała wersja PostScript Level 2 wzbogacona o obsługę kolorów. Kolejne wersje wniosły dodatkowe rozszerzenia i wbudowane czcionki. Z racji powszechności drukarek wyposażonych w jego interpreter oraz programowych interpreterów, a także precyzyjnego odwzorowania wydruku, PostScript stał się popularnym formatem wymiany gotowych materiałów [Czajkowski, 2002, s. 485; Freedman, 2004, s. 621-622].

EPS (Encapsulated Postscript) – jest określany jako hermetyzowany Postscript i służy do przenoszenia ilustracji graficznych pomiędzy aplikacjami oraz platformami. Pliki EPS zawierają kod Postscript zapisany za pomocą tekstu w formacie ASCII, a także opcjonalną ilustrację umożliwiającą podgląd plików w innych formatach (np. w TIFF). Typowe zastosowanie formatu EPS polega na eksporcie ilustracji utworzonej w formacie graficznym do formatu EPS oraz imporcie tego pliku do programu służącego do składu publikacji. Pliki w formacie EPS są znacznie większe niż pliki w innych formatach graficznych, jednak z uwagi na fakt, iż są to pliki tekstowe, można je skompresować i uzyskać około jednej czwartej rozmiaru pierwotnego [Freedman, 2004, s. 221].

Z racji powszechnego dostępu do dokumentacji formatu EPS oraz niezależności systemowej, nadaje się on do długoterminowej archiwizacji. Jednak przy wyborze formatu archiwizacyjnego należy również uwzględnić duże zapotrzebowanie na moc obliczeniową podczas operacji wykonywanych na tym formacie (w tym także operacji odczytu i prezentacji zawartości).

PDF (Portbale Document Format) – czyli tzw. przenośny format dokumentu, jest językiem opisu stron, który bazuje na języku PostScript i został wzbogacony o hipertekstowość. Obsługiwany przez aplikację Adobe Acrobat, PDF jest stosowany do opisu układu dokumentów i/lub obrazów graficznych w sposób niezależny od platformy lub drukarki. Pliki PDF mogą zawierać zarówno grafikę rastrową, jak i wektorową, osadzone czcionki, mechanizmy wyszukiwania, podpisy cyfrowe oraz wiele innych zaawansowanych funkcji [Leśniewski, 2002]. Format PDF zachowuje informacje nie tylko o treści dokumentu, ale również o jego wyglądzie (rozmiary ilustracji, marginesy, rozkład elementów na stronie) [Kowalska, 2007, s. 34-35].

PDF doczekał się standaryzacji międzynarodowej. Specjalnie dla celów długoterminowej archiwizacji powstała norma ISO dotycząca formatu PDF/A jako standardu archiwizacyjnego dla plików cyfrowych (ISO/CD 19005-1) [ISO, 2005]. PDF/A został pomyślany jako format dokumentu zdolny zachować połączenie tekstu, obrazów, grafiki wektorowej. Powinien również umożliwiać zdefiniowanie właściwości oraz możliwości systemów, które znajdują zastosowanie przy odczytywaniu, reprodukcji i wyszukiwaniu pełnotekstowym. Należy spodziewać się, że będzie zawierać wszelkie metadane niezbędne do opisu oraz użytkowania dokumentów [Fülle i Ott, 2006, s. 29]. Możliwości formatu PDF/A są stale rozwijane z myślą o różnych

typach dokumentów cyfrowych. Po PDF/A-2 w 2012 r. została opublikowana wersja PDF/A-3 umożliwiająca archiwizowanie np. zasobów skrzynek e-mailowych.

DjVu jest formatem charakteryzującym się przede wszystkim bardzo wydajną metodą kompresji obrazu. Metoda ta została opracowana przez naukowców amerykańskiego koncernu AT&T do kompresji dokumentów skanowanych w kolorze. Zamierzeniem twórców formatu DjVu było umożliwienie tworzenia cyfrowych bibliotek, w których książki byłyby przechowywane na nośnikach elektronicznych. Cel taki łączył się z koniecznością stworzenia formatu plików graficznych, który umożliwiałby przechowywanie zeskanowanego tekstu z jakością odpowiadającą papierowemu oryginałowi przy jednoczesnym względnie małym rozmiarze pliku. Z racji istotnych cech formatu DjVu jego popularność i zastosowanie w polskich bibliotekach cyfrowych, od początku ich rozwoju, stale rosły [Bednarek, 2006; Siedlarz, 2012, s. 423-436]. W oświadczeniu PCSS z 2015 r. potwierdzono, że DjVu znalazł zastosowanie w polskich bibliotekach cyfrowych właśnie z racji bardzo dobrych parametrów kompresji obrazu, przy zachowaniu wystarczająco dobrej jakości oraz z możliwością zapisu wielostronicowego obiektu w postaci wieloplikowej. Było to kluczowe w początkowej fazie rozwoju polskich bibliotek cyfrowych, kiedy prędkość połączeń internetowych była niewielka. Przyznano, że stosowanie tego formatu przyjmowane było (często bez refleksji nad zachodzącymi zmianami technologicznymi) przez kolejne projekty digitalizacyjne, co zaowocowało dominacją plików DjVu w polskich kolekcjach cyfrowych. Jednak z racji zmian wprowadzonych w przeglądarce internetowej Google Chrome wystąpiły utrudnienia odczytu tych plików. Dlatego zaproponowano ograniczenie stosowania formatu DjVu na rzecz PDF [Oświadczenie, 2015]. Wg opinii polskich specjalistów DjVu nie jest i nigdy nie był formatem przeznaczonym do archiwizacji trwałej. Zwrócili oni uwagę na opinię użytkowników bibliotek cyfrowych o formacie DjVu, bolesną dla wielu bibliotekarzy i organizatorów kolekcji cyfrowych, stanowiącą też naukę na przyszłość nt. całościowego zrozumienia procesów digitalizacji i otwartości na różne modele prezentacji zasobu cyfrowego w sieci WWW. Format DjVu został uznany za martwy; nie będzie wspierany technologicznie i nie stanowi przedmiotu dyskusji. Obecnie wspieranym formatem jest PDF[2].

Formaty zapisu cyfrowych dokumentów multimedialnych

Pojęciem *multimediów* bądź *dokumentów/publikacji multimedialnych* określa się obiekty łączące różnorodne formy przekazu treści, tj.: tekst, dźwięk, grafikę, animację, wideo, najczęściej wzajemnie się uzupełniające, użyte w jednym przekazie [Dubisz, 2003, s. 740].

[2] Opinie polskich specjalistów zebrane w 2017 r.

Spośród wielu możliwych formatów zapisu plików audio, tylko nieliczne gwarantują bezstratną kompresję oraz traktowane są jako standard bądź posiadają status normy ISO. Do takich zaliczają się WAV (Windows WAVeform) oraz AIFF (Audio Interchange File Format) [Grossmann, 1997, s. 319]. Uruchamiają się na rozmaitych platformach, a ich dokumentacja jest powszechnie znana; z tego względu mogą pełnić i pełnią funkcję formatów archiwalnych. Ich alternatywą jest format ALS (Audio Lossless Coding) stanowiący część standardu MPEG-4. Nie jest wykluczone, że format MPEG-4-ALS stanie się w przyszłości standardem w archiwizacji plików audio [Formats, b.d.].

Format MP3, którego rzeczywista nazwa to MPEG-1 Layer III, zdobył popularność tym, że w stosunku do nieskompresowanego formatu WAV, przy niewielkim pogorszeniu jakości dźwięku pozwala na ok. dziesięciokrotne zmniejszenie objętości danych. Zastosowany w MP3 algorytm kompresji usuwa z sygnału najsłabsze dźwięki, które nie są odbierane przez większość ludzi. Ze względu na utratę części kodowanego sygnału, do zastosowań archiwizacyjnych format MP3 lepiej zastąpić bezstratnymi formatami zapisu dźwięku jak choćby FLAC (Free Lossless Audio Codec).

Wszystkie dane zapisane w formacie FLAC podlegają procesowi kompresji, jednakże można odzyskać ich pierwotną, niezmienioną postać, dokonując dekompresji. Format FLAC został stworzony specjalnie do zmniejszania objętości plików dźwiękowych, dlatego pozwala otrzymać stosunkowo wysoki stopień kompresji, dochodzący nawet do 60%. „Koder/dekoder" stratnej kompresji może uzyskać stopień kompresji nawet 90%, jednak jest to związane z usunięciem pewnych danych.

Cyfrowe pliki wideo w zasadzie mogą być zapisane w każdym formacie grafiki rastrowej, jednak w praktyce przy ich zapisie zastosowanie znajdują formaty wideo. W obszarach cyfrowej telewizji i kina rozwiązania normatywne wynikają z norm CCIR 601 [Marshall, 1999] oraz ANSI/SMPTE 268M-1994 [SMPTE, 2015; Digital Moving, 2017]. W przypadkach kiedy konieczne jest kompresowanie plików wideo, stosuje się format MPEG-2 określony normą ISO. Umożliwia on wysoką kompresję danych w plikach różnej jakości i rozdzielczości, także w połączeniu z różnego pochodzenia danymi audio. Format MPEG-2 powstał z myślą o profesjonalnych zastosowaniach w studiach telewizyjnych i radiowych. Rozpowszechnił się jednak głównie jako format zapisu filmów na płytach DVD. Bywa też coraz częściej używany jako format wideo w Internecie (sieciowe zapisy wideo). Nowszy standard MPEG-4 nie ma funkcji zastępczej dla MPEG-2, lecz powstał z myślą o optymalizacji strumieniowego przesyłania obrazu przy ograniczonej przepustowości łącza, np. na potrzeby nagrań wideo przez telefon komórkowy [Grossmann, 1997, s. 201].

W procesach długoterminowej archiwizacji plików wideo za optymalny uznaje się format MPEG-2 wraz z metadanymi w formacie MPEG-7 [Martínez, 2004].

Dla potrzeb archiwizacji plików HQ zaleca się stosowanie formatów standardowych CCIR 601 wraz z kontenerem wymiennym MXF [Fülle i Ott, 2006, s. 33].

Formaty zapisu dokumentów interaktywnych

Do zapisu dokumentów zawierających animacje, gry, zabawy, formularze interaktywne i inne stosowany jest najczęściej format Flash [Adobe, b.d.], który został opracowany przez firmę Macromedia i podlega jej kontroli (aktualnie właścicielem praw do formatu jest firma Adobe). To format dobrze udokumentowany, a pozyskanie narzędzi alternatywnych nie stanowi większego problemu. Obejmuje środowisko programistyczne umożliwiające tworzenie animacji z zastosowaniem grafiki wektorowej. Powstałe pliki można odtwarzać za pomocą przeglądarki internetowej z zainstalowaną odpowiednią wtyczką (Flash Player) lub w oddzielnym programie. Alternatywą dla formatu Flash może być format SVG lub SMIL. SMIL (wym. [smail], ang. *Synchronized Multimedia Integration Language*) – standard zalecany przez W3C do opisu prezentacji multimedialnych przygotowanych w XML [SMIL, 2008].

SMIL określa, m.in.: znaczniki synchronizacji, układu, animacji, przejść obrazu oraz zagnieżdżania; pozwala na wstawianie obrazu, animacji, muzyki oraz filmów i umożliwia synchronizację tych elementów [SMIL, 2008].

Do zapisu dokumentów interaktywnych używany jest także format VRML (Virtual Reality Modelling Language – tłumaczony jako język modelowania rzeczywistości wirtualnej) [VRML, 1995] oraz X3D [X3D, 2017]. VRML umożliwia wyświetlanie trójwymiarowych obrazów na stronach WWW. Rywalizuje z językiem HTML jako niezależny od danej platformy język, umożliwiający tworzenie scen rzeczywistości wirtualnej, które użytkownik przemierza podobnie jak konwencjonalne strony WWW, korzystając z łączy. Za pomocą VRML można stworzyć np. wirtualne muzeum z elementami naśladującymi prawdziwy budynek muzealny: korytarzami, gablotami, pokazami multimedialnymi, czy supermarket z półkami pełnymi towarów, którymi można manipulować za pomocą myszy, bądź wirtualną bibliotekę, w której „książki" wyjmuje się z „półek" [Grossmann, 1997, s. 316-317; X3D, 2017].

2.1.9. Metadane dokumentów cyfrowych i formaty ich zapisu

Metadane to dane opisujące inne dane, ich strukturę, atrybuty, modyfikację [Freedman, 2004, s. 456]. Stanowią ustrukturyzowaną informację związaną z obiektem w celu umożliwienia jego wyszukania, opisu, użytkowania, administrowania i zarządzania [Nahotko, 2006, s. 184]. W przypadku archiwów cyfrowych metadane stanowią bardzo ważny komponent obiektów archiwalnych, umożliwiający zapew-

nienie ich funkcjonalności [Ross, 2003, s. 32]. Zdaniem specjalistów w każdej strategii konserwacji zasobów cyfrowych metadane odgrywają zasadniczą rolę, ponieważ są jedynym sposobem uchwycenia kontekstu archiwizowanych dokumentów [Ross, 2003, s. 31], czyli wszelkich informacji o twórcy, pochodzeniu, procesie powstawania, także o czasie i celu powstania, o dotychczasowej historii, warunkach dostępu, sposobach użytkowania i wszelkich powiązaniach archiwizowanego dokumentu z innymi komponentami pozostającymi w archiwum bądź poza nim. Metadane powinny wspomóc procesy migracji rekordów przez kolejne generacje sprzętu komputerowego i oprogramowania, umożliwiać rekonstrukcję procesu decyzyjnego dotyczącego prac na obiektach cyfrowych, dostarczyć rejestr kontroli rekordu przez cały cykl jego życia [Lupovici i Masanès, 2000, s. 3-4; Ross, 2003, s. 32-33].

Metadane zwykle ujmowane są w następujące grupy: opisowe, techniczne, administracyjne, strukturalne oraz prawne (niekiedy określane również jako użytkowe) [Fülle i Ott, 2006; Januszko-Szakiel, 2006, s. 141; Metadata, 2010].

Metadane opisowe dokumentów cyfrowych są odpowiednikiem klasycznych danych bibliograficznych. Mają za zadanie informować o autorstwie, dacie opublikowania dokumentu, wydawcy czy ewentualnych dokumentach towarzyszących. Zwykle ich podstawowa funkcja realizuje się w procesach wyszukiwania, toteż niekiedy bywają określane jako metadane wyszukiwawcze [Ross, 2003, s. 32].

Techniczne metadane z kolei udzielają informacji przede wszystkim o platformie sprzętowo-programowej potrzebnej do odczytu i prezentacji treści dokumentu. Zawierają również opis zastosowanego formatu zapisu publikacji wraz z informacjami o dokumentacji formatu; informują o zastosowanym nośniku danych. Metadane techniczne bywają określane jako konserwacyjne, gdyż zawierają informacje o planowanych bądź przeprowadzonych dotychczas pracach konserwatorskich na publikacjach [Ross, 2003, s. 32]. Dodatkowo mogą być uzupełnione o dokumentację sporządzaną podczas takich prac bądź do niej odsyłać. Techniczne metadane są też istotnym czynnikiem umożliwiającym automatyzację określonych procesów zarządzania i prac konserwatorskich na obiektach archiwalnych. W dużej mierze przyczyniają się do realizacji procesów migracji i emulacji. Mogą też dostarczać dokładny opis zastosowanego identyfikatora trwałego do obiektu archiwalnego [Neuroth i in., 2009].

Kolejny rodzaj metadanych – nazywany administracyjnymi – w najogólniejszym ujęciu dotyczy zarządzania publikacjami cyfrowymi w archiwum. Znajdują się w nich np. informacje o istniejących wersjach określonej publikacji czy o sporządzanych kopiach archiwizowanych publikacji. Ważnym elementem są informacje o takich parametrach publikacji jak przykładowo: integralność, autentyczność, wyniki sporządzania sum kontrolnych [Fülle i Ott, 2006, s. 38]. Metadane administracyjne informują również o uprawnieniach poszczególnych pracowników archiwum do wykonywania określonych czynności na obiektach archiwalnych.

Metadane strukturalne opisują relacje pomiędzy obiektami (np. między obrazami graficznymi poszczególnych stron a książką jako całością) lub między pojedynczym obiektem muzealnym a kolekcją. Mają zatem istotne znaczenie dla jednoznacznego przyporządkowania każdego elementu składowego archiwizowanego obiektu, tak aby poprawnie tworzyły spójną całość [Płoszajski, 2009, s. 102-103].

Z kolei metadane prawne, nazywane także użytkowymi, informują o zakresie użytkowania archiwizowanych zasobów – niekiedy ograniczonym, z racji przepisów prawa autorskiego. Prawne metadane dotyczą warunków udostępniania, powielania, przekazywania archiwalnych obiektów [Fülle i Ott, 2006, s. 39].

Istnieje wiele formatów metadanych dla dokumentów cyfrowych [Day, 2002]. Najczęściej stosowane i przyjęte w skali światowej jako normatywy to Dublin Core DC [Dublin Core, 2017; Nahotko, 2006, s. 177; Radwański, 2005, s. 105], MARC [Marc Standards, 2017] oraz TEI [TEI, b.d.]. Jednak z racji trendu unifikowania rozwiązań stosowanych w archiwizacji zasobów cyfrowych, poczyniono starania na rzecz standaryzacji formatu metadanych. W 2001 r. w ramach amerykańskiej inicjatywy Digital Library Federation został opracowany standard METS, który ma stanowić pewnego rodzaju szkielet metadanych dla pakietów informacyjnych (ang. *Information Packages*) zgodnych ze standardem OAIS. METS ma umożliwiać dostosowanie rozmaitych danych, metadanych i wszelkich innych informacji zapisanych w różnych formatach, do postaci zgodnej i kompatybilnej z założeniami przyjętymi w OAIS. METS rozpowszechnił się i jest często stosowanym formatem metadanych w dziedzinie archiwistyki na całym świecie, szczególnie dobrze sprawdza się jako format opisu czasopism elektronicznych [Fülle i Ott, 2006, s. 58, 90; METS, b.d.]. Nieustanny rozwój formatu przejawia się w pojawiających się co jakiś czas kolejnych wersjach. W formacie METS opis dokumentu cyfrowego został ujęty w cztery grupy: metadane opisowe, techniczne, administracyjne oraz strukturalne [Fülle i Ott, 2006, s. 58, 90; METS, b.d.].

Przy formatach opisu dokumentów cyfrowych warto także wspomnieć o standardzie MPEG-21 służącym do opisu i wymiany dokumentów multimedialnych [ISO, 2006a; Fülle i Ott, 2006, s. 74]. MPEG-21 umożliwia zebranie danych technicznych, organizacyjnych oraz prawnych, istotnych szczególnie w procesach wymiany multimediów. Format MPEG-21 jest standardem ISO (ISO/IEC-Norm 21000) [ISO, 2006a]. Główną koncepcją twórców było umożliwienie użytkownikom tworzenia tzw. Digital Item, czyli odpowiednika Pakietu Informacyjnego w rozumieniu standardu OAIS. Pod pojęciem użytkownika w założeniach formatu MPEG-21 rozumie się zarówno konsumentów, jak i dystrybutorów dokumentów multimedialnych. W ramach MPEG-21 mają oni do dyspozycji tzw. sześciofunkcyjną obsługę. Mogą samodzielnie tworzyć, modyfikować, oferować, użytkować,

wyszukiwać oraz dostosowywać pakiet informacyjny. Dostosowanie oznacza, że format daje możliwość uwzględnienia indywidualnych potrzeb, oczekiwań oraz wymagań ze strony użytkowników, sprzętu i otoczenia sieciowego. Ponadto w MPEG-21 uwzględnia się metadane opisowe, techniczne oraz prawne. Jest zatem nie tylko formatem opisu dokumentów multimedialnych, ale również narzędziem, które umożliwia interakcję użytkowników z techniczną infrastrukturą. Dodatkowo zawiera metadane definiujące profil użytkownika, sprzętu technicznego i kanału transferu dokumentu. To format otwarty, czyli zdolny do przyjęcia kolejnych segmentów metadanych pochodzących z innych formatów [Fülle i Ott, 2006, s. 75]. MPEG-21 jest przydatny zarówno w procesach archiwizacyjnych, jak i w procesach opisu oraz wymiany multimedialnych dokumentów sieciowych pomiędzy archiwami bądź archiwami a użytkownikami, np. w inicjatywach zdalnego nauczania, a także pomiędzy wydawcami a instytucjami odbiorczymi, czyli księgarniami i bibliotekami. Podobną funkcję pełnią rozwiązania DID (Digital Item Declaration) [ISO, 2006a; MPEG-21, 2001], IMS Content Packaging Specification [Content Packaging, b.d.], CCSDS Packaging Standard [Recommended Standards CCSDS, b.d.], ONIX [ONIX, 2014].

2.1.10. Trwałe identyfikowanie dokumentów w archiwach cyfrowych

Istotnym atrybutem archiwów cyfrowych jest m.in. jednoznaczna identyfikacja przechowywanych w nich obiektów. Dostępność oraz czytelność obiektów archiwalnych powinna być zagwarantowana pomimo wszelkich technologicznych i organizacyjnych zmian, m.in. poprzez jednoznaczne adresowanie i identyfikowanie. W przypadku dokumentów tradycyjnych (np. opublikowanych przez wydawców) system identyfikowania jest powszechnie znany. Polega na przydzielaniu publikacjom znormalizowanych, jednoznacznych i niepowtarzalnych numerów: ISBN, ISSN, ISAN bądź ISMN[3].

Podobne systemy identyfikacyjne są stosowane dla obiektów cyfrowych. W celu dotarcia do dokumentów opublikowanych w Internecie najczęściej korzysta się

[3] ISBN – International Standard Book Number (Międzynarodowy Znormalizowany Numer Książki), ISSN – International Standard Serial Number (Międzynarodowy Znormalizowany Numer Wydawnictwa Ciągłego), ISAN – International Standard Audiovisual Number (Międzynarodowy Znormalizowany Numer Utworów Audiowizualnych), ISMN – International Standard Music Number (Międzynarodowy Znormalizowany Numer Druku Muzycznego). Szczegółowe informacje o identyfikatorach dokumentów zamieszcza w swoim serwisie Biblioteka Narodowa: http://www.bn.org.pl/index.php [Dostęp: 10.07.2017].

z adresów URL (Uniform Resource Locators), które umożliwiają wyszukanie dokumentu oraz służą jako identyfikator w procesach cytowania i bibliograficznych odesłań do dokumentów internetowych. Mogą być również stosowane w bazach danych, katalogach, indeksach, rejestrach i wszelkich innych typach bibliograficznych wykazów odsyłających do pełnych tekstów dokumentów internetowych bądź ich metadanych. Jednak zmiana miejsca dokumentu sieciowego powoduje, że zastosowane odesłanie w postaci URL jest nieużyteczne – obiekt cyfrowy przestaje spełniać podstawowe kryterium dostępności. W związku z tym powszechnie stosowany URL nie powinien być określany mianem identyfikatora, lecz raczej „lokalizatora" obiektu sieciowego, ponieważ wskazuje jedynie jego lokalizację, a nie identyfikuje jednoznacznie samego obiektu.

Połowicznym rozwiązaniem jest stosowanie metod zapewniających tzw. stabilność okresową obiektów cyfrowych. Do metod tych zalicza się:

- zastosowanie systemu adresowania URL, w którym serwer dynamicznie ustala miejsce zapisu obiektu sieciowego, korzystając z odpowiednich skryptów oraz baz danych zawierających bieżącą lokalizację dokumentów;
- zastosowanie odpowiedniej konfiguracji serwera Web, która umożliwi przekierowanie z nieaktualnego do nowego adresu w formie tzw. redirects lub aliases;
- przeprowadzanie okresowej kontroli dostępności adresów i powiązanych z nimi obiektów (ang. *URL-Checks*) przez administratora i wykonanie uaktualnienia odwołań do dokumentów [Persistent Identifier, 2008; Schöning-Walter, 2008, s. 32-38].

Zreferowana wcześniej metodologia stanowi jednak rozwiązanie krótko- lub średniookresowe, z pewnością nietrwałe. Dzieje się tak z wielu powodów, głównie z uwagi na bardzo prawdopodobne zmiany w metodologii adresowania, wynikające np. z technicznych modyfikacji otoczenia systemowego. Za sensowne uznaje się okresową kontrolę URL, jednak tylko przy założeniu „konsekwentnej pielęgnacji", która oznacza, że w przypadku stwierdzenia, iż hiperłącze nie odsyła do pożądanego obiektu, należy ustalić źródło błędu, odszukać właściwy adres do obiektu i nanieść stosowne zmiany we wszelkich wykazach, katalogach, bibliografiach, portalach i innych, które do danego obiektu odsyłają. Są to, co jasne, zabiegi pracochłonne. Okresową niedostępność adresów URL mogą też powodować błędy sieciowe lub niestabilne połączenia z serwerem. Wreszcie dokumenty sieciowe ulegają zmianom w wyniku procesów zachodzących w instytucjach, w których są zlokalizowane, toteż ich identyfikacja oraz adresowanie za pomocą samego URL mogą okazać się zawodne [Persistent Identifier, 2008].

W związku z tym zachodzi potrzeba zastosowania trwałego mechanizmu archiwizacji obiektów cyfrowych. Zaproponowane rozwiązanie to identyfikatory trwałe (ang. *persistent identifiers*) (PI).

Identyfikator trwały (PI) to niezmienna (określana też jako stabilna, unikalna, permanentna) nazwa, którą przyporządkowuje się do obiektu sieciowego jeden raz na cały cykl jego istnienia i funkcjonowania. Zadaniem PI jest jednoznaczna i trwała identyfikacja obiektu cyfrowego oraz przynależnych do niego metadanych, niezależnie od miejsca (instytucji), w którym obiekt został zapisany i jest archiwizowany, z uwzględnieniem różnorodnych systemów – ich ograniczeń (granic), zmian – oraz w obliczu występowania obiektów cyfrowych w różnych wersjach, postaciach, formach reprezentacji [Schöning-Walter, 2009]. Na podstawie PI system powinien umożliwić zlokalizowanie dokumentu i jego odczyt. Obecnie stosowane są głównie trzy systemy PI, tj.: PURL, Handle System i URN. Bez względu na wybór zastosowanego systemu, ważne jest, aby identyfikatory pozostawały niezmienne. Istotne jest także, by dany system obsługi PI miał podbudowę instytucjonalną.

PURL – Persistent URL

System Persistent Uniform Resource Locator (PURL) jest rozwinięciem koncepcji URL i funkcjonalnie pozostaje z nim tożsamy. Wykorzystuje adresy URL, które zamiast wskazywać na określony obiekt, wskazują na usługę przekierowującą do danego obiektu. Tym samym PURL składa się z adresu serwera usługi przekierowującej oraz identyfikatora obiektu, do którego chcemy uzyskać dostęp. Adresy PURL stosuje się wówczas, gdy przewiduje się częste zmiany położenia poszczególnych obiektów WWW. Pełnią one rolę oficjalnych adresów, pod którymi można znaleźć żądane zasoby, a odpowiednimi przekierowaniami zajmuje się serwer [Freedman, 2004, s. 667]. Baza danych serwera usługi przekierowań zawiera wszystkie identyfikatory zarejestrowane w danym systemie wraz z przypisanymi im aktualnymi lokalizacjami dokumentu. Można więc powiedzieć, że w systemie tym odróżnia się *identyfikatory* od *lokalizatorów*, czyli adresów lokalizacji, w których przechowywane są kopie danego obiektu. W przypadku gdy mamy do czynienia z obiektem sieciowym, lokalizatory mają postać aktualnych URL poszczególnych kopii obiektu.

System PURL został wprowadzony przez Online Computer Library Center (OCLC) w 1995 r. w ramach inicjatywy Internet Cataloging Projects, której celem było poprawienie (dookreślenie, uściślenie) adresów internetowych zasobów wykazywanych w katalogach bibliotecznych.

Składnia adresu PURL wygląda następująco: <Protocol><RA><Name>, przy czym: Protocol to standardowy protokół, np.: http; RA to adres serwera usługi przekierowującej do wybranego obiektu; Name to nazwa wskazująca na określony obiekt.

Przykład: http://purl.oclc.org/keith/home
gdzie:
http – protokół
purl.oclc.org – adres serwera przekierowującego
/keith/home – nazwa zasobu

System PURL znalazł zastosowanie m.in. w Bibliotece Kongresu oraz United States Goverment Printing Office (GPO), eksperymentalnie również w OCLC [Schröder, 2009]. Obecnie nie jest już rozwijany, natomiast zasady jego działania posłużyły przy opracowywaniu bardziej kompleksowych systemów – takich jak Handle System i URN.

Najszerzej stosowaną implementacją założeń systemu PURL jest Archival Resource Key – ARK [CDlib, 2008]. Stanowi on schemat identyfikacyjny służący do trwałej dostępności cyfrowych obiektów. Identyfikator ARK jest stosowany jako link odsyłający od obiektu cyfrowego do organizacji (do której obiekt przynależy) czy łączący obiekt cyfrowy z jego metadanymi, odsyłający do treści obiektu bądź jego kopii.

System ARK znalazł zastosowanie w 15 repozytoriach, m.in. w: California Digital Library, Library of Congress, National Library of France.

Trwałość w tym systemie identyfikacyjnym jest zapewniana przez usługodawcę, a nie składnię nazwy. ARK wskazuje metadane o obiekcie, nie daje gwarancji trwałości identyfikatora, zezwala na zintegrowanie z innymi schematami.

Składnia ARK jest następująca: http://<NMAH>/ark:/<NAAN>/<Name>, przy czym: NMAH to adres serwera usługi przekierowującej; NAAN to identyfikator instytucji nadającej poszczególnym obiektom identyfikatory we własnej przestrzeni nazw; Name to nazwa (identyfikator) przydzielona do danego zasobu.

Przykład: http://bnf.fr/ark:/13030/tf5p30086k

Handle System

Handle System jest systemem identyfikatorów przypisywanych obiektom cyfrowym niezależnie od ich fizycznego umiejscowienia. Założenia systemu zostały opracowane przez Corporation for National Research Initiatives CNRI[4] i opisane w dokumencie RFC 3650 [Handle System, 2017; Sun i in., 2003].

[4] CNRI – to amerykańska organizacja non profit, założona w 1986 r., której głównym celem jest wspieranie rozwoju kluczowych technologii przetwarzania i udostępniania wiedzy z użyciem sieci komputerowych. Zob.: *Corporation for National Research Initiatives* [online]. CNRI. Dostępny w WWW: http://www.cnri.reston.va.us/about_cnri.html [Dostęp: 10.07.2017].

Autorzy zdefiniowali m.in. zasadę budowy identyfikatorów, na które składa się prefiks oraz sufiks. Prefiks jest numerycznym kodem oznaczającym instytucję, która została zarejestrowana w Global Handle Service (instytucji nadzorującej system) jako upoważniona do nadawania obiektom identyfikatorów we własnej przestrzeni nazw. Sufiks identyfikatora jest nazwą (identyfikatorem) danego obiektu, unikatową w przestrzeni nazw danej instytucji i może składać się z dowolnej liczby znaków zgodnych z systemem ASCII.

Składnia Handle System wygląda następująco: Handle: <HNA> <HLN>, przy czym: HNA – prefix instytucji nadawany przez Global Handle Service; HNL – identyfikator obiektu w przestrzeni nazw danej instytucji.

W momencie rejestracji obiekt otrzymuje identyfikator, do którego przypisane są informacje uzupełniające. Handle System nie narzuca sztywnej struktury metadanych powiązanych z obiektem, więc zarówno rodzaj, jak i zakres tych informacji determinowany jest przez instytucję rejestrującą oraz typ obiektu cyfrowego. Wśród informacji o obiekcie najczęściej znajdują się dane właściciela (autora), opis dokumentu (tytuł, słowa kluczowe) oraz co najmniej jeden wpis pozwalający na dostęp do kopii danego obiektu.

Identyfikatory wraz z powiązanymi metadanymi przechowywane są w centralnej, ogólnodostępnej bazie danych umożliwiającej szybkie uzyskanie podstawowych informacji na temat określonych obiektów poprzez usługi dostępne w sieciach komputerowych. Funkcje systemu umożliwiają jednostkom rejestrującym dystrybucję, administrację oraz rozwiązywanie (likwidację) identyfikatorów.

Z Handle System korzysta obecnie wiele instytucji i firm. Przykładem zastosowania Handle System są m.in. CORDA/ADL i DVIA, czyli systemy Departamentu Obrony Stanów Zjednoczonych, rejestrujące i zarządzające dokumentami związanymi z obronnością Stanów Zjednoczonych. Handle System jest również użyteczny w projekcie DSpace realizowanym przez MIT, w ramach którego tworzona jest baza danych na temat materiałów edukacyjnych powstających we wszystkich wydziałach i jednostkach tej instytucji. Kolejnym projektem stosującym opisywany system jest The National Digital Library – program, którego założeniem jest digitalizacja i utworzenie bazy danych dzieł zgromadzonych w bibliotekach publicznych i uczelnianych w Stanach Zjednoczonych [Handle, b.d.].

Struktura identyfikatorów Handle System pozwala na rejestrację tzw. rejestratorów lokalnych. Zarejestrowana instytucja ma możliwość rejestracji instytucji sobie podległych, które dysponują własną przestrzenią nazw dla obiektów cyfrowych. W takim przypadku prefiks identyfikatora składa się z dwóch numerycznych członów oddzielonych kropką (np. 10.1000), przy czym pierwszy człon określa instytucję nadrzędną zarejestrowaną przez Global Handle Service, natomiast drugi jest identyfikatorem lokalnego rejestratora. Drzewiasta struktura

Handle System pozwoliła na powstanie podsystemów identyfikacyjnych, z których najpopularniejszym jest Digital Object Identifier (DOI).

DOI to identyfikator dokumentu elektronicznego, który jest do niego na stałe przypisany i – w odróżnieniu od identyfikatora URL – nie zależy od fizycznej lokalizacji dokumentu. Digital Object Identifier to rozwiązanie pozwalające na przydzielanie dokumentom, publikacjom i wszelkim innym zasobom dostępnym w Internecie stałych, niezmiennych nazw zamiast adresów URL [Freedman, 2004, s. 143]. Podstawowym założeniem systemu DOI jest identyfikacja oraz wymiana obiektów cyfrowych. Trwają również prace nad organizacyjnymi oraz technicznymi rozwiązaniami, umożliwiającymi zarządzanie obiektami cyfrowymi oraz powiązanie producentów i dostawców obiektów z użytkownikami [DOI System, 2017].

Zarządzaniem systemem identyfikatorów DOI zajmuje się Międzynarodowa Fundacja DOI (International DOI Foundation IDF), która jest organizacją non profit, finansowaną ze składek członkowskich oraz sprzedaży prefiksów i numerów DOI. Fundacja DOI sprawuje kontrolę nad instytucjami i firmami, które uzyskały prawo do pełnienia roli agencji rejestracyjnych DOI (DOI Registration Agency, RA). Podstawowym zadaniem tych agencji jest przydzielanie identyfikatorów wydawcom (Publisher ID) i zapewnienie im infrastruktury umożliwiającej tworzenie identyfikatorów obiektów (Item ID) oraz zarządzanie metadanymi przypisanymi identyfikatorom DOI. Od RA oczekuje się promocji systemu DOI oraz współpracy na rzecz jego rozwoju.

Struktura DOI stanowi od 2001 r. standard ANSI/NISO (Z39.84), a jej komponenty są implementacją założeń Handle System. System DOI składa się z następujących elementów: metadane, DOI jako identyfikator trwały (PI) oraz techniczna implementacja Handle System. Identyfikatory DOI zgodnie z założeniami Handle System są ciągami znaków ASCII. Składają się z przedrostka i końcówki.

Przykład: 10.1000/182, przy czym: 10.1000 to przedrostek, w którym znaki 10 informują, że chodzi o identyfikator DOI; 1000 to numer przypisany przez IDF wydawcy (Publisher ID); natomiast suffix 182 to końcówka, która jest przypisana do określonego dzieła (Item ID).

Publisher ID jest przypisywany wydawcom, którzy zdecydowali się zarejestrować i korzystać z systemu DOI przez agencję posiadającą do tego prawo. Item ID jest nadawany przez samego wydawcę, który powinien zagwarantować, że ID będzie unikalne dla każdej wydanej przez niego publikacji. Item ID może być, ale nie musi, numerem katalogowym publikacji pochodzącym z innych systemów rejestrowania, np.: ISBN, ISSN. Poprawny sposób podawania odnośników do źródeł wygląda następująco: doi: 10.1000/182.

System DOI jest stosowany, m.in.: w agencjach praw autorskich, wydawnictwach i bibliotekach. Typowym przykładem zastosowania DOI jest identyfikowanie elek-

tronicznych wersji publikacji naukowych, np. w repozytorium SpringerLink. Identyfikator DOI może otrzymać artykuł, całe czasopismo naukowe, rozdział w książce, plik multimedialny, program komputerowy i inne.

URN – Uniform Resource Name

Historia systemu URN rozpoczęła się w 1990 r. i ma związek z projektowaniem architektury World Wide Web (WWW). URN został wprowadzony jako ujednolicona forma oznaczania zasobów internetowych. Formy i kierunki rozwoju sieci Internet są kontrolowane przez organizację Internet Assigned Numbers Authority (IANA) [IANA, b.d.]. To właśnie IANA oraz ściśle związana z nią grupa robocza o nazwie Internet Engineering Task Force (IETF) stanowią siłę napędową rozwoju Internetu i de facto dyktują standardy, spośród których najbardziej znane są publikacje pod tytułem *Requests for Comments* (RFCs). W dokumencie RFC 1737 z 1994 r. dość precyzyjnie określono wymagania dotyczące schematu URN [Sollins i Masinter, 1994], natomiast trzy lata później, w publikacji RFC 2141 z 1997 r., zostały wymienione cele rozwoju identyfikatorów trwałych PI [Moats, 1997].

System URN został świadomie pomyślany jako schemat otwarty, zdolny do integracji z systemami istniejącymi – choćby z identyfikatorami ISBN albo URL. Od ponad 10 lat URN funkcjonuje jako standard adresowania obiektów w instytucjach objętych obowiązkiem takiego identyfikowania zasobów, aby były one dostępne trwale oraz niezależnie od tego, w której instytucji są przechowywane [Uniform Resource, 1996].

System URN cieszy się dużą popularnością. Przykładowo jest stosowany w narodowych bibliotekach krajów takich jak: Finlandia, Holandia, Austria, Szwajcaria czy Wielka Brytania. Istnieje możliwość integracji identyfikatorów URN z istniejącymi numerycznymi systemami identyfikacyjnymi dokumentów, np.: ISAN, ISSN, ISBN.

Identyfikatory URN składają się z kilku hierarchicznie ułożonych elementów, tj.: z *Namespace IDentifier NID* (tzw. identyfikatora przestrzeni nazw) oraz z podporządkowanych mu subelementów (*SNID, NSS*).

Składnia identyfikatora wygląda następująco: urn: <NID> [: SNID] : <NSS>, przy czym: NID – identyfikator przestrzeni nazw; SNID – identyfikator podprzestrzeni nazw (jeśli występuje); NSS – unikalny dla danej podprzestrzeni identyfikator zasobu (łańcuch znaków).

Jedną z podprzestrzeni nazw systemu URN jest system NBN – National Bibliographic Number. Został opracowany w celu wyszczególnienia w bibliografiach narodowych publikacji cyfrowych, np. czasopism elektronicznych, rozpraw doktorskich i habilitacyjnych, ale także innych publikacji stanowiących narodowe

dziedzictwo cyfrowe i podlegających obowiązkowi wieczystej archiwizacji. Koncepcja systemu NBN zrodziła się w ramach popularnych inicjatyw bibliotek narodowych, Conference of Directors of National Libraries (CDNL) oraz Conference of European National Librarians (CENL).

NBN jest implementacją założeń systemu URN, w związku z czym składnia jego identyfikatorów wygląda następująco: urn: NBM : <ICC> [:SNS] NBNstring, gdzie: ICC to dwuliterowy kod kraju wg ISO 3166; SNS to podprzestrzeń nazw; NBNstring to identyfikator w podanej przestrzeni nazw.

Przykład: urn:NBN:de:kobv:23-2312

System NBN jest używany wyłącznie w bibliotekach narodowych i stosowany do jednoznacznej, trwałej identyfikacji zarówno dokumentów cyfrowych, jak i fizycznych. Biblioteki narodowe przyjmują obowiązek zarządzania przestrzeniami nazw w obrębie danego kraju.

Podsumowując – składowanie i archiwizacja zasobów nauki i kultury w sieci ma sens wówczas, gdy zasoby te w każdej chwili, obecnie i w najbardziej odległej przyszłości, mogą być udostępniane i użytkowane. Instytucje tworzące archiwa cyfrowych zasobów decydują się na rozmaite systemy ich trwałego identyfikowania. Zaleca się, aby w procesie decyzyjnym, dotyczącym wyboru systemu identyfikacji, instytucje uwzględniły następujące kryteria [Persistent Identifier, 2008]:

- standaryzacja – instytucje powinny skłaniać się do stosowania systemów, które zostały zaakceptowane jako standard, najlepiej o światowym zasięgu;
- wymagania funkcjonalne – wybierane systemy identyfikacyjne powinny charakteryzować się trwałością, jednoznacznością, światowym zasięgiem, niezależnością od miejsca składowania; identyfikatory trwałe powinny odsyłać równocześnie do wielu kopii jednego obiektu;
- elastyczność, skalowalność – stosowane systemy powinny być skalowalne oraz zdolne do rozszerzenia o nowe funkcje, bez zaburzenia ich zgodności z przyjętym standardem;
- niezależność technologiczna i kompatybilność – systemy identyfikacyjne powinny być standardowe, niezależne od protokołów i technologii, a także kompatybilne z funkcjonującymi instalacjami i usługami;
- instalacje, polecenia (rekomendacje) – przy wyborze systemu należy uwzględnić jego akceptację i popularność w skali międzynarodowej;
- koszty oraz trwałość – kryterium wyboru systemu powinny być koszty systemu (zarówno wstępne, jak i dalszego utrzymania) oraz jego niezawodność.

Opisane systemy trwałej identyfikacji obiektów sieciowych w zasadzie spełniają wszystkie z wymienionych kryteriów i są najczęściej implementowane w profesjonalnych archiwach cyfrowych. Należy jednak zaznaczyć, że obok nich

istnieje także szereg innych, mniej popularnych rozwiązań. Charakterystyka ich wszystkich przekroczyłaby ramy niniejszych rozważań, ale warto choćby je wymienić. Są to: ERRoL – Extensible Repository Resources Lokator [Technology Reports, b.d.], GRI – Grid Resource Identifier [Parastatidis i in., 2003], GUUID/UUID – Globally Unique Identifier/Universal Unique Identifier [GUID, 2005], InfoURI [Sompel i in., 2006], LSID – Life Science Identifier [LabKey, b.d.], POI – PURL-Based Object Identifier [OASIS, b.d.], XRI – Extensible Resource Identifier [OpenXRI, 2009].

W Polsce tematyka jednoznacznego identyfikowania zasobów cyfrowych nie jest obca. Propozycje wprowadzania do informatycznych systemów archiwalnych zasady centralnej rejestracji dokumentów cyfrowych i nadawania im numeru ewidencyjnego, podobnego do ISBN, pojawiły się jakiś czas temu [Radwański, 2005, s. 104-105], a do praktyki weszły wraz z budową polskich bibliotek cyfrowych. Zgodnie z informacjami udostępnianymi przez twórców serwisu Federacji Bibliotek Cyfrowych, udostępnienie protokołu OAI-PMH w sieci polskich repozytoriów i bibliotek cyfrowych zaowocowało powstaniem automatycznie nadawanych, unikalnych w skali światowej identyfikatorów udostępnianych obiektów. Mechanizm identyfikatorów (OAI Identifier) zawarty jest w standardzie opisującym protokół OAI-PMH. Format identyfikatora wygląda następująco: oai:<domena repozytorium>:<identyfikator zasobu w repozytorium>, a przykładowa jego instancja ma formę: oai:www.wbc.poznan.pl:8711.

Serwis FBC posiada mechanizm rozpoznawania identyfikatorów OAI Identifier, a tym samym pozwala na uzyskanie informacji oraz aktualnego adresu obiektu cyfrowego na podstawie jego unikalnego identyfikatora. Mechanizm ten może być również przydatny do utworzenia trwałej referencji do obiektu cyfrowego np. na potrzeby odwołań w bibliografii. Referencja taka ma postać: http://fbc.pionier.net.pl/id/, a przykładowe odwołanie wygląda w ten sposób: http://fbc.pionier.net.pl/id/oai:www.wbc.poznan.pl:8711.

Otwarcie takiego adresu w przeglądarce WWW spowoduje wyświetlenie podstawowych metadanych obiektu cyfrowego o podanym identyfikatorze oraz odnośników do metadanych i/lub treści tego obiektu. Utworzona w ten sposób referencja może być trwałym i poprawnym odwołaniem do cyfrowego dokumentu, niezależnie od zmian wprowadzonych w oprogramowaniu biblioteki cyfrowej, która ten dokument udostępnia. Informacje o identyfikatorach poszczególnych obiektów cyfrowych, które można użyć do tworzenia referencji, udostępniane są zazwyczaj przez poszczególne repozytoria na stronach z opisem obiektów [Lewandowska i in. 2007].

2.2. Ekonomiczne zagadnienia trwałej ochrony zasobów cyfrowych

Organizowanie prac związanych z ochroną zasobów cyfrowych wymaga jednoznacznego ustalenia źródeł ich finansowania. Dużym ułatwieniem byłby posiadający moc prawną dokument, określający podstawy długoterminowego finansowania działań archiwizacyjnych. Koniecznym jest również, choćby orientacyjne, wskazanie rozmiaru kosztów, z którymi należy liczyć się, podejmując takie działania.

Kalkulowanie kosztów długoterminowej ochrony zasobu cyfrowego określonej objętości jest wciąż kwestią otwartą. Próby takie już podejmowano, a ich przykładem jest The LIFE Project realizowany w Wielkiej Brytanii w latach 2005-2006 przez The British Library oraz University College London – z głównym założeniem opracowania modelu zarządzania kosztami długoterminowej archiwizacji zasobów cyfrowych [LIFE, 2010]. W wyniku projektu stworzono wprawdzie formułę szacowania kosztów archiwizacji, jednak sporo kwestii budzi w niej wątpliwości. Zaproponowana formuła ma status dyskusyjnej i otwartej: $L_T = A_q + + I_T + M_T + Ac_T + S_T + P_T$, gdzie:

- L – Complete lifecycle cost over time (koszt całego cyklu „życia" publikacji cyfrowej w czasie 0 do T)
- Aq – Acquisition (gromadzenie)
- I – Ingest (wprowadzenie, przyjęcie do archiwum)
- M – Metadata (metadane)
- Ac – Access (udostępnianie)
- S – Storage (przechowywanie)
- P – Preservation (archiwizacja, ochrona).

Każdy z wymienionych w formule parametrów może zostać podzielony na bardziej szczegółowe elementy odzwierciedlające indywidualne wymagania instytucji. W ten sposób istnieje możliwość dopasowania formuły do specyficznych potrzeb, działań i zadań poszczególnych instytucji archiwizujących.

Podobna formuła pojawiła się także w wyniku innego przedsięwzięcia – w ramach projektu narodowej biblioteki Holandii oraz firmy IBM, dotyczącego budowy repozytorium cyfrowego dla zasobu Holandii. Próbowano wówczas szacować koszty zastosowania dwóch podstawowych metod archiwizacji zasobów cyfrowych, tj. emulacji i migracji [Oltmans, 2004, s. 380-392; Oltmans i Kol, 2005]. Autorzy obliczeń odwołali się do formuły obliczania kosztów archiwizacji ma-

teriałów niecyfrowych [Shenton, 2003, s. 254-272] i podjęli próbę zastosowania jej do obliczenia kosztów archiwizacji materiałów cyfrowych. Formuła składa się z następujących parametrów: K(t,a) = s(a) + i(a) + h(t,a), gdzie:

- K(t,a) – (total cost of holding *a* objects for a period of *t* years) całkowity koszt utrzymania (a) obiektów cyfrowych w okresie (t) czasu
- s – (selection) selekcja
- i – (ingest) wprowadzenie (przyjęcie do archiwum)
- h – (storage) przechowywanie.

Ta formuła zastosowana dla obliczenia kosztów migracji ma następującą postać: K(t,a) = h(t,a) + m(t,a), gdzie:

- K(t,a) – całkowity koszt utrzymania (a) obiektów cyfrowych w okresie (t) czasu
- h – koszt przechowywania
- m – koszt migracji.

Autorzy formuły zakładają, że koszty migracji cyfrowych obiektów są zależne od czasu przechowywania *t* (im dłużej obiekt będzie przechowywany w archiwum, tym więcej razy będzie poddawany procesowi migracji) oraz od ilości archiwizowanych obiektów *a* (im więcej obiektów archiwalnych, tym więcej akcji migrowania należy dokonać).

Ta sama formuła zastosowana dla obliczenia kosztów emulacji wygląda następująco: K(t,a) = h(t,a) + E + e(t), gdzie:

- K(t,a) – całkowity koszt utrzymania (a) obiektów cyfrowych w okresie (t) czasu
- h – koszt przechowywania
- E – koszt opracowania i wykonania urządzenia emulującego
- e – koszty utrzymania urządzenia emulującego.

Na podstawie założonych wartości poszczególnych kosztów autorzy obliczeń wykazali, że zastosowanie metody emulacji w długim okresie jest bardziej efektywne kosztowo, niż zastosowanie strategii migracji. Należy jednak zauważyć, że strategia opierająca się na emulacji pierwotnego środowiska programowo-sprzętowego skutkuje ograniczeniem dostępu do zgromadzonych zasobów jedynie do stanowisk obsługujących emulację bezpośrednio w określonej instytucji. Trudne do zrealizowania i raczej niemożliwe jest udostępnienie „na zewnątrz" dokumentu odczytanego metodą emulacji. Decyzja o zastosowaniu tej strategii pociąga za sobą także konieczność poniesienia wysokich kosztów związanych z opracowaniem i wykonaniem narzędzi emulujących.

Z danych publikowanych przez biblioteki narodowe Holandii i Niemiec wiadomo, że budowa emulatora jest bardzo droga. Na zlecenie narodowej biblioteki Holandii firma IBM zaimplementowała w 2004 r. demonstracyjną wersję urządzenia

emulującego. Jego projekt i budowa wraz z potrzebnymi badaniami oraz testami pochłonęła 32 tygodnie czterdziestogodzinnej pracy, przy stawce godzinowej w wysokości 120 USD, czyli łącznie ok. 150 tys. USD. Koszty implementacji emulatora obliczono na 200 tys. USD. Są to koszty wersji demonstracyjnej.

O kosztach początkowych archiwizacji można dowiedzieć się również ze źródeł niemieckojęzycznych. Niemiecka biblioteka narodowa, podobnie jak biblioteka Holandii postanowiła zbudować system archiwalny dla niemieckiego zasobu cyfrowego. Pierwszym krokiem był projekt Kopal [Kopal, b.d.] realizowany w latach 2004-2007, w celu opracowania organizacyjnej i technicznej strategii archiwizacji zasobu cyfrowego Niemiec. W efekcie powstał system archiwizacji zasobów cyfrowych, którego koszt wyniósł 4,2 mln euro.

Przedstawione dotychczas dane są oczywiście wyrywkowe i tylko w ograniczonym zakresie pozwalają zorientować się, z jakiego rzędu kosztami powinny liczyć się instytucje pamięci. Podobnie jak szereg aspektów natury organizacyjnej, technicznej i prawnej, tak i kwestia prognozowania kosztów zabezpieczenia i konserwacji zasobów cyfrowych pozostaje obszarem wymagającym badań.

Na podstawie opracowań przedmiotu oraz dokumentacji zrealizowanych projektów [Beucke, 2010, s. 26-37; Kopal, b.d.; National Library of Australia, 2003, s. 66-67; Neuroth i in., 2009] możliwe jest wymienienie jedynie podstawowych grup kosztów w archiwizacji:

1. Koszty początkowe:
   - pozyskanie wiedzy o stosowanych rozwiązaniach w archiwistyce cyfrowej;
   - utworzenie stanowisk pracy;
   - pozyskanie wyposażenia (głównie technicznego) dla nowych stanowisk pracy;
   - planowanie, konsultowanie, pisanie, ogłaszanie przetargów na projekty.
2. Koszty zaopatrzenia:
   - zaprojektowanie i zbudowanie systemu depozytowego wraz z potrzebną infrastrukturą;
   - zaprojektowanie i stworzenie stosownego oprogramowania bądź opłaty licencyjne na system zarządzania archiwum; możliwe jest również skorzystanie z oprogramowania typu Open Source Software, ale wówczas kosztowne może okazać się jego dostosowanie do zakładanych potrzeb;
   - zaprojektowanie i stworzenie odpowiednich środków zabezpieczenia dostępu do zasobów cyfrowych;
   - zatrudnienie nowych pracowników (specjalistów) bądź szkolenie, kształcenie i dokształcanie personelu już zatrudnionego.
3. Koszty bieżącego działania:
   - wprowadzenie zgromadzonych już zasobów cyfrowych do archiwum;
   - przyjmowanie i wprowadzanie do archiwum nowych zasobów cyfrowych;

- bieżące koszty działania archiwum (np.: energia, sporządzanie kopii zapasowych, regularna kontrola dostępności zasobów, środki bezpieczeństwa, ochrony archiwum i jego zasobów);
- utrzymanie, konserwacja i zwiększanie pojemności repozytorium;
- bieżące koszty licencyjne, ewentualne składki z racji przynależności do określonych konsorcjów, ugrupowań, organizacji oraz inne.

Koszty długoterminowej archiwizacji stanowiły przedmiot badań pn. „Keeping Research Data Save". Celem było oszacowanie procentowego udziału kosztów procesu długoterminowej archiwizacji danych cyfrowych pochodzących z badań naukowych. Badaczki ustaliły, że największy udział w kosztach archiwizacji zajmują czynności wytworzenia i przyjęcia dokumentu do archiwum (42%). Na poziomie 35% kosztów kształtuje się zapewnienie dostępu do dokumentu. Najtańsze są ochrona treści i kodu zero-jedynkowego zgromadzonego zasobu cyfrowego (23%) [Beagrie i in., 2008, s. 5].

Koszty programów ochrony dziedzictwa cyfrowego, zwłaszcza długoterminowych, trudno oceniać z powodu mnogości czynników niepewnych, takich jak: przyszłe zmiany w technologiach bądź bardzo długie horyzonty czasowe działań. Należy jednak liczyć się z tym, że koszty organizacji i utrzymania archiwum cyfrowego są wysokie [National Library of Australia, 2003, s. 34] i będą kształtować się rozmaicie w zależności od specyfiki i celów działania poszczególnych archiwów – przede wszystkim od ilości, różnorodności i stanu technicznego cyfrowej kolekcji archiwalnej, a także od oczekiwań deponentów i użytkowników oraz zobowiązań archiwów wobec nich [Neuroth i in., 2009].

Koszty ochrony zasobów cyfrowych są wysokie, ale prawdopodobne koszty konsekwencji braku tychże działań są jeszcze większe. Pierwsze szacunki w tym zakresie pochodzą z 1995 r. i zostały zaprezentowane podczas spotkania grupy roboczej ISO Archiving Standards. Podano wówczas do wiadomości, że roczny koszt zachowania jednego megabajta zapisów elektronicznych stworzonych w sektorze inżynieryjnym wynosi 5-7 USD, ale aż 1250 USD kosztowałoby ich odtworzenie, gdyby zapisy te zostały utracone lub zniszczone [Ross, 2003, s. 17]. Z nowszych doniesień wynika, że koszty przechowywania plików cyfrowych są znacznie wyższe, niż się powszechnie sądzi. Zwrócono uwagę, że „w miarę wzrostu ilości przechowywanych informacji szybciej rośnie obciążenie ekonomiczne. Fakt, że pojemność nośników co roku się podwaja, powoduje błędne przekonanie, iż ceny przechowywania szybko spadają. Sprawdza się to w krótkim okresie, czyli z reguły krótszym niż 5 lat, ponieważ utrzymywanie dostępu do plików wymaga niewiele wysiłku. Jednak w perspektywie długoterminowej koszty zarządzania będą stale rosnąć. Mówi o tym Jim Gray, szef Bay Area Research Center firmy Microsoft: «Prawdziwe koszty generowane są jednak przy zarządzaniu. Ludzie

z Wall Street powiedzieli mi, że zarządzając przechowywanymi danymi co roku wydają 300 tys. USD na terabajt. Na jeden terabajt mają więcej, niż jednego administratora danych. Inne firmy mówią o jednym administratorze na 10 TB, Google i Internet Archive mają jednego na 100 TB. Koszt tworzenia kopii, odtwarzania, archiwizowania, reorganizowania, powiększania i zarządzania pojemnością zdaje się przerastać koszt sprzętu. To prawdziwe wyzwanie dla speców od oprogramowania. Jeżeli przyjąć powyższe normy, to petabajt wymagałby 1000 administratorów»" [Palm, 2011].

Podsumowując, niezbędnym elementem składowym programów ochrony jest model ich finansowania, ale równie istotna jest ciągła dostępność źródeł finansowych. Menadżerowie programów powinni opracowywać długoterminowe modele działań i jednocześnie znajdować źródła finansowania z gwarancją ich płynnego następstwa w długim czasie [National Library of Australia, 2003]. Nie ma obecnie pewności, jak będą kształtować się koszty procesu długoterminowej archiwizacji w przyszłości, jednak pewne jest, że na samym początku muszą zostać wyasygnowane potężne sumy na zaprojektowanie, stworzenie i uruchomienie profesjonalnego systemu archiwalnego. Finansowanie realizowanych na świecie projektów długoterminowej archiwizacji publikacji elektronicznych odbywa się głównie z udziałem dotacji państwowych. Autorzy projektów wnioskują o ich finansowanie bądź współfinansowanie do organizacji, instytucji, fundacji wspierających badania i rozwój sektora nauki oraz kultury. Przede wszystkim dąży się do zaangażowania w tego typu projekty władz państwowych.

Problematyce szacowania kosztów i finansowania zadań trwałej ochrony zasobów cyfrowych dedykowano projekt 4C Collaboration to Clarify the Costs of Curation. Było to przedsięwzięcie o charakterze międzynarodowym, zainicjowane w 2013 r. na gruncie współpracy Digital Preservation Coalition (DPC) z Komisją Europejską w celu skutecznego inwestowania i zarządzania finansami cyfrowej ochrony dokumentów, głównie w krajach europejskich. W ramach projektu 4C podejmowano kompleksowe badania obejmujące różne finansowe aspekty zarządzania projektami archiwizacji. Zdaniem inicjatorów, przez ścisłą współpracę i wymianę doświadczeń licznych instytucji i organizacji łatwiejsze jest zrozumienie ekonomicznych uwarunkowań długoterminowej archiwizacji cyfrowej. Bardzo ważna jest skuteczność kontrolowania i zarządzania własnymi zasobami cyfrowymi w długim czasie, ale możliwe i opłacalne są również nowe rozwiązania i usługi komercyjnie świadczone dla podmiotów zainteresowanych [4C, 2013]. Projekt 4C zakończył się w 2015 r. Na jego gruncie została utworzona branżowa platforma społecznościowa o nazwie The Curation Costs Exchange (CCEx), w której funkcjonowanie są zaangażowane DPC jako założyciel i koordynator, Nestor, NCDD. Specjaliści różnych krajów i instytucji dzielą się informacjami poufnymi o plano-

wanych zadaniach i ich kosztach. W ten sposób powstaje odpowiednio zarządzana baza, której dane po przetworzeniu umożliwiają wnioskowanie i tworzenie wiedzy o wysokości kosztów poszczególnych elementów składowych trwałej archiwizacji kolekcji zasobów cyfrowych [Curation Costs, b.d.].

2.3. Prawne zagadnienia trwałej ochrony zasobów cyfrowych

Zachowanie dziedzictwa cyfrowego wymaga opracowania oraz zatwierdzenia solidnych podstaw prawnych. Ich tworzenie prawdopodobnie należałoby rozpocząć od rewizji obowiązujących w tym zakresie aktów prawnych, przede wszystkim ustaw o: bibliotekach, obowiązkowych egzemplarzach bibliotecznych, narodowym zasobie archiwalnym i archiwach, prawie autorskim i prawach pokrewnych oraz ustawy o bazach danych.

Niezbędne jest także sformułowanie szeregu dodatkowych przepisów tworzących „prawną platformę" dla wszelkich starań, planów oraz przedsięwzięć podejmowanych przez instytucje pamięci na rzecz długoterminowej archiwizacji zasobów cyfrowych. W obcojęzycznym piśmiennictwie przedmiotu zbiór takich norm prawnych jest określany mianem *preservation policy* i niekoniecznie ma formę spójnego dokumentu, lecz bywa tworzony z pojedynczych ustaw, rozporządzeń, zarządzeń, wytycznych, umów, itp. [Preservation Policy, 2009]. Zaleca się, aby narodowe programy ochrony dziedzictwa cyfrowego bazowały na jednoznacznych przepisach regulujących pobieranie, kopiowanie, modyfikowanie, przechowywanie i zapewnianie dostępu do materiałów cyfrowych [National Library of Australia, 2003, s. 32]. Tym samym w każdym kraju instytucje pamięci stoją przed zadaniem opracowania i przyjęcia pakietu zapisów o mocy prawnej, stanowiących podstawę realizacji zadań długoterminowej archiwizacji cyfrowych zasobów (ang. national preservation policy).

2.3.1. Ustawy o bibliotekach, o obowiązkowych egzemplarzach bibliotecznych, o narodowym zasobie archiwalnym i archiwach a długoterminowa archiwizacja zasobów cyfrowych

Właściwie we wszystkich krajach w czasie planowania i przystępowania przez nie do działań archiwizacyjnych obowiązujące w nich ustawy regulujące działalność instytucji pamięci nie zawierały odniesień do zasobów sieciowych (z materiałów elektronicznych uwzględnione były jedynie publikacje na cyfrowych nośnikach przenośnych) [Goebel i in., 2004]. Podobnie jest w naszym kraju. W obowiązujących w Polsce ustawach dotyczących instytucji pamięci

ustawodawca posługuje się pojęciami: materiały, dokumenty, publikacje, utwory, wśród których wymieniane są, m.in., dokumenty elektroniczne na nośnikach fizycznych [Szczepańska, 2007, s. 57; Ustawa, 1996; Ustawa, 1997] bądź informatycznych [Ustawa, 1983; Ustawa, 2005], brak natomiast precyzyjnego ujęcia dokumentów sieciowych, np.: naukowych czasopism elektronicznych publikowanych w Internecie, zasobów bibliotek cyfrowych, innych kolekcji cyfrowych dokumentów udostępnianych w sieci i wartych trwałego zachowania. W opinii znawców przedmiotu jednak biblioteki cyfrowe stanowią biblioteki w rozumieniu ustawy o bibliotekach i wszelkie przepisy szczególne dotyczące bibliotek tradycyjnych odpowiednio lub wprost proporcjonalnie będą miały zastosowanie do bibliotek cyfrowych [Barta i Markiewicz, 2004, s. 116; Stanisławska-Kloc, 2005; Szczepańska, 2007, s. 57], tym samym zapewne i do kolekcji sieciowych czasopism naukowych.

W świetle obowiązującego prawa wydawcy oraz wszelkie instytucje prowadzące działalność polegającą na wytwarzaniu dokumentów cyfrowych, stanowiących dziedzictwo nauki i kultury, nie mają obowiązku zgłaszania i odsyłania do instytucji archiwizujących publikacji sieciowych, natomiast instytucje archiwizujące nie są zobligowane do ich gromadzenia, archiwizacji oraz organizacji dostępu do nich. Inicjatywy związane z tworzeniem, ochroną i udostępnianiem kolekcji zasobów sieciowych są podejmowane i realizowane na mocy pozaustawowych postanowień. Należałoby wobec tego jednoznacznie uregulować procesy gromadzenia, archiwizacji i udostępniania zasobów sieciowych.

Szczególnie poważne wyzwania są związane z gromadzeniem i archiwizacją witryn internetowych przedstawiających nierzadko unikatową wartość dla dziedzictwa narodowego – tzw. *harvesting* bądź *webharvesting*. Wiele krajów, np. Wielka Brytania, Australia, także kraje Skandynawii i Francja, zdecydowało się na dokumentowanie i archiwizowanie stron WWW należących do ich domeny narodowej. Ustalenia wymagają nie tylko kryteria gromadzenia stron (np. na podstawie domeny albo określonej tematyki), ale przede wszystkim konieczne są podstawy prawne umożliwiające instytucjom pamięci archiwizację tego typu publikacji cyfrowych. Do czasu obowiązywania odpowiednich aktów prawnych prawdopodobnie najrozsądniejszym rozwiązaniem jest uzyskanie zgody dysponenta autorskich praw majątkowych na zarządzanie nimi w ramach procesów archiwizacji. Możliwe jest również – w zależności od umowy z dysponentem praw – wykonanie kopii archiwalnej albo okresowe sporządzanie tzw. zrzutów ekranowych strony WWW i umieszczanie ich w archiwum[5]. W 2008 r. plany archiwizacji stron WWW ogłosiła polska Biblioteka Narodowa. Podczas konferencji „Polskie Biblioteki Cyfrowe"

[5] Na podstawie informacji dostępnych na stronach bibliotek narodowych wymienionych krajów.

w Poznaniu w dniu 25 listopada 2008 r. przedstawicielka BN zwróciła uwagę na potrzebę takich działań, motywując to faktem ulotności, często bezpowrotnej utraty zasobów sieciowych o unikatowej wartości. W wypowiedzi podkreślono, że do realizacji tych działań potrzebne są podstawy prawne i wiedza. Próby gromadzenia polskich witryn WWW miały rozpocząć się w 2009 r. [Ślaska, 2009]. Z raportu MKiDN z września 2009 r. wynika, że w ramach planów archiwizowania polskiego Internetu Biblioteka Narodowa w Warszawie przyłączyła się do inicjatywy International Internet Preservation Consortium (IIPC), której celem jest wspieranie idei archiwizacji zasobów internetowych oraz budowa narzędzi i technologii archiwizacyjnych. W tym samym roku w BN zostało wdrożone oprogramowanie Heritrix i rozpoczęto pilotażowe prace przy archiwizacji domen internetowych [Program digitalizacji, 2009]. Początkowe doświadczenia BN z realizacji tego projektu wskazywały na znikomą świadomość dotyczącą znaczenia archiwizacji dokumentów sieciowych oraz takie samo zainteresowanie osób i organizacji, którym powinno zależeć na archiwizacji przestrzeni Internetu bezpośrednio ich dotyczącej. Ponadto pojawia się szereg innych problemów natury prawnej i technicznej [Ślaska i Wasilewska, 2012].

Do problematyki archiwizacji polskiego Internetu odniesiono się w treści raportu MKiDN z września 2009 r. Zwrócono uwagę, że „ulotność i zmienność zasobów internetowych powoduje konieczność systematycznej archiwizacji, dzięki której będzie można zachować dla współczesnych i przyszłych badaczy bogactwo polskiego Internetu. Obowiązujące w Polsce prawo uniemożliwia jednak gromadzenie i udostępnianie archiwizowanych zasobów internetowych bez zgody ich właścicieli, co znacznie utrudnia i wydłuża proces archiwizacji. Zmiana legislacji w tym zakresie powinna być jednym z najważniejszych celów strategicznych instytucji odpowiedzialnych za ochronę dziedzictwa cyfrowego" [Program digitalizacji, 2009, s. 43; Ślaska i Wasilewska, 2012]. W programie konferencji IFLA organizowanej we Wrocławiu w sierpniu 2017 r. zaplanowano sesję poświęconą archiwizacji Internetu. Korzystając z obecności ekspertów (w tym z Internet Archive oraz British Library) MKiDN zaplanowało organizację roboczego spotkania (25 sierpnia 2017 r., w Warszawie) poświęconego archiwizacji Internetu. Jest to inicjatywa podkreślająca znaczenie tej tematyki we współczesnym świecie oraz jej uwzględnienie w programach działania polskich instytucji[6].

Polskie ustawodawstwo dotyczące funkcjonowania instytucji pamięci wymaga zmian i uzupełnienia. Przede wszystkim potrzebna jest redefinicja pojęć: *materiały biblioteczne*, *materiały archiwalne, narodowy zasób archiwalny*, *dziedzictwo nauki i kultury*, *utwór*, *publikacja*, służąca uwzględnieniu dokumentów sieciowych.

[6] Na podstawie informacji uzyskanych w MKiDN w sierpniu 2017 r.

Ustawodawca powinien obligować nie tylko wydawców, lecz wszelkie instytucje, organizacje oraz osoby prywatne tworzące cyfrowe dokumenty (które posiadają wartość naukową) do przekazywania ich instytucjom pamięci w celu archiwizacji. Instytucje pamięci natomiast powinny opracować i opublikować procedurę standaryzowanego sposobu przekazywania dokumentów sieciowych w celu trwałej archiwizacji.

Wyznaczone instytucje pamięci w Polsce na mocy obowiązujących aktów prawnych są zobligowane do szczególnej ochrony i wieczystego archiwizowania zasobu. Potrzebne są jednak ustawowe uprawnienia instytucji archiwalnych do dokonywania prac konserwatorskich na zgromadzonych zasobach cyfrowych (np. migracji i emulacji) oraz do sporządzania ich zapasowych kopii bezpieczeństwa. Dla wieczystej archiwizacji zasobów cyfrowych, występujących zarówno na przenośnych mediach cyfrowych, jak i w sieci, są to zabiegi konieczne.

Uregulowania wymaga także kwestia warunków użytkowania archiwizowanych zasobów cyfrowych, głównie tych w postaci sieciowej. Dla zachowania wiarygodności instytucji archiwizujących zasady udostępniania materiałów, w szczególności sieciowych, powinny być zdefiniowane możliwie jednoznacznie. Kwestie podstawowe to miejsce udostępnienia – tzw. dostęp wewnętrzny (ze stanowisk komputerowych wewnątrz instytucji) bądź dostęp zewnętrzny (z prywatnych stanowisk użytkowników w domach lub miejscach pracy) oraz okres czasu – od momentu przekazania dokumentu do instytucji archiwizującej przez twórcę do jej pierwszego udostępnienia użytkownikom.

2.3.2. Ustawa o prawie autorskim i prawach pokrewnych a długoterminowa archiwizacja zasobów cyfrowych

Zagwarantowanie długoterminowej użyteczności dokumentów cyfrowych wymaga okresowego poddawania ich określonym czynnościom, np.: kopiowaniu, emulowaniu bądź migrowaniu. Działania te są z jednej strony niezbędne dla osiągania założonych celów archiwum, z drugiej natomiast kłopotliwe, ponieważ stawiają archiwum w sprzeczności z zapisami ustawy o prawie autorskim i prawach pokrewnych. Konflikt z prawem jest zaś poważnym zagrożeniem dla wizerunku wiarygodnej instytucji archiwizującej.

Nienaruszalność treści i formy utworu

Art. 16 ustawy o prawie autorskim i prawach pokrewnych, definiujący zakres autorskich praw osobistych zawiera przepisy dotyczące nienaruszalności treści

i formy utworu [Ustawa, 1994]. Tymczasem z racji zabiegów konserwatorskich niezbędnych w procesach archiwizacji, głównie migracji oraz emulacji, prawdopodobna, a czasem nieunikniona jest jakościowa bądź ilościowa zmiana danych w archiwizowanym obiekcie. Potrzebne wydaje się ustalenie i udokumentowanie, czy zmiany spowodowane powyższymi zabiegami wywołują zagrożenie dla twórczych elementów w utworze i mogą być postrzegane jako naruszenie autorskich praw osobistych. Nie jest wykluczone, że mogą one zostać uznane za zmiany konieczne, którym twórca nie ma podstawy się sprzeciwiać [Goebel i in., 2004; Neuroth i in., 2009]. Należałoby zastanowić się nad dodatkowymi przepisami w ustawie bądź wprowadzeniem standardów umownych, na których mocy instytucjom archiwizującym zezwala się i precyzuje warunki prac na obiektach archiwalnych oraz przyznaje prawo do ewentualnych, uzasadnionych modyfikacji dokumentu.

Udostępnienie utworu publiczności

W art. 28 ustawy o prawie autorskim i prawach pokrewnych ustawodawca udziela bibliotekom, archiwom i szkołom prawa do nieodpłatnego udostępniania – w zakresie swoich zadań statutowych – egzemplarzy utworów rozpowszechnionych [Ustawa, 1994 ; Ustawa, 2004]. Ze względu na możliwość udostępniania jedynie egzamplarzy utworów rozpowszechnionych dozwolony użytek z art. 28, zgodnie z obowiązującą wykładnią, jest ograniczony do wydań papierowych. Udostępnianie publikacji cyfrowych odbywa się na podstawie umowy licencyjnej. Zdarza się, że twórcy bądź właściciele praw autorskich (np. wydawcy naukowych czasopism elektronicznych) zastrzegają, że udostępnienie treści może nastąpić po określonym czasie od momentu rozpowszechnienia. Może zdarzyć się również – w zależności od typu instytucji archiwizującej i celów, które realizuje – że deponent przekazuje do archiwum dokument, którego udostępnienie publiczności z różnych powodów jest wykluczone bądź ograniczone do grona użytkowników upoważnionych. Potrzebne jest zatem zdefiniowanie kręgu takich użytkowników oraz momentu pierwszego udostępnienia archiwizowanych materiałów. Regulacji wymaga też kwestia miejsca udostępniania zasobów archiwalnych. W zależności od typu materiałów możliwe jest zastosowanie koncepcji udostępniania wewnętrznego (przy terminalach instytucji archiwizujących) bądź zewnętrznego (określanego jako otwarte, czyli pozostające w zgodzie z nowoczesnymi założeniami udostępniania, poprzez wymianę dokumentów na skalę światową).

Nadzór nad sposobem korzystania z utworu

W kontekście tematu niniejszych rozważań warto zwrócić uwagę na zagadnienie technicznych zabezpieczeń, które umożliwiają kontrolę bądź ograniczenie użytkowania dzieła w środowisku cyfrowym – np. Digital Rights Management (DRM) oraz Technological Protection Measures (TPM) [Matlak, 2007, s. 21-22; Szczepańska, 2007, s. 61-63]. Zabezpieczenia są definiowane jako wszelkie technologie, urządzenia lub ich elementy, których przeznaczeniem jest zapobieganie działaniom lub ograniczenie działań umożliwiających korzystanie z utworów bądź artystycznych wykonań z naruszeniem prawa. Bywają one również określane jako skuteczne, techniczne zabezpieczenia, czyli takie, które dodatkowo umożliwiają podmiotom uprawnionym kontrolę nad korzystaniem z chronionego utworu lub artystycznego wykonania poprzez zastosowanie kodu dostępu lub mechanizmu zabezpieczenia, w szczególności szyfrowania, zakłócania lub każdej innej transformacji utworu bądź artystycznego wykonania, bądź mechanizmu kontroli zwielokrotniania, spełniających cel ochronny [Szczepańska, 2007, s. 61].

Obecnie obowiązująca w Polsce ustawa o prawie autorskim i prawach pokrewnych udziela podmiotom uprawnionym do korzystania z utworów na podstawie licencji ustawowych prawa do obchodzenia technicznych środków zabezpieczających bez uzyskiwania zgody właścicieli praw autorskich, wprowadzając równocześnie zakaz posiadania urządzeń służących do obchodzenia zabezpieczeń. Prowadzi to do konieczności uzyskania pliku ze zdjętymi zabezpieczeniami bezpośrednio od wydawcy, co jest trudne w realizacji (nieefektywne). Dodatkowo, w sytuacji, gdy biblioteka posiadająca dzieło zechce je skopiować, do czego ma prawo na podstawie licencji ustawowej dla bibliotek, a właściciel praw autorskich jest nieznany, uzyskanie zgody na usunięcie (obejście) zabezpieczenia stanie się niemożliwe. Po wtóre – włączenie pod ochronę utworów już należących do domeny publicznej. Jeśli zostaną one umieszczone na nośniku, który będzie zabezpieczony, możliwość ich kopiowania zostanie uzależniona od woli uprawnionego. Jeden raz założony system DRM pozwala kontrolować chronione treści niezależnie od tego, czy czas ochrony autorskich praw majątkowych minął, czy nie. Po trzecie – ograniczenie praw bibliotek otrzymujących egzemplarz obowiązkowy jako instytucji zachowujących dziedzictwo narodowe. Ograniczenie dostępu do nabytego dzieła będzie uniemożliwiać realizację statutowych obowiązków, polegających na archiwizowaniu i udostępnianiu tegoż dziedzictwa.

W związku z tym zasadne zdają się postulaty instytucji pamięci o wsparcie legislacyjne w zakresie możliwości zdejmowania zabezpieczeń z dokumentów cyfrowych znajdujących się w ich zbiorach [Siewicz, 2013]. Uzasadniona byłaby również dokładniejsza analiza, czy i w jaki sposób, oprócz kopiowania, systemy

zabezpieczeń technicznych mogą utrudniać inne prace konserwatorskie na dokumentach cyfrowych.

2.3.3. Ustawa o ochronie baz danych a długoterminowa archiwizacja zasobów cyfrowych

W ustawie o ochronie baz danych termin *baza danych* jest definiowany jako „zbiór danych lub jakichkolwiek innych materiałów i elementów zgromadzonych według określonej systematyki lub metody, indywidualnie dostępnych w jakikolwiek sposób, w tym środkami elektronicznymi, wymagający istotnego, co do jakości lub ilości, nakładu inwestycyjnego w celu sporządzenia, weryfikacji lub prezentacji jego zawartości" [Ustawa, 2001]. W świetle tego zapisu, system biblioteczno-archiwalny, podobnie jak biblioteka cyfrowa, stanowi niewątpliwie bazę danych [Szczepańska, 2007, s. 60]. Instytucje tworzące i przechowujące archiwalne kolekcje powinny zatem zastanowić się nad zleceniem opinii prawnej, w których obszarach ich działalności archiwizacyjnej i w jakim zakresie mają zastosowanie zapisy ustawy o ochronie baz danych; jakie uprawnienia i jakie zobowiązania z ustawy dla nich wynikają.

2.3.4. Preservation Policy

Zbiór dokumentów regulujących realizację wszelkich procesów długoterminowej archiwizacji zasobów cyfrowych określony został w piśmiennictwie anglojęzycznym terminem *preservation policy*. W języku polskim odpowiednikiem mógłby być termin *wytyczne* lub *założenia prawne dla długoterminowej archiwizacji zasobów cyfrowych*. Bardziej prawdopodobne jest jednak przyjęcie do polskiej terminologii pojęcia *polish preservation policy*. W dalszej części rozważań stosowane będzie określenie *preservation policy* oraz *polish preservation policy.*

Przypomnijmy jeszcze o odmienności znaczenia pojęć *strategia archiwizacji* oraz *preservation policy*. W odróżnieniu od strategii archiwizacji, w której definiuje się, jak należy archiwizować zasoby cyfrowe, w *preservation policy* znajdują się zapisy określające, co, dlaczego, gdzie i jak długo powinno być archiwizowane. *Preservation policy* jest więc prawną podstawą dla strategii archiwizacji [Neuroth i in., 2008].

Cechą charakterystyczną *preservation policy* powinna być długoterminowość. Raz przyjęte wytyczne nie powinny zmieniać się w zależności od cyklów rozwoju technologicznego, zmian politycznych w kraju lub w zarządzie instytucji archiwizujących, lecz obowiązywać długoterminowo. W zależności od rodzaju i skali projektów archiwizacyjnych opracowywane są instytucjonalne, lokalne, narodowe

oraz międzynarodowe *preservation policies*. Za przykład międzynarodowej *preservation policy* uznano Kartę Ochrony Dziedzictwa Cyfrowego z dnia 17 października 2003 r. przyjętą podczas 32 sesji Konferencji Ogólnej UNESCO w Paryżu [Neuroth i in., 2008].

Najistotniejsze zapisy *preservation policy* powinny dotyczyć następujących kwestii [Neuroth i in., 2008]:

- deklaracji przyjęcia obowiązku długoterminowej archiwizacji zasobów cyfrowych;
- gwarancji dostępności i użyteczności zasobów cyfrowych w przyszłości;
- podstaw prawnych dla działań na rzecz długoterminowej archiwizacji zasobów cyfrowych;
- zapewnienia źródeł finansowania długoterminowej archiwizacji zasobów cyfrowych;
- podziału odpowiedzialności i kompetencji; dla usprawnienia działań w zakresie długoterminowej archiwizacji należy uregulować strukturę współpracy – ustanowić zespoły kompetencyjne, przypisać im określone obowiązki;
- kryteriów oceny i selekcji, czyli typowania dokumentów do kolekcji archiwalnych.

W krajach zaangażowanych w procesy długoterminowej archiwizacji zasobów cyfrowych od lat trwają prace nad utworzeniem narodowych *preservation policies*, jednak z uwagi na potrzebę uwzględnienia różnych interesów społecznych, politycznych oraz możliwości wykonawców jest to proces długotrwały i skomplikowany. W piśmiennictwie przedmiotu niejednokrotnie podkreślono, że utworzenie narodowej polityki dotyczącej długoterminowego archiwizowania dziedzictwa kulturowego jest zadaniem trudnym, w związku z czym proponuje się tworzenie instytucjonalnych *preservation policies*, zorientowanych głównie na lokalne, specyficzne potrzeby indywidualnych instytucji archiwizujących. Instytucje powinny zastanowić się nad własnymi zadaniami archiwizacyjnymi i opracować wewnętrzne procedury umożliwiające im realizację zadań, przynajmniej tych najpilniejszych. Opracowanie instytucjonalnej *preservation policy* jest zwykle postrzegane jako dowód akceptacji nowych funkcji i zadań oraz podjęcia działań na rzecz ich realizacji. Zaleca się, aby w instytucjonalnej *preservation policy* uwzględnić indywidualne cele działania, politykę gromadzenia oraz potrzeby użytkowników. Istotne znaczenie ma sprecyzowanie, które zbiory, w jakiej formie, komu oraz na jakich warunkach mogą zostać udostępnione. Instytucjonalna *preservation policy* może również zawierać indywidualne wytyczne dotyczące sposobu ochrony obiektów cyfrowych, bezpieczeństwa sieciowego, zabezpieczeń pojedynczych stanowisk komputerowych, czy wreszcie wytyczne dotyczące postępowania w razie ewentualnych katastrof. Dodatkowo instytucjonalne *preservation policy* zawierają zwykle zapisy dotyczące stopnia zabezpieczenia autentyczności oraz integralności cyfrowych obiektów. Dopuszczalne jest – w przypadku niektórych instytucji oraz

potrzeb ich użytkowników – ustalenie akceptowalnego stopnia utraty danych, wywołującego, np. przekłamania w formie lub treści dokumentów. Wszelkie odstępstwa od pierwotnej wersji dokumentu powinny być starannie dokumentowane [Building, 2013; Neuroth i in., 2008; OCLC, 2004; PADI, b.d.].

W instytucjonalnej *preservation policy* koniecznym jest ustalenie sposobów i źródeł finansowania zadań długoterminowej archiwizacji zasobów cyfrowych oraz powołanie zespołu osób wraz z rozdzieleniem ich kompetencji, biorąc pod uwagę, iż długoterminowa archiwizacja to zadanie stałe, wymagające permanentnych starań i nakładów finansowych. Instytucjonalna *preservation policy* powinna zawierać zapisy dotyczące ewentualnych potrzeb przekazania zadań długoterminowej archiwizacji instytucjom partnerskim, zastępczym bądź następczym.

Z uwagi na fakt, że *preservation policy* zawiera zapisy dotyczące technicznych aspektów strategii długoterminowej ochrony obiektów cyfrowych, powinna podlegać okresowym, regularnym rewizjom w celu dostosowywania do panujących standardów oraz konfrontacji z indywidualnymi możliwościami instytucji.

Każdy kraj tworzący narodową *preservation policy* powinien, obok narodowych aktów prawnych, uwzględnić dyrektywy nadrzędne. Kraje unijne, np. implementują rozporządzenia Komisji Europejskiej [Building, 2013; Neuroth i in., 2008; PADI, b.d.].

3. Organizacyjne zagadnienia długoterminowej archiwizacji zasobów cyfrowych

Przedstawiony w książce stan wiedzy z zakresu długoterminowej archiwizacji zasobów cyfrowych wynika z międzynarodowej współpzracy licznych organizacji i instytucji, głównie bibliotecznych, archiwalnych i muzealnych, oraz doświadczeń pochodzących z wielu krajowych i międzynarodowych projektów oraz eksperymentów.

3.1. Budowanie świadomości i inicjowanie prac badawczych dotyczących trwałej ochrony zasobów cyfrowych

Organizowanie procesów długoterminowej archiwizacji zasobów cyfrowych w instytucjach pamięci na świecie rozpoczęto w latach 1993-1995 ubiegłego stulecia. W 1993 r. dyskusję na temat potrzeby działań archiwizacyjnych podjęto w National Library of Australia. Na posiedzeniu przedstawicieli australijskich instytucji pamięci, głównie bibliotek i archiwów, podjęto próbę ustalenia sposobu działań na rzecz ochrony materiałów zapisanych w cyfrowej postaci. Została wówczas powołana australijska grupa PADI – Preserving Access to Digital Information, w której skład weszli przedstawiciele bibliotek, muzeów, archiwów, galerii, instytucji nauki, badań i rozwoju [PADI, b.d.].

Za cel funkcjonowania grupy przyjęto przede wszystkim projektowanie i rozwój narodowej strategii oraz opracowywanie wytycznych dla długoterminowej ochrony dostępności i użyteczności cyfrowych zasobów w Australii. Ponadto do zadań PADI włączono tworzenie i utrzymywanie serwisu WWW stanowiącego platformę informacyjną oraz promocyjną wszelkich działań na rzecz długoterminowej archiwizacji. PADI identyfikuje, relacjonuje i propaguje nowe inicjatywy archiwizacyjne podejmowane przez instytucje pamięci na świecie. Pełni rolę forum współpracy środowisk uczestniczących w systemie tworzenia, ochrony i udostępniania zasobów kultury i nauki [NLA, b.d.].

W 1993 r. prace badawcze nad archiwizowaniem dokumentów cyfrowych podjęto także w Stanach Zjednoczonych. Z inicjatywą utworzenia archiwum zasobów cyfrowych wyszła The National Archives and Records Administration (NARA). Po mozolnych próbach odczytu informacji zapisanych w najstarszych plikach z 1960 r., uzmysłowiono sobie, że zachowanie użyteczności dokumentów elektronicznych wymaga przemyślanej, długoterminowej taktyki działań. W konsekwencji w 2000 r. rozpoczęły się intensywne prace nad tworzeniem Electronic Records Archives (ERA) oraz ustaleniem programu jego działania [NARA, b.d.].

Z kolei Library of Congress archiwizacją zasobów cyfrowych zainteresowała się w toku realizacji projektu digitalizacji zbiorów bibliotecznych American Memory, zapoczątkowanego w 1990 r. [LoC, b.d. a]. Na lata 1996-2000 datowany był projekt National Digital Library Program, którego przedmiotem było przyjęcie odpowiedzialności za kolekcję digitalizatów powstałą w projekcie American Memory i zdefiniowanie sposobów jej ochrony w długim czasie [Arms, 2000]. W 2000 r. Kongres Stanów Zjednoczonych oficjalnie zatwierdził amerykański program ochrony zasobów cyfrowych – National Digital Information Infrastructure and Preservation Program (NDIIPP) [LoC, b.d. b].

W Ameryce badania w zakresie długoterminowej archiwizacji dokumentów cyfrowych zostały podjęte również przez Research Library Group (RLG). W 1994 r. powołano grupę roboczą Task Force on Archiving of Digital Information, której celem było zidentyfikowanie i zdefiniowanie najważniejszych zagadnień związanych z ochroną zasobów cyfrowych gromadzonych głównie w bibliotekach. Raport z wynikami prac grupy roboczej, opublikowany w maju 1996 r. przez RLG, jest prekursorskim opracowaniem w całości odnoszącym się do najistotniejszych kwestii długoterminowej ochrony zasobów cyfrowych. Stał się podstawą prac archiwizacyjnych nie tylko w Ameryce, ale również w wielu instytucjach pamięci innych krajów [Preserving, 1996].

Równolegle problematyka długoterminowej archiwizacji cyfrowych zbiorów instytucji pamięci stała się przedmiotem uwagi w instytucjach europejskich.

Zarówno w Holandii, jak i w Niemczech pierwsze dyskusje na temat ochrony publikacji elektronicznych podjęto w 1995 r.; nieco później, bo na przełomie lat 1998-2000, w tych krajach rozpoczęto organizowanie konkretnych działań archiwizacyjnych. W 2001 r. do intensywnych prac nad długoterminową ochroną zbiorów cyfrowych przystąpiono również w Wielkiej Brytanii. Powołana została wówczas Digital Preservation Coalition (DPC), której powierzono odpowiedzialność za ochronę brytyjskiego zasobu cyfrowego [DPC, 2009]. Za podstawę swoich działań DPC przyjęła wyniki prac badawczych, podejmowanych w Wielkiej Brytanii w połowie lat 90. XX w., a koncentrujących się wokół zagadnień tworzenia i zabezpieczania zasobów cyfrowych. Wyniki badań

opublikowano w 1998 r. Na ich podstawie opracowano pierwszą wersję strategii ochrony cyfrowych zasobów w Wielkiej Brytanii [Beagrie i Greenstein, 1998].

Dla działań archiwizacyjnych w Niemczech kluczowa okazała się przedmiotowa dyskusja, której wynikiem była konstatacja, iż biblioteki i archiwa, jeśli chcą być „pamięcią ludzkości", muszą podjąć wyzwanie długoterminowej archiwizacji publikacji elektronicznych z równą powagą, jak to ma miejsce w przypadku mediów konwencjonalnych [Hauffe, 1998, s. 68]. Poprzez pisemne i ustne rozpowszechnianie tego poglądu, kształtowało się w przedstawicielach niemieckich instytucji pamięci przekonanie, że podjęcie działań na rzecz długoterminowej ochrony materiałów cyfrowych jest koniecznością.

Swoje przemyślenia i konkretne rozpoznania na tym gruncie mieli już Holendrzy [Steenbakkers, 1999, s. 93-105], Anglicy [Muir, 2000, s. 151-165], a także Australijczycy [Berthon i Howell, 2000] i Amerykanie [Preserving, 1996]. Nadal jednak postrzegano długoterminową archiwizację narodowych zasobów cyfrowych jako kwestię dotyczącą instytucji pamięci; nie było wówczas mowy o zbiorowej odpowiedzialności za zachowanie cyfrowej kolekcji narodowego dziedzictwa kultury i nauki. Brakowało jednoznacznych stwierdzeń, czy długoterminowa archiwizacja cyfrowego dziedzictwa to indywidualny problem każdego narodu, czy też istnieje szansa konsolidacji sił i szerszej dyskusji, wymiany poglądów, doświadczeń, by ustalić wspólną, ponadnarodową strategię postępowania.

Na 31 sesji Konferencji Plenarnej UNESCO w 2002 r. przyjęto rezolucję na temat ciągłego przyrostu dziedzictwa cyfrowego i potrzeby międzynarodowej kampanii na rzecz zachowania zagrożonej „pamięci cyfrowej" [National Library of Australia, 2003, s. 15]. Rok później w Paryżu została zatwierdzona Karta Ochrony Dziedzictwa Cyfrowego zawierająca szereg zasad dotyczących postępowania w zakresie ochrony światowych zasobów cyfrowych. Poprzez zapisy w Karcie zachęcono organizacje rządowe, pozarządowe, międzynarodowe, publiczne i prywatne do przypisania ochronie dziedzictwa cyfrowego wysokiego priorytetu na szczeblu polityki krajowej i międzynarodowej. UNESCO opracowuje strategię upowszechniania projektów archiwizacji informacji cyfrowych koncentrującej się na szeroko rozumianych konsultacjach, upowszechnianiu zaleceń technicznych, wdrażaniu projektów pilotażowych oraz przyjęciu Karty ochrony dziedzictwa cyfrowego, której towarzyszą Zalecenia opracowane na zlecenie UNESCO przez Bibliotekę Narodową Australii i zawierające ogólne zasady konieczne do uwzględnienia w każdym programie ochrony dziedzictwa [Karta, 2003; Piotrowicz, 2005, s. 45-52]. Z preambuły Karty UNESCO wynika, iż oczywistym jest fakt, że utrata dziedzictwa istniejącego w jakiejkolwiek postaci powoduje zubożenie dziedzictwa wszystkich narodów oraz że zachowanie dziedzictwa dla dobra obecnych i przyszłych pokoleń jest priorytetem o znaczeniu międzynarodowym. UNESCO podjęło się pełnienia roli źródła

informacji i forum współpracy państw członkowskich, organizacji krajowych i międzynarodowych, stowarzyszeń publicznych i sektora prywatnego w opracowywaniu celów, instrumentów polityki i projektów ochrony dziedzictwa cyfrowego [National Library of Australia, 2003, s. 23, 27]. Ustalenia w zakresie długoterminowej archiwizacji zasobów cyfrowych, wynikające z działalności wymienionych instytucji oraz opublikowanych rezultatów prac, przedstawiono w dniach 20-21 kwietnia 2007 r. podczas konferencji zorganizowanej we Frankfurcie nad Menem przez Deutsche Nationalbibliothek oraz niemiecką grupę roboczą do spraw długoterminowej archiwizacji Nestor. Konferencja odbyła się pod hasłem przewodnim: „Długoterminowa archiwizacja. Strategie i praktyka współpracy europejskiej" [Jehn, 2007, s. 10-11]. W jednej z prezentowanych wypowiedzi dokonano krótkiej charakterystyki stanu wiedzy na temat długoterminowej archiwizacji zbiorów cyfrowych. Długoterminowa ochrona kolekcji cyfrowych dokumentów była postrzegana jako wyzwanie przed dziesięcioma laty i tak jest nadal. Zabezpieczenie cyfrowego dziedzictwa kultury i nauki wciąż stanowi dylemat rozpatrywany na łamach piśmiennictwa, w audycjach radiowych i telewizyjnych, a także podczas licznych konferencji. Wyzwanie polega na pokonaniu szeregu organizacyjnych, technicznych, prawnych oraz ekonomicznych problemów związanych z zabezpieczeniem dostępu i użyteczności ogromnej, wciąż wzrastającej ilości heterogenicznych zasobów cyfrowych. Podczas konferencji stwierdzono, że ilość cyfrowych dokumentów rośnie w szybkim tempie i wiele z nich nie ma odpowiedników w wersji tradycyjnej. Jest oczywiste, że instytucje pamięci potrzebują wsparcia w działaniach na rzecz długoterminowej archiwizacji, a programy i narzędzia długoterminowej archiwizacji , choć mają charakter prowizoryczny i nie są pozbawione błędów, powinny być wdrażane, gdyż stanowią podstawę rozwoju i doskonalenia programów docelowych. Ponadto ustalono, że zachowanie dziedzictwa cyfrowego ma sens tylko wówczas, gdy nie stanowi ono idei samej w sobie, lecz jest osadzone w szerszym, europejskim, czy nawet światowym kontekście polityki informacyjnej, odbywa się z założeniem, że ma służyć realizowanym obecnie i planowanym w przyszłości procesom edukacyjnym, naukowym i badawczym na całym świecie. O ile jeszcze do niedawna o długoterminowej archiwizacji toczono rozmowy w bardziej hermetycznych kręgach, tak obecnie zapotrzebowanie na wiedzę z tego zakresu jest bardzo szerokie. Archiwizacja zasobów cyfrowych przestała być zadaniem dla pojedynczych instytucji lub krajów, teraz dostrzega się olbrzymią potrzebę konsolidacji sił i kooperacji na skalę światową. Stało się oczywiste, że edukacji i nauce światowej będą potrzebne wszelkie zasoby cyfrowe, więc bez względu na podziały i różnice instytucje pamięci powinny połączyć siły na rzecz opracowania światowej strategii długoterminowej ochrony cyfrowego dziedzictwa. Każda instytucja pamięci na świecie powinna wypełniać swój obowiązek ochrony dziedzictwa. Najistotniejszą kwestią rozważań jest szukanie

szans na alians ewentualnych strategii narodowych. Nie wypracowano rozwiązań gotowych i gwarantujących sukces w zakresie długoterminowej archiwizacji zasobów cyfrowych, które mogłyby wdrażać instytucje pamięci. Istnieje wiele propozycji, rozmaitych organizacyjno-technicznych konstelacji działań. Znaczącym krokiem naprzód okazało się ustalanie kryteriów wiarygodności dla systemów archiwizacji zasobów cyfrowych. Wiadomo również, że tylko nieliczne kraje podejmują inicjatywy na rzecz ochrony dziedzictwa cyfrowego; ograniczenia działań wielu krajów wynikają z biernej postawy podyktowanej oczekiwaniem na odgórne normy międzynarodowe. Należy jednak liczyć się z możliwym problemem ich dopasowania do lokalnych możliwości. W wielu krajach działania archiwizacyjne są utrudnione bądź niepodejmowane, z racji deficytów w obszarze finansowania.

Konstatacje z 2007 r. są wciąż aktualne. Niewątpliwie w ostatnich latach poziom świadomości, zasób wiedzy i doświadczeń w omawianym temacie wzrosły znacząco. W wielu krajach, także w Polsce, zrealizowano ważne projekty, opublikowano szereg raportów, artykułów i książek, zorganizowano liczne konferencje i warsztaty. Wciąż jednak w wypowiedziach znawców tematu obecne są opinie, w myśl których w archiwistyce cyfrowej jest wiele nierozwiązanych problemów, wiele pytań czeka na odpowiedzi, wiele zadań – na wykonanie, a przyszłość i rozwój technologiczny niewątpliwie będą przynosić kolejne wyzwania.

Istotna jest świadomość potrzeby działań i odpowiednie nastawienie środowiska instytucji pamięci do zadań długoterminowej ochrony. Ważne jest, aby tych zadań nie odsuwać w czasie. Wypowiedzi polskich ekspertów, zgromadzone w 2017 r., rysują nieoczywisty obraz przedmiotowej sytuacji w Polsce. W jednej z rozmów wybrzmiało niezadowolenie, że zbyt dużo się tylko rozmawia i teoretyzuje wobec potrzeby konkretnych działań. Rozmówca uznał, że w wielu instytucjach na konkretne działania trzeba będzie jeszcze poczekać, ponieważ wymagają one podejmowania decyzji restrukturyzujących instytucje pod kątem ich cyfrowego istnienia; decyzji kategorycznych, na których podjęcie jeszcze nie wszyscy są gotowi. Aby myśleć o cyfrowych instytucjach trzeba, w opinii uczestnika badania, zerwać z wieloma paradygmatami charakterystycznymi dla analogowych instytucji, a to w wielu przypadkach „zakrawa o naruszenie świętego status quo". W wypowiedziach zwrócono też uwagę na brak fachowców, brak zaangażowania i wspólnej wizji. Kilku instytucjom udało się „wybić technologicznie" i te „próbują coś zrobić". Przyznano jednak, że „na poziomie krajowym jest to niewielki procent w morzu potrzeb". Doświadczenia rozmówców pokazują, że w większości instytucji, szczególnie na peryferiach, jest źle. Finansowanie w omawianym obszarze jest nadal słabe. Koncepcje w tej materii bazują na całkowitej przebudowie istniejących systemów dokumentów cyfrowych i stanowią asumpt do składania wniosków o dofinansowanie tych działań. o dofinansowanie tych działań.

3.2. Instytucje i organizacje zaangażowane w rozwój długoterminowej archiwizacji zasobów cyfrowych

Liczba instytucji i organizacji współpracujących w zakresie długoterminowej archiwizacji zasobów cyfrowych jest obszerna. Według różnych wykazów od 15 do 50 instytucji, organizacji, zrzeszeń, fundacji, osób prywatnych udziela się w różnych przedsięwzięciach.

Do instytucji inicjujących działania i intensywnie pracujących na dotychczasowy stan wiedzy i doświadczenia w omawianym obszarze należą:

• Research Library Group (RLG) – amerykańska organizacja bibliotek naukowych powołana w 1974 r. jako wspólne przedsięwzięcie biblioteki publicznej Nowego Yorku oraz uniwersytetów Columbii, Harvardu i Yale. Głównym celem przedsięwzięcia było prowadzenie wspólnej polityki w zakresie gromadzenia oraz udostępniania zbiorów. Od momentu fuzji RLG z OCLC w 2006 r. mówi się o programie RLG-OCLC obejmującym ponad 140 bibliotek, archiwów, muzeów oraz innych instytucji pamięci; jego celem jest tworzenie zasobów dla nauki i edukacji, a także zapewnienie ich długoterminowej i wiarygodnej archiwizacji [RLG, b.d.]. RLG jest jedną z pierwszych organizacji na świecie zajmujących się tematyką audytu i certyfikacji repozytoriów cyfrowych. Rozpoznania grupy RLG oraz współpracujących z nią ściśle organizacji OCLC i NARA stanowią niewątpliwie podstawę dla wszelkich późniejszych inicjatyw z tego zakresu tematycznego.

• Online Computer Library Center (OCLC) – amerykańska organizacja o charakterze badawczym i usługowym, powstała w 1967 r. pod nazwą Ohio College Library Center. Jej zadaniem jest przede wszystkim wspieranie procesów automatyzacji, a następnie także cyfryzacji bibliotek; obecnie w zakresie bibliotekarstwa cyfrowego zajmuje się przede wszystkim poprawą dostępu do zasobów informacyjnych oraz obniżaniem kosztów wynikających z ich użytkowania. W 2006 r. miało miejsce połączenie OCLC i RLG [OCLC, 2009], jednak już wcześniej obie instytucje współpracowały głównie w zakresie charakterystyki stabilnych, wiarygodnych archiwów cyfrowych budowanych na podstawie modelu referencyjnego OAIS.

• National Archives and Records Administration (NARA) – narodowe archiwum Stanów Zjednoczonych, które zostało założone w 1934 r. w celu ujednolicenia polityki ochrony stanowych zasobów archiwalnych. W 2003 r. NARA i RLG zainicjowały utworzenie międzynarodowej grupy roboczej do spraw certyfikacji archiwów cyfrowych [Task Force, b.d.].

• Library of Congress – National Digital Information Infrastructure and Preservation Program (NDIIPP) – inicjatywa biblioteki Stanów Zjednoczonych podjęta w 2000 r. w celu rozwoju narodowej strategii gromadzenia, przechowywania

i ochrony cyfrowych zasobów dla obecnych oraz przyszłych pokoleń użytkowników. Obecnie program NDIIPP skupia ponad 130 instytucji partnerskich z całego świata [LoC, b.d. b].

• The Digital Preservation Coalition (DPC) – organizacja utworzona w 2001 r. w celu działania na rzecz ochrony zasobów cyfrowych Wielkiej Brytanii oraz współpracy z organizacjami innych krajów dla tworzenia światowej bazy wiedzy o ochronie dziedzictwa cyfrowego [DPC, 2010a; DPC, 2010b].

• Nestor – Network of Expertise in Long-term Storage of Digital Resources (niem. *Kompetenznetzwerk Langzeitarchivierung*) niemiecka organizacja utworzona w 2003 r. przez siedem niemieckich instytucji bibliotecznych, archiwalnych oraz muzealnych w celu wymiany doświadczeń w zakresie długoterminowej archiwizacji zasobów cyfrowych. Nestor aktywnie działa w pracach międzynarodowej grupy założonej przez RLG i NARA; jest autorem niemieckojęzycznej wersji prezentowanego w niniejszym tekście katalogu kryteriów oceny wiarygodności archiwów cyfrowych [Nestor, 2009]. Publikuje także Newsletter w wersji sieciowej, którego zadaniem jest bieżące dostarczanie informacji o rozwoju sytuacji w zakresie długoterminowej archiwizacji publikacji elektronicznych.

• U. S. Center for Research Libraries (CRL) – założone w 1949 r. jako Midwest Inter-Library Center (MILC) przez amerykańskie uczelnie oraz biblioteki w celu utworzenia wspólnej depozytowej biblioteki gromadzącej i archiwizującej różnego typu zasoby potrzebne w procesach badawczych oraz edukacyjnych. Obecnie CRL skupia ponad dwieście instytucji partnerskich, które wspiera merytorycznie w ich bieżącej działalności oraz rozwoju. Na podstawie doświadczeń pochodzących z licznych prac badawczych, analiz i testów dotyczących funkcjonowania archiwów cyfrowych CRL pełni funkcję doradcy w wielu inicjatywach związanych z archiwizacją dokumentów cyfrowych [CRL, 2010].

• Digital Curation Center (DCC) – konsorcjum czterech brytyjskich instytucji partnerkich (the University of Edinburgh, the University of Glasgow, the University of Bath, the Science and Technology Facilities Council). Powołane w 2004 r. w celu merytorycznego wsparcia brytyjskich instytucji naukowych, badawczych, edukacyjnych i innych generujących oraz przechowujących zasoby cyfrowe. DCC ma doradzać w zakresie metod i narzędzi długoterminowego oraz stabilnego zarządzania zasobami cyfrowymi [DCC, 2010]. Bardzo aktywnie działa w międzynarodowej grupie do spraw audytu i certyfikacji archiwów cyfrowych. Obecnie wspiera prace w zakresie standaryzacji procesów audytu i certyfikacji wiarygodnych archiwów cyfrowych. Ściśle współpracuje z Digital Preservation Coalition.

• Preserving Access to Digital Information (PADI) – przedsięwzięcie narodowej biblioteki Australii, stanowiące wsparcie merytoryczne dla inicjatyw związanych z zarządzaniem i zabezpieczeniem dostępu do cyfrowych zasobów

w przyszłości. PADI to także platforma międzynarodowej współpracy ponad dwudziestu instytucji i organizacji sektora głównie archiwów i bibliotek całego świata [PADI, b.d.].

3.3. Wybrane projekty z zakresu trwałej ochrony zasobów cyfrowych

Prace koncepcyjne i badawcze w zakresie długoterminowej archiwizacji zasobów cyfrowych są podejmowane od połowy lat 90. XX w. Wniosły cenne doświadczenia i dawały asumpt do kolejnych inicjatyw. Celem tej książki nie jest opisanie ich wszystkich, jednak warto wymienić choćby nieliczne z nich, realizowane w ostatnich latach.

• PARSE.Insight – Permanent Access to the Records of science in Europe – międzynarodowy projekt współfinansowany przez Unię Europejską w ramach 7. Programu Ramowego (7.PR) realizowany w latach 2008-2010. W projekcie uczestniczyli pracownicy bibliotek, uniwersytetów oraz instytucji naukowo-badawczych z pięciu państw europejskich. Celem projektu było nie tyle szukanie technicznych rozwiązań dla archiwizacji cyfrowej, ile opracowanie pewnego rodzaju „mapy" przeznaczonej dla infrastruktury e-Nauki (ang. *e-Science*). Podczas badań skupiono się na czterech interesariuszach (ang. *stakeholders*): naukowcach, menedżerach danych, wydawcach oraz fundatorach. Centralny punkt mapy stanowią pracownicy nauki, którzy dostarczają wydawcom i menedżerom obiekty archiwizacji (dane cyfrowe będące wynikami prac naukowych) finansowane przez fundatorów, jednocześnie zaś są użytkownikami tychże danych. Wśród grupy zostały rozesłane ankiety, na które otrzymano ok. 2000 odpowiedzi dotyczących praktyki, potrzeb oraz pomysłów związanych z dokumentami cyfrowymi i ich archiwizacją. Wszyscy interesariusze byli zgodni, że ochrona dorobku naukowego jest kwestią priorytetową [Jakubiec i Pazdur, 2013, s. 45-58; PARSE.Insight, 2011].

• DigiCULT – Technology Challenges for Digital Culture – projekt realizowany od 2002 do 2004 r. [Digi Cult, b.d.]. Celem projektu była organizacja licznych warsztatów poświęconych obserwowaniu zmian technologicznych i ocenie starań instytucji pamięci w zakresie zachowania dostępu do zasobów dziedzictwa kultury i nauki.

• ERPANET – Electronic Research Preservation and Access – projekt realizowany w latach 2001-2004, finansowany ze środków UE. Jednym z założeń było gromadzenie doświadczeń i wymiana informacji o najlepszych rozwiązaniach dotyczących ochrony wyników prac naukowych publikowanych w wersji cyfrowej oraz cyfrowych dokumentów stanowiących dziedzictwo kultury. W ramach pro-

jektu instytucje pamięci wraz z jednostkami badawczo-rozwojowymi oraz firmami informatycznymi zostały powołane do utworzenia platformy wiedzy i wymiany doświadczeń o podejmowanych pracach w zakresie trwałej ochrony zasobów cyfrowych oraz ich efektach. Wyniki prac pochodzące z projektu ERPANET dostarczyły podstaw dla przedsięwzięcia Digital Preservation Europe (DPE) [Digital Preservation Europe, b.d.; ERPANET, 2004; Ross, 2004].

• Preservation and long-term access through networked services (PLANETS) – Zachowanie i długoterminowy dostęp poprzez usługi sieciowe – to czteroletni projekt naukowo-badawczy prowadzony w ramach 6. Programu Ramowego Badań i Rozwoju Technicznego (6.PR) Unii Europejskiej, realizowany w latach 2001--2006. Zrzeszono ekspertów z europejskich bibliotek, archiwów, uniwersytetów oraz firm technologicznych; koordynatorem zadania była British Library, a łącznie wzięło w nim udział 16 instytucji. Celem podstawowym projektu było zbudowanie usług i narzędzi do zapewnienia długoterminowego dostępu do cyfrowych danych kultury i nauki pozwalających semantycznie wzbogacić zasoby oraz automatycznie identyfikować formaty. Skupiono się na opracowaniu metodologii, narzędzi i usług potrzebnych do charakterystyki obiektów cyfrowych, na opracowaniu innowacyjnych rozwiązań dla działań konserwatorskich, na zapewnieniu konsekwentnej i spójnej bazy danych dla obiektywnej oceny różnych protokołów, narzędzi, usług oraz kompletnych planów konserwacji, na automatyzacji procesu opracowania i przechowywania oraz na stworzeniu metod oceny stopnia strat danych cyfrowych [Jakubiec i Pazdur, 2013, s. 45-58; Planets, 2007].

• Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR) – projekt realizowany w ramach 6. Programu Ramowego Badań i Rozwoju Technicznego (6.PR) Unii Europejskiej w latach 2001-2006. Międzynarodowy zintegrowany projekt badawczy poświęcony zapisowi, udostępnianiu oraz wyszukiwaniu informacji w sferze nauki, kultury i sztuki. W ramach projektu współpracowali naukowcy, eksperci w dziedzinie kultury i sztuki, inżynierii wiedzy i ochrony informacji, firmy komercyjne oraz instytucje akademickie z pięciu państw Europejskich: Czech, Francji, Grecji, Wielkiej Brytanii i Włoch oraz z Izraela. Założeniem projektu było stworzenie, wdrożenie i weryfikacja systemu gwarantującego długoterminową archiwizację cyfrowych zasobów kultury i nauki. Główny akcent stanowiło przechowywanie informacji i wiedzy. Po opracowaniu teoretycznych podstaw przyjętego rozwiązania, zaplanowano przeprowadzenie praktycznych testów przyjętych metod i narzędzi, ze szczególnym uwzględnieniem zmieniającego się otoczenia (w tym sprzętu i oprogramowania), zmieniającej się grupy odbiorców (ang. *Designated Communities*) oraz poziomu jej wiedzy (ang. *Knowledge Base*). Głównym celem zatem było opracowanie systemu archiwizacji wzbogaconego o interpretację treści [Lamb i in., 2009].

• Creative Archiving at Michigen & Leeds: Emulating the Old on the New (CAMiLEON Project) – projekt brytyjsko-amerykański, realizowany w latach 1999-2003, który zaowocował szczególnie ważnymi rozpoznaniami dotyczącymi emulacji jako jednej z metod trwałej archiwizacji zasobów cyfrowych. Był też okazją do dyskusji i ustaleń nt. wariantów migracji mających istotne znaczenie w strategii archiwizacji trwałej [CAMiLEON, b.d.]. CAMiLEON stanowił uzupełnienie projektu CEDARS realizowanego w latach 1998-2002, a dedykowanego również szukaniu rozwiązań dla trwałej archiwizacji zasobów cyfrowych [CEDARS , b.d.].

• Sustaining Heritage Access through Multivalent Archiving (SHAMAN) – projekt zainicjowany przez Komisję Europejską, a realizowany w latach 2008-2011 przez instytucje z sektora kultury, nauki i edukacji oraz gospodarki i biznesu. Jego celem było opracowanie koncepcji infrastruktury dla systemu długoterminowej archiwizacji, który byłby w stanie połączyć zasoby cyfrowe wytworzone i składowane w wymienionych sektorach [Dindorf i Schrimpf, 2012, s. 33-35].

• Keeping Emulation Environments Portable (KEEP) – projekt realizowany w latach 2009-2012 na podstawie zaleceń Komisji Europejskiej dotyczących prowadzenia badań i rozwoju nowych narzędzi długoterminowej ochrony zasobów cyfrowych. W projekcie KEEP skoncentrowano się na rozwijaniu narzędzi emulujących, których zadaniem jest umożliwienie odczytu treści statycznych i dynamicznych obiektów cyfrowych przy zastosowaniu metody emulacji. Celem była integracja różnych emulatorów na jednej platformie, tak aby zapewnić odczyt i prezentację dokumentów cyfrowych zapisanych w różnych formatach i na różnych nośnikach. W ramach projektu powstało również narzędzie umożliwiające generowanie obrazów zapisów zakodowanych na dyskach optycznych i magnetycznych. Treści z najstarszych dyskietek i płyt mogą zostać „wyciągnięte" z oryginalnego środowiska zapisu i wdrożone do archiwum cyfrowego. W projekcie zajmowano się też badaniem podstaw prawnych do emulowania otoczenia sprzętowego i programistycznego oraz transferowania treści obiektów cyfrowych, głównie w kontekście prawa europejskiego. Oprogramowanie opracowane jako efekt projektu zostało udostępnione jako Open Source Software [Dindorf i Schrimpf, 2012, s. 33-35; KEEP, 2011].

• Alliance Permanent Access to the Records of Science in Europe Network (APARSEN) – projekt realizowany w latach 2011-2014 przez różne instytucje z trzynastu krajów, sfinansowany przez Komisję Europejską. W ramach projektu nastąpiła wymiana doświadczeń wynikających z prowadzonych przez partnerów projektu prac testowych i eksperymentów dotyczących różnych obszarów długoterminowej archiwizacji zasobów cyfrowych. Instytucje zrzeszone w ramach projektu wspólnie opracowywały system trwałej identyfikacji obiektów archiwalnych. Poddano testom istniejące systemy identyfikacyjne dla obiektów cyfrowych, twórców, instytucji oraz poczyniono próby ich powiązania w jedną „usługę".

APARSEN miał służyć również wprowadzeniu programu certyfikowania wiarygodnych archiwów cyfrowych. Podstawą programu były normy DIN 31644 oraz ISO 16363 [APARSEN, 2013; Dindorf i Schrimpf, 2012, s. 33-35].

• Opportunities for Data Exchange (ODE) – projekt realizowany w latach 2010-2012, którego głównym celem było szukanie rozwiązań dla swobodnej wymiany oraz wykorzystywania danych pochodzących z prac badawczych. W efekcie powstał raport charakteryzujący obecny stan wiedzy oraz istniejące narzędzia i programy umożliwiające integrację zasobów naukowych z perspektywy bibliotek i wydawnictw naukowych. Skupiono też uwagę na problematyce ochrony danych osobowych i innych danych o statusie wrażliwych, a także na obawach przed utratą kontroli nad „własnymi" danymi [Dindorf i Schrimpf, 2012, s. 33-35].

• Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics (PERICLES) – projekt finansowany przez Unię Europejską w ramach 7. Programu Ramowego (ICT Call 9); realizuje cel programu ICT-2011.4.3 Ochrona cyfrowa. Przedsięwzięcie czteroletnie zapoczątkowane w 2013 r. miało na celu sprostanie wyzwaniu zapewnienia dostępu do treści cyfrowych w środowisku podlegającym ciągłym zmianom, obejmującym nie tylko zmiany technologiczne, ale także zmiany w semantyce, praktyce akademickiej i zawodowej lub samym społeczeństwie. Próbowano uchwycić i interpretować zmiany mogące wpływać na postawy i interesy różnych zainteresowanych stron, wchodzących w interakcje z treściami cyfrowymi. Założono gromadzenie i utrzymywanie szczegółowych, złożonych informacji na temat treści cyfrowych, środowiska, w którym istnieją, a także procesów i zasad, którym podlegają. W rezultacie opracowano zestaw modeli, usług, narzędzi i propozycji najlepszych praktyk, nazwany ekosystemem cyfrowym, w celu wspierania procesów archiwizacji, konserwacji i zarządzania cyklem życia obiektów archiwizowanych długoterminowo. Projekt zakończył się w marcu 2017 r. [Pericles, b.d.].

3.4. Przygotowania do zadań długoterminowej archiwizacji zasobów cyfrowych w Polsce

Podczas 31. sesji Konferencji Plenarnej UNESCO w 2001 r. przyjęto Rezolucję 34 zwracającą uwagę na przyrost dziedzictwa cyfrowego w świecie i na potrzebę międzynarodowej kampanii na rzecz jego zachowania. Istotnym krokiem było opracowanie projektu Karty Ochrony Dziedzictwa Cyfrowego [National Library of Australia, 2003, s. 23-27] oraz zachęcenie organizacji rządowych, pozarządowych, międzynarodowych, publicznych i prywatnych do przypisania ochronie dziedzictwa cyfrowego wysokiego priorytetu na szczeblu polityki krajowej. W 2003 r.

podczas 32. sesji generalnej Konferencji UNESCO Karta Ochrony Dziedzictwa Cyfrowego została zatwierdzona, jednak nie zyskała charakteru dokumentu obligatoryjnego. Pomimo to jej treści są respektowane oraz uwzględniane w fachowych dyskusjach i planowanych oraz realizowanych projektach, a także cytowane w opracowaniach przedmiotu. Karta UNESCO przysłużyła się uświadomieniu rangi problemu różnorodnym środowiskom na całym świecie, dostarczając uzasadnienia potrzeby ochrony informacji cyfrowych, wskazując na istniejące zagrożenie utraty dostępu do dziedzictwa cyfrowego oraz określając, kto powinien podjąć wyzwanie długoterminowej ochrony dziedzictwa cyfrowego i za pomocą jakich działań należałoby je realizować [Janczewska-Sołomko, 2006].

Zapisy Karty UNESCO dotyczące ochrony dziedzictwa cyfrowego posłużyły również upowszechnieniu tematyki długoterminowej archiwizacji w Polsce. Stanowią część przetłumaczonego na język polski opracowania australijskiej biblioteki narodowej: *Guidelines for the Preservation of Digital Heritage*, znanego w Polsce jako *Ochrona dziedzictwa cyfrowego: zalecenia*, wydanego w 2003 r. Opracowanie stanowi źródło informacji przydatnych w każdej fazie prac nad długoterminową archiwizacją zasobów cyfrowych, a w szczególności na etapie uświadamiania problemu, rozpoznawania jego istoty oraz przystępowania do wstępnych dyskusji i badań przedmiotu.

Interesujący dla rozważań o świadomości istnienia potrzeby ochrony polskich zasobów cyfrowych okazał się raport z września 2009 r., przygotowany na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego. Zauważono w nim, że „wytwarzane przez polskie instytucje obiekty cyfrowe, na których powstanie przeznaczono znaczne fundusze, nie zawsze są przechowywane w sposób zapewniający ich bezpieczeństwo oraz długoterminową ochronę. W wielu instytucjach posiadających obiekty cyfrowe istnieje niski stopień świadomości, dotyczącej zasad przechowywania dokumentów cyfrowych, co może spowodować w perspektywie najbliższych kilku lat bezpowrotną utratę wielu obiektów cyfrowych, przechowywanych np. na mających niską trwałość płytach CD lub DVD" [Program digitalizacji, 2009, s. 44]. Warty przytoczenia jest również zapis, w którym stwierdzono, że „naturalne dokumenty elektroniczne stanowią ważny składnik polskiej kultury współczesnej i powinny być zabezpieczone dla przyszłych pokoleń nawet z większą dbałością niż odwzorowania cyfrowe powstające w wyniku skanowania, posiadające pierwowzór analogowy, do którego w większości przypadków będzie można wrócić. Dlatego też niezbędne jest opracowanie szczegółowych zasad dotyczących archiwizacji i udostępniania dokumentów elektronicznych oraz systematyczne ich przenoszenie z zagrożonych degradacją nośników fizycznych oraz z Internetu do bezpiecznych repozytoriów cyfrowych" [Program digitalizacji, 2009, s. 42]. Istotny zarówno dla rozwoju świadomości, jak i ochrony polskich

zasobów cyfrowych jest sam fakt opracowania takiego raportu z inicjatywy Ministerstwa. Na podstawie lektury można stwierdzić, że jest to rodzaj ogłoszenia potrzeby przystąpienia do zadań długoterminowej ochrony polskiego dziedzictwa wynikającej ze świadomości polskich władz odnośnie do istoty sprawy. Raport jest również zapowiedzią przystąpienia do zadań archiwizacyjnych w Polsce w latach 2009-2020 [Program digitalizacji, 2009, s. 34-37].

3.4.1. Prace badawcze poświęcone trwałej ochronie polskich zasobów cyfrowych

Zasoby cyfrowe i świadomość ich ochrony

W latach 2007-2009 prowadzono w Polsce badania dotyczące poziomu świadomości polskiego środowiska instytucji pamięci na rzecz opracowania strategii długoterminowej ochrony cyfrowego dziedzictwa narodowego oraz rozwoju rodzimych działań w tym zakresie [Januszko-Szakiel, 2010, s. 405-428; Januszko-Szakiel, 2011a, s. 21-46; Januszko-Szakiel, 2011b, s. 211-230; Januszko-Szakiel, 2012, s. 131-149]. Polscy bibliotekarze i archiwiści byli wówczas zaangażowani przede wszystkim w procesy digitalizacji i tworzenia bibliotek cyfrowych, uczyli się ich planowania i budowy, natomiast archiwizowanie zasobów cyfrowych traktowali jako krok kolejny, zadanie na przyszłość. Brakowało polskich opracowań przedmiotu, fachowej kadry doradców i decydentów. Odczuwalny był niedobór szkoleń i innych form propagowania wiedzy z tego zakresu. Nawet przedstawiciele instytucji pamięci, w których opinii poziom takiej świadomości był zadowalający, przyznawali, iż oprócz świadomości niezbędna jest wiedza z zakresu metod i narzędzi archiwizacji, sposobów jej organizowania i pozyskiwania środków na jej realizację.

Z prowadzonych badań oraz analizy piśmiennictwa wynika, że w Polsce pojęcie długoterminowej archiwizacji utożsamiano, do 2009 r., przede wszystkim z procesami digitalizacji. Digitalizacja sama w sobie miała zapewnić długoterminową użyteczność materiału cyfrowego. Nie uświadamiano sobie, że samo zgromadzenie materiału nie jest jego długoterminowym zabezpieczeniem, gdyż materiał cyfrowy potrzebuje być może staranniejszej ochrony, szczególniejszych warunków przechowywania i metod konserwacji niż dokumenty analogowe. W Polsce fazę pierwszą kształtowania się świadomości i wiedzy o archiwizacji zasobów cyfrowych można datować na lata 1999-2004. Jej początek jest związany z pierwszymi projektami digitalizacji najbardziej zagrożonych zbiorów drukowanych i pierwszymi sieciowymi kolekcjami dokumentów cyfrowych. Koniec zaś to okres pojawienia się w polskiej literaturze przedmiotu wypowiedzi dotyczących długoterminowej archiwizacji zbiorów cyfrowych oraz projektów z tego zakresu

realizowanych w krajach zachodnich. Krzywdząca byłaby jednak powyższa charakterystyka w odniesieniu do całego środowiska związanego zawodowo z instytucjami pamięci. Z pewnością istniały wówczas w Polsce jednostki, które wykazywały się trafną znajomością tematu i świadomością potrzeby archiwizowania dokumentów cyfrowych we właściwym znaczeniu tego terminu.

Przypuszczalnie w okresie prowadzenia omawianych badań świadomość potrzeby archiwizacji dokumentów cyfrowych w Polsce znajdowała się w fazie drugiej, w której pod pojęciem długoterminowej archiwizacji zasobów cyfrowych rozumiano ochronę i zabezpieczenie na przyszłość głównie zasobów bibliotek cyfrowych zgromadzonych i przechowywanych na serwerach. Archiwizację utożsamiano z zabiegami sporządzania ich kopii zapasowych, z okresowym sprawdzaniem sum kontrolnych, z terytorialnym oddaleniem jednej kopii od miejsca archiwum macierzystego, przy czym były to zadania cedowane na informatyków bądź techników bibliotecznych i archiwalnych. Rola bibliotekarzy i archiwistów w tym procesie kończyła się na sporządzeniu metodą digitalizacji cyfrowego dokumentu, który miał być udostępniany użytkownikom. Wśród pracowników instytucji kultury charakterystyczny był wówczas sposób myślenia i działania, zgodnie z którym „podstawowym celem działalności digitalizacyjnej jest udostępnienie zasobów cyfrowych. Powoduje to umieszczanie w bibliotekach i archiwach cyfrowych plików o niskiej rozdzielczości, podczas gdy pliki macierzyste często są archiwizowane na nośnikach nietrwałych (dyski wymienne, płyty DVD)" [Januszko-Szakiel, 2011a, s. 21-46]. W tej fazie sporadyczne były wiedza i świadomość o potrzebie objęcia ochroną dokumentów zapisanych i przechowywanych na fizycznych nośnikach przenośnych. Bardzo rzadko pojawiały się opinie o pilnej potrzebie prób odczytu dokumentów najstarszych.

W drugiej fazie rozwoju świadomości raczej nie mówiło się o roli twórców, a głównie wydawców publikacji elektronicznych w procesach archiwizacji; nie prowadziło się prac w zakresie opracowywania wytycznych dotyczących publikowania elektronicznego. Zwracano uwagę na wytyczne i standardy, ale w odniesieniu do procesów digitalizacji oraz tworzenia bibliotek i archiwów cyfrowych. Przypuszczalnie druga faza poziomu świadomości występowała w dość wąskich kręgach ludzi związanych zawodowo z instytucjami bibliotecznymi, archiwalnymi i muzealnymi. W Polsce istniała wówczas potrzeba upowszechniania zagadnień związanych z długoterminową archiwizacją zasobów cyfrowych. Mało znane były oferty szkoleń, kursów czy konferencji z tego zakresu tematycznego. Respondenci podawali przykłady szkoleń i konferencji dotyczących tworzenia bibliotek cyfrowych, co jest potwierdzeniem, że wówczas istniała skłonność do utożsamiania długoterminowej archiwizacji z digitalizacją i tworzeniem zasobów bibliotek i archiwów cyfrowych. Osobiste rozmowy w ramach wywiadów dostarczyły dowo-

dów, że w temacie archiwizacji istniał terminologiczny nieporządek. W polskim środowisku bibliotecznym używano pojęcia *elektroniczna archiwizacja*, które rozumiano jako digitalizację w sensie zabezpieczenia materiałów drukowanych. Prawdopodobnie przyjmowano, że zdigitalizowanie dokumentu i składowanie w cyfrowej postaci zapewni jego przechowanie i użyteczność w długim czasie. Tymczasem zapewnienie długoterminowej dostępności i użyteczności cyfrowej postaci tego materiału to domena archiwistyki cyfrowej, czyli odpowiednich zabiegów głównie natury technicznej. Rozmówcom należało uświadamiać, że długoterminowa archiwizacja nie powinna oznaczać tylko ochrony materiałów zdigitalizowanych, składowanych na serwerach i na bieżąco udostępnianych, jak to ma miejsce w przypadku bibliotek i archiwów cyfrowych. Równie (o ile nie bardziej) pilne jest objęcie procesami archiwizacji oryginalnych dokumentów cyfrowych zapisywanych na fizycznych nośnikach przenośnych, zwłaszcza tych najstarszych. Po takim wyjaśnieniu rozmówcy przyznawali, że temat długoterminowej archiwizacji zasobów cyfrowych podejmowano w Polsce bardzo rzadko i że konieczne jest jego nagłaśnianie.

Na podstawie badań wnioskowano, że świadomość potrzeby archiwizowania dokumentów cyfrowych osiągnie w Polsce pełną dojrzałość wówczas, gdy procesami archiwizacji zostaną objęte zarówno zasoby zdigitalizowane, jak i pochodzące z procesów publikowania elektronicznego. Nastąpi wówczas trzecia faza rozwoju świadomości, charakterystyczna dla krajów i instytucji pamięci, w których odczytuje się regularnie treści dokumentów cyfrowych i – w zależności od przyjętej strategii archiwizacji – przenosi je na serwery archiwalne bądź pozostawia w oryginalnym środowisku. Jednak wykonuje się odświeżanie bądź zmianę generacji nośnika, planuje się – w razie potrzeby – operacje migrowania i emulowania. Bez względu na typ nośnika sporządza się ich kopie zapasowe. Wszelkie operacje i zabiegi konserwatorskie przeprowadzane na całych kolekcjach cyfrowych bądź poszczególnych jej obiektach są szczegółowo dokumentowane. Powołuje się centralną instytucję koordynującą działania archiwizacyjne w skali kraju, organizuje szkolenia, zarządza diagnozę stanu zasobów cyfrowych w instytucjach pamięci, prowadzi badania i prace w zakresie tworzenia narodowej strategii długoterminowej archiwizacji narodowego zasobu cyfrowego zgodnie z aktualnymi wytycznymi i standardami oraz uświadamia instytucjom rządowym i pozarządowym potrzebę finansowania działań archiwizacyjnych.

Obecnie świadomość potrzeby ochrony polskich zasobów cyfrowych kształtuje się na zupełnie innym poziomie. Choć wciąż występuje rzadko i w ściśle określonych środowiskach, to jednak udaje się identyfikować inicjatywy stanowiące przyczynek do rozpowszechniania problematyki ochrony zasobów cyfrowych, szukania wzorców i metod dla rodzimych działań oraz ustalania źródeł ich

finansowania [Januszko-Szakiel, 2013]. Istotne jest, aby bibliotekarze i archiwiści jako przeszłe potraktowali czasy, w których za zachowanie dziedzictwa narodowego odpowiadały wyłącznie instytucje pamięci [Program digitalizacji, 2009, s. 16]. Odpowiedzialnością za ochronę dziedzictwa narodowego obarcza się w równej mierze organizacje rządowe, pozarządowe, stowarzyszenia publiczne i prywatne, wszelkie instytucje tworzące dorobek nauki i kultury oraz korzystające z niego. Oczywiste jest jednak, że to instytucje pamięci, głównie biblioteki, archiwa i muzea, w szczególności powinny przyjąć tę odpowiedzialność. Nie tylko z racji ich ustawowego powołania, ale i dotychczasowych doświadczeń w służbie na rzecz zachowania dziedzictwa narodowego oraz znajomości potrzeb użytkowników. Muszą jednak otrzymać wsparcie merytoryczne, prawne oraz finansowe, umożliwiające im rozszerzenie dotychczasowych obowiązków i objęcie należytą ochroną narodowego dziedzictwa w postaci cyfrowej.

Eksperyment odczytu najstarszych publikacji elektornicznych

W toku badań z lat 2007-2009, długoterminowa archiwizacja zasobów cyfrowych została określona przez respondentów jako zadanie „na później". Nasunął się zatem wniosek, iż bardzo pilne jest przeciwdziałanie takiemu podejściu poprzez uświadamianie, że archiwizacja dokumentu cyfrowego rozpoczyna się w toku jego projektowania, a na pewno w momencie tworzenia i zapisu, w dodatku wymaga ciągłości i systematyczności działań przez cały okres istnienia dokumentu; z pewnością ochrony długoterminowej nie zapewni włożenie nośnika z danymi „do szuflady" [National Library of Australia, 2003, s. 30]. Opisany wątek badawczy stał się inspiracją do przeprowadzenia eksperymentu, którego celem było ustalenie, w jakim stanie są najstarsze publikacje elektroniczne przechowywane w polskich bibliotekach. Eksperyment polegał na identyfikacji oraz próbie prezentacji treści najstarszych materiałów bibliotecznych opublikowanych na przenośnych mediach typu dyskietki. Na miejsce eksperymentu wybrano jedną z bibliotek uprawnionych do otrzymywania egzemplarza obowiązkowego opublikowanych dokumentów. Próba odczytu miała miejsce w czerwcu 2009 r. [Januszko-Szakiel, 2012, s. 131-149].

Na podstawie przeglądu katalogu komputerowego biblioteki stwierdzono, że identyfikacja najstarszych publikacji elektronicznych jest trudna, a w przypadku wielu tytułów niemożliwa. Podczas rozmowy z dyżurnym bibliotekarzem ustalono, że najstarsze publikacje elektroniczne zgromadzone i przechowywane w bibliotece pochodzą z 1994 r. i są zapisane na dyskietkach typu 5,25 oraz 3,5 cala. Zachowane dyskietki stanowią materiał uzupełniający do publikacji wydanych drukiem, adnotację zaś o nich w opisie katalogowym dokumentu zamieszczano

sporadycznie. Nie było wówczas instrukcji opisu bibliograficznego tego typu materiałów bibliotecznych, zatem w zależności od decyzji katalogującego bibliotekarza informacja o elektronicznym dodatku do publikacji drukowanej była zamieszczana bądź pomijana. Materiał uzupełniający w postaci dyskietki był oddzielany od dokumentu drukowanego i przechowywany osobno.

Na pytanie, czy możliwy jest dostęp do najstarszych zachowanych dyskietek, dyżurny bibliotekarz odpowiedział twierdząco i zgodził się je udostępnić, prosząc o chwilę czasu. Po około pięciu minutach do czytelni dokumentów audiowizualnych wyposażonej w rozmaite sprzęty do odczytu i prezentacji multimediów zostały przyniesione dwa tekturowe podłużne opakowania. W jednym z nich znajdowało się około 100 dyskietek typu 5,25 cala, ułożonych chronologicznie, w drugim znajdowała się podobna ilość dyskietek typu 3,5 cala, także w układzie chronologicznym. Kolejne pytanie skierowane do dyżurnego bibliotekarza dotyczyło dostępności sprzętu i oprogramowania, które umożliwią odczyt treści zapisanych na dyskietkach. W odpowiedzi bibliotekarz stwierdził, że w przypadku tych materiałów możliwy jest tylko ich ogląd. W bibliotece nie ma już ani jednego komputera ze stacją i oprogramowaniem, które czytają dyskietki 5,25 cala. W przypadku dyskietek 3,5 calowych są dostępne komputery z odpowiednią stacją dysków, ale problemem okazał się brak oprogramowania. Bibliotekarz przyznał, że nie zachowano programów, które byłyby w stanie odczytać kod binarny i zaprezentować go na ekranie w postaci treści zrozumiałej dla użytkownika. Nie wykonywano także kopii zapasowych tych dokumentów; nie przenoszono treści z najstarszych dyskietek na nośniki nowszych generacji. Nie było polityki planowego okresowego odczytu tychże dokumentów, więc ewentualne trudności odczytu i prezentacji ich treści mogły zostać stwierdzone jedynie w procesie udostępniania użytkownikom. Jest bardzo prawdopodobne, że publikacje, których nie zamawiali czytelnicy, były odczytywane tylko jeden raz – w procesie opracowania publikacji drukowanej, do której zostały załączone jako materiał towarzyszący. W bibliotece nie udało się również odnaleźć dokumentacji omawianych zasobów.

Brak charakterystyki najstarszych dokumentów skutkuje poważnymi trudnościami w procesach oceny wartości i przydatności zawartych w nich treści. Tymczasem decyzja o kosztownych zabiegach archiwizacyjnych, których celem miałoby być odtworzenie zawartości najstarszych nośników danych cyfrowych, powinna być poprzedzona staranną analizą ich przydatności. Brak metadanych utrudnia również identyfikację środowiska programowego odczytu treści najstarszych publikacji cyfrowych.

Podjęta w eksperymencie próba odczytu najstarszych polskich publikacji elektronicznych nie powiodła się. Prawdopodobnie bezpowrotnie utracono dostęp do ich treści. Przyczyną był niedostateczny poziom świadomości dotyczący

potrzeby okresowego odczytu treści zapisanych na nośnikach danych cyfrowych, stałego obserwowania rozwoju technologicznego oraz reakcji na zachodzące zmiany [Januszko-Szakiel, 2012, s. 131-149]. Z rozmów z pracownikami polskich instytucji nauki i kultury prowadzonych w latach 2012-2017 wynika, że nadal nie podejmuje się i nie planuje w naszym kraju prób odczytu i ratowania treści najstarszych publikacji. Jeśli takie próby występują, mają charakter inicjatyw bardzo rzadkich[1]. Na tej podstawie wytworzy się lub już wytworzyła tzw. dziura cyfrowa w polskim zasobie nauki i kultury[2].

Stan rzeczy w tym zakresie oddają opinie ekspertów z 2017 r. W jednej z instytucji kultury wykonano migrację archiwum obiektów zdigitalizowanych, przez Pracownię Reprografii, z płyt CD i DVD na macierze. Nie dotyczy to publikacji cyfrowych gromadzonych w zbiorach ogólnych. Z kolejnego przekazu wynika, że w instytucji jest tworzone archiwum, którego pomysł narodził się kilka lat temu. Większa część dokumentów cyfrowych powstałych w latach 2004-2008 została przeniesiona z dysków optycznych na serwery i częściowo również opracowania tych dokumentów (metadane). Niewielka część zasobu została jednak utracona. Jeszcze inny respondent zadeklarował, że zbiory elektroniczne w instytucji, którą reprezentuje są sprawdzane okresowo i jest z tym duży problem, ponieważ części z nich niedługo nie będzie można odczytać. Prowadzone są prace zabezpieczające, polegające na przenoszeniu zbiorów na macierze dyskowe i archiwizowaniu ich na taśmach. Niestety brakuje środków, aby zdążyć zabezpieczyć cały posiadany zasób. Ze względu na koszty prace są prowadzone tylko w minimalnym zakresie. W opinii rozmówcy, dziura cyfrowa była i będzie. Jest to nieuniknione; takie są koszty cyfryzacji. Za większy problemem uznaje się niski poziom techniczny digitalizacji i przede wszystkim ciągle niski procent zdigitalizowanego i udostępnianego materiału.

---

[1] Rzecz dotyczy planów Biblioteki Narodowej w zakresie organizacji prac nad analizą ilościową i jakościową dokumentów elektronicznych. Jest to wprawdzie dopiero wstępna faza koncepcji organizacji tych prac, ale decyzja o zasadności ich realizacji zapadła. Informacje na ten temat pochodzą z wywiadu telefonicznego prowadzonego w 2017 r.

[2] Jednym z propagatorów coraz popularniejszego sformułowania „dziura cyfrowa" jest Maciej Dziubecki, prezes spółki Aleph Polska specjalizującej się w dystrybucji m.in. oprogramowania archiwów cyfrwych, uczestnik II Krakowskiej Konferencji Bibliotek Naukowych, która odbyła się w Krakowie w dniach 24-25 października 2012 r. Konferencja była poświęcona tematyce długoterminowej archiwizacji polskiego dziedzictwa cyfrowego. Pojęcie wymieniono w toku dyskusji podczas konferencji, później też w treści wpisu dotyczącego konferencji na blogu M. Dziubeckiego: https://www.aleph.pl/ii-krakowska-konferencja-bibliotek-naukowych/ [Dostęp: 27.10.2017]. Wcześniejsze zastosowanie tego terminu miało miejsce w opracowaniu *Cyfrowa czarna dziura* [Palm, 2011].

Dokumentacja zasobów cyfrowych

Gromadzenie i przechowywanie dokumentów cyfrowych w magazynach to za mało, aby przyszłym pokoleniom użytkowników zapewnić dostęp do ich treści. Instytucje pamięci powinny przystąpić do prac nad opracowaniem strategii długoterminowej archiwizacji zgromadzonych zasobów cyfrowych. Idąc śladem instytucji krajów zachodnich, zaawansowanych w działaniach archiwizacyjnych [Hedstrom i Montgomery, 1998], pierwszym praktycznym zadaniem powinno być rozpoznanie stanu ilościowego i jakościowego zgromadzonych zasobów cyfrowych oraz stworzenie dokumentacji informującej o:

- liczbie dokumentów cyfrowych zgromadzonych i przechowywanych w instytucji;
- łącznej objętości tychże dokumentów;
- nośnikach, na których są one zapisane;
- formatach ich zapisu;
- najstarszych dokumentach cyfrowych znajdujących się w zasobach instytucji;
- ewentualnych utrudnieniach odczytu przechowywanych dokumentów cyfrowych;
- liczbie dokumentów cyfrowych, do których z różnych powodów utracono dostęp oraz znaczeniu ich treści i stopniu zapotrzebowania na nie ze strony użytkowników.

Próbę takiej diagnozy na rodzimym gruncie poczyniono w ramach jednego z przedsięwzięć badawczych [Januszko-Szakiel, 2010, s. 405-428]. Do instytucji pamięci skierowano pytanie o prowadzenie dokumentacji dotyczącej przechowywanych zasobów cyfrowych oraz o strategię ich archiwizacji. Dwie z siedmiu badanych instytucji przyznały, iż pomimo gromadzenia i przechowywania publikacji cyfrowych nie posiadają wiedzy o ich liczbie i kondycji. Instytucje te nie prowadziły dokumentacji o własnych zasobach cyfrowych i nie stosowały żadnej strategii ich archiwizacji. Prawdopodobnym podłożem tego niepokojącego faktu była nieświadomość lub brak wiedzy o postępowaniu ze zbiorami cyfrowymi. Nie był to jednak odosobniony przypadek – podobna sytuacja miała miejsce w niemieckich instytucjach bibliotecznych w połowie lat 90. XX w. Wówczas to Deutsche Nationalbibliothek jako główna instytucja pamięci przyjęła obowiązek długoterminowej archiwizacji narodowego zasobu cyfrowego i zobligowała niemieckie instytucje biblioteczne do ilościowego oraz jakościowego zdiagnozowania stanu przechowywanych w nich publikacji cyfrowych. Okazało się, że większość bibliotek tylko gromadziła publikacje cyfrowe – były starannie opracowywane i składowane w magazynach, jednak brakowało świadomości, że kiedyś ich odczyt i użytkowanie mogą być zagrożone [Henze, 1999, s. 15-17]. To był początek prac nad długoterminową ochroną dokumentów cyfrowych w Niemczech.

Dalsza analiza wyników rodzimego badania pozwoliła wyciągnąć mniej niepokojące wnioski [Januszko-Szakiel, 2010, s. 405-428]. Otóż w kolejnych instytucjach prowadzono dokumentację, dzięki której możliwe było oszacowanie liczby oraz łącznej objętości zgromadzonych dokumentów cyfrowych. Możliwe było także wskazanie formatów, w jakich są one zapisane, wraz z podaniem odsetka dokumentów określonego formatu. Sondowane instytucje zastrzegały jednak, że ich cyfrowe kolekcje nieustannie rozrastają się – zarówno poprzez gromadzenie dokumentów cyfrowych, jak i wskutek digitalizacji materiałów analogowych. Istotne było pytanie o stosowane w projektach digitalizacji formaty zapisu uzyskanych wersji cyfrowych. Celem było ustalenie, czy instytucje decydują się na formaty standardowe, otwarte, obsługiwane przez powszechnie dostępny sprzęt i oprogramowanie oraz posiadające cechy zwiększające prawdopodobieństwo bezpiecznego przechowania dokumentów w czasie. Konsekwentnie wymieniane w odpowiedziach formaty to: TIFF, GIF, DjVu, JPEG, rzadziej PDF. Celem projektów digitalizacji powinno być tworzenie tzw. dobrych obiektów cyfrowych [Bednarek-Michalska, 2006], czyli takich, które są zapisane w formatach ułatwiających tworzenie kopii i które mogą być wymieniane pomiędzy platformami. Ma to niebagatelne znaczenie z racji krótkotrwałości zapisów cyfrowych i bardzo prawdopodobnej potrzeby zmiany platformy, migrowania dokumentów i metadanych czy innych zabiegów natury archiwizacyjnej. Każdy z wymienianych formatów lepiej lub gorzej spełnia określone funkcje, także w różnym stopniu nadaje się do celów archiwizacji, zwłaszcza długoterminowej. Jednak format umożliwiający bezproblemowe przeprowadzenie obiektu cyfrowego przez kolejne generacje otoczenia sprzętowo-programowego, tzw. format archiwizacyjny, jest wciąż poważnym deficytem w dziedzinie archiwistyki cyfrowej.

Obok poznania faktycznego gabarytu zgromadzonego dotychczas zasobu cyfrowego, instytucjom archiwalnym potrzebne jest też oszacowanie tempa jego przyrostu. Ma to ogromne znaczenie dla ustalenia pojemności archiwum cyfrowego i jego skalowalności. Przy szacowaniu pojemności archiwum należałoby rozstrzygnąć, czy długoterminową archiwizacją zostaną objęte wszelkie zgromadzone zasoby cyfrowe, czy tylko te spośród nich, które przedstawiają znaczącą wartość dla nauki i kultury i są zaliczane do dziedzictwa narodowego. Ocena i selekcja zasobów cyfrowych stanowi kolejne, niełatwe zadanie w ramach ich długoterminowej archiwizacji.

W czterech z siedmiu ankietowanych instytucji możliwa była identyfikacja najstarszych dokumentów elektronicznych. Tylko w jednej ankiecie wskazano, że najstarsze cyfrowe publikacje pochodzą z lat 1998-1999, natomiast cyfrowe kolekcje pozostałych trzech instytucji wzięły swój początek od dokumentów publikowanych po 2000 r. Dwie instytucje deklarowały utratę dostępu do materiału

cyfrowego – w jednym przypadku dotyczyło to dokumentów zapisanych na dyskietkach 5,25 cala, w drugim – chodziło o kolekcję skanów zapisaną w formacie TIFF na płytach CD, które uległy uszkodzeniu. W jednym i drugim przypadku nie podano szczegółów utraty dostępu, przypuszczalnie był to wynik nieprzestrzegania warunków trwałości nośników i warunków ich przechowywania bądź też nieuświadomienia potrzeby okresowego odczytu tychże dokumentów.

Okresowy odczyt dokumentów cyfrowych

Długoterminowa bezstratna archiwizacja zasobów cyfrowych zależy od wielu różnych czynników, ale najistotniejsze to rodzaj zastosowanego nośnika oraz jego okresowy odczyt, niezależnie od prognozowanej trwałości [Daszewski, 2006, s. 85-94]. Powszechnie wiadomo, że np. nośnik CD należałoby po 5 latach od daty zapisu na nim danych poddać procesowi odświeżenia.

Wobec tego, że okresowy odczyt materiałów cyfrowych jest traktowany jako istotny element strategii ich długoterminowej archiwizacji, w badaniu z 2007 r. zapytano, czy i z jaką częstotliwością dokonuje się okresowego odczytu zgromadzonych dokumentów cyfrowych w celu identyfikacji ewentualnych utrudnień ich użytkowania. Wyniki świadczyły o braku przemyślanej strategii w tym zakresie. Odczyt dokumentu zasadniczo następował w procesie bieżącego udostępniania, a ewentualne nieprawidłowości funkcjonowania zgłaszali użytkownicy. Brak planowego odczytu dokumentów mógł spowodować, że w instytucjach tych przetrwały tylko te zasoby, które były najczęściej udostępniane. Faktem jest, że nie każdy dokument charakteryzuje się wysoką potrzebą i częstotliwością użytkowania oraz wartością ponadczasową i nie każdy musi przetrwać do czasu, kiedy ochronę nad nimi przejmą profesjonalne certyfikowane archiwa cyfrowe. Jednak częstotliwość korzystania z dokumentu, choć nie jest tu bez znaczenia, z całą pewnością nie powinna stanowić jedynego kryterium decydującego o jego przetrwaniu. W tym miejscu znów należałoby podkreślić znaczenie procesu oceny i selekcji dokumentów szczególnie ważnych i bezwzględnie zasługujących na trwałą ochronę.

Do badanych instytucji skierowano również pytanie o procesy odświeżania nośnika oraz migracji danych cyfrowych do aktualnych formatów, a także o sporządzanie kopii bezpieczeństwa. Tylko w dwóch instytucjach nie realizowano wówczas ani procesów odświeżania nośnika, ani migracji danych do aktualnych formatów. Odpowiedzi tych nie uzupełniono żadnym dodatkowym komentarzem zdradzającym motywy takiego podejścia. W pozostałych instytucjach istniały różne ustalenia na temat tych dwóch procesów, jednak ani odświeżanie nośników, ani migrowanie danych nie było w nich wówczas potrzebne; są to instytucje, w których wszystkie dokumenty cyfrowe mogły być w czasie badania bez problemu

użytkowane. Z dwóch odpowiedzi wynikało, że instytucje liczą się z ewentualną potrzebą migracji danych i traktują ten proces jako przyszłe zadanie do wykonania. Wątpliwość budzi natomiast kwestia procesu odświeżania nośnika. Tylko w dwóch ankietach udzielono odpowiedzi – w jednej instytucji postanowiono odświeżać nośnik jeden raz w roku, w drugiej zaś minimum raz na pięć lat. Z powodu braku uzasadnienia takiej decyzji i szczegółowych informacji na temat technicznych parametrów zastosowanych nośników, trudno komentować, czy ustalenia te były właściwe, czy też nie, z całą pewnością jednak decyzja o częstotliwości procesów odświeżania powinna być rezultatem starannych studiów dotyczących trwałości poszczególnych nośników, na których zapisane są dokumenty.

Wiedza i techniczne warunki zarządzania dokumentami cyfrowymi

W ankiecie z 2007 r. zamieszczono cykl pytań służących wysondowaniu, jaka jest wiedza oraz techniczne możliwości pracowników instytucji pamięci w zakresie zarządzania zbiorami cyfrowymi. Ze wszystkich odpowiedzi wynika jednoznaczna zgodność co do tego, że materiał cyfrowy wymaga szczególnych warunków oraz metod przechowywania, w przeciwnym razie treści dokumentów zapisanych cyfrowo zostaną utracone. Przedstawiciele badanych instytucji zadeklarowali obecność fachowców (informatyków, archiwistów, bibliotekarzy bądź pracowników informacji specjalizujących się w organizacji oraz obsłudze bibliotek i archiwów cyfrowych), a także dostępność narzędzi (otoczenie programowo-sprzętowe), potrzebnych do archiwizacji i zarządzania zasobami cyfrowymi. W niektórych instytucjach deklarowano zatrudnienie informatyków orientujących się w omawianej problematyce, ale mieli oni inny zakres obowiązków; ponadto instytucje te nie były w posiadaniu technologii informatycznych umożliwiających im profesjonalną ochronę i zarządzanie zasobami cyfrowymi.

Z uzyskanych wypowiedzi wynikało, że zarządzanie zasobami cyfrowymi powinno służyć przede wszystkim ich sprawnemu wyszukiwaniu i udostępnianiu, przy jednoczesnej dbałości o zabiegi natury konserwatorskiej oraz respektowaniu praw własności intelektualnej. Dopuszczano, aby opiekę techniczną nad zasobami cyfrowymi zlecić specjalistom zewnętrznym. Wyraźnie w wypowiedziach respondentów podkreślany był temat potrzeby sporządzania kopii bezpieczeństwa dla dokumentów cyfrowych. Można przypuszczać, że ich tworzenie było wówczas najmocniejszą stroną strategii ochrony zasobów cyfrowych w polskich instytucjach pamięci.

W toku prowadzonych badań ustalono, że w polskich instytucjach pamięci, nawet w tych, w których najstarsze dokumenty cyfrowe pochodzą z lat 1994-1996, nie było do 2009 r. strategii długoterminowej ochrony ich użyteczności. Odczyt

i prezentacja treści najstarszych dokumentów były w wielu instytucjach utrudnione, a nawet niemożliwe z racji braku potrzebnego sprzętu i oprogramowania. Mniej niepokojący okazał się stan zbiorów cyfrowych wytworzonych po 2000 r. Odczyt i prezentacja ich treści były w zasadzie wykonalne, aczkolwiek konieczne i pilne już wtedy było podjęcie decyzji o sposobie długoterminowego utrzymania ich użyteczności. Z badań wynikało, że w dobrej kondycji były np. zbiory polskich bibliotek cyfrowych. Z uwagi na to, że kolekcje te były i są nadal tworzone, składowane i zarządzane w profesjonalnych systemach informatycznych oraz posiadają kopie bezpieczeństwa, mają duże szanse przetrwania do czasu, kiedy zostaną umieszczone i archiwizowane w profesjonalnych systemach depozytowych.

Analizując zgromadzony materiał badawczy, zauważono, że wśród dokumentów cyfrowych przechowywanych w polskich instytucjach pamięci istnieje wyraźny podział na dwie grupy. Pierwszą grupę stanowią dokumenty najstarsze, które należałoby określić jako „nieużyteczne" bądź „nieczytelne", wydane w latach 1994-1999, na nośnikach i w formatach, które obecnie nie są w powszechnym użytku i których odczyt oraz prezentacja są niemożliwe z powodu braku starszych wersji sprzętu i oprogramowania. Druga grupa natomiast to dokumenty „użyteczne", „czytelne", stworzone po 2000 r., w zdecydowanej większości w formatach i na nośnikach uznanych za standardowe, możliwe do odczytania i prezentacji przy użyciu powszechnie dostępnej platformy sprzętowo-programowej. W drugiej grupie bardzo poważną część stanowią zasoby bibliotek i archiwów cyfrowych przechowywane na serwerach, nad którymi fachową opiekę sprawują techniczni pracownicy bibliotek i archiwów, często informatycy przygotowani do prac nad zarządzaniem i archiwizacją kolekcji cyfrowych dokumentów. Archiwizacja ta polegała zazwyczaj na zautomatyzowanej kontroli sporządzanych sum kontrolnych oraz tworzeniu kopii zapasowych. Ze studiów rodzimej literatury przedmiotu oraz badań wynika, że kolekcje te charakteryzowały się dużą szansą przetrwania i zachowania swej użyteczności w długim czasie. Potwierdzeniem tej opinii może być wypowiedź z wywiadu przeprowadzanego w Bibliotece Narodowej: „W Bibliotece Narodowej prace nad kwestiami bezpiecznej i długotrwałej archiwizacji trwają od 2006 r. Ze względu na obowiązek statutowy BN polegający na wieczystej archiwizacji jednego egzemplarza polskich materiałów bibliotecznych od początku budowy zasobów cyfrowych kwestia ich archiwizacji była traktowana z całą powagą i w oparciu o najlepsze wzorce opracowane w innych europejskich bibliotekach narodowych". Podobnie dzieje się w Bibliotece Jagiellońskiej, tzn. zbiory zdigitalizowane podlegają ochronie w sensie poddawania ich zabiegom typowym dla przechowywania danych w systemie informatycznym, np. sporządzanie kopii zapasowych.

O dużej dbałości i profesjonalnym zarządzaniu zasobami polskich bibliotek cyfrowych świadczą prace Poznańskiego Centrum Superkomputerowo-Sieciowego.

Od początku były one realizowane z uwzględnieniem przyjętych w świecie standardów [Konferencja, 2008]. Pomimo znacznie skromniejszych możliwości finansowych, polskie biblioteki cyfrowe są organizowane i funkcjonują na poziomie podobnych kolekcji w świecie. Wprawdzie ich ochrona zapewniająca długoterminową użyteczność jest tematem podejmowanym dopiero od 2008 r. i wymaga dużego nakładu pracy, ale przemyślane, staranne i konsekwentne działania pozwalają osiągnąć pożądany efekt. O rozwoju świadomości pracowników PCSS oraz ich aktywnej postawie wobec problematyki trwałej ochrony użyteczności zasobów cyfrowych świadczy usługa dArceo dedykowana trwałej archiwizacji zasobów cyfrowych [Mazurek i in., 2013, s. 101-111].

Na podstawie opinii respondentów zgromadzonych w latach 2007-2009 nasunął się wniosek, że w środowisku polskich bibliotekarzy i archiwistów temat długoterminowej archiwizacji zasobów cyfrowych wywoływał skojarzenia związane z digitalizacją i budowaniem kolekcji cyfrowych; nie wspominano o ochronie oryginalnych dokumentów cyfrowych, zwłaszcza tych starszych, zapisanych na fizycznych nośnikach przenośnych, a to właśnie w ich przypadku potrzeba działań jest najpilniejsza. Jak już wspomniano, należałoby podjąć starania na rzecz ustalenia ich ilości, objętości, rodzajów nośników oraz formatów zastosowanych do ich zapisu, a także ustalić sprzęt i oprogramowanie potrzebne do ich odczytu i prezentacji, podjąć ewentualnie decyzję o zastosowaniu metody migracji bądź emulacji w celu odtworzenia treści tych dokumentów. Przede wszystkim jednak należałoby zastanowić się, czy treść tych dokumentów jest warta ponoszonych kosztów takich czynności oraz czy jest zgłaszana potrzeba ich użytkowania.

Ważne, aby nie doprowadzić do sytuacji, w której zasoby użyteczne dołączają do grupy zasobów nieużytecznych. Jedynie w Bibliotece Narodowej deklarowano istnienie planów dotyczących okresowego odczytu publikacji elektronicznych, a także rozpoczęcie prac związanych z procesami odświeżania nośników publikacji elektronicznych. Z pozostałych odpowiedzi wynika, że procesy odświeżania nośników nie były do 2007 r. przeprowadzane, a kopie bezpieczeństwa mają tylko kolekcje dokumentów zdigitalizowanych przechowywanych na serwerach bądź macierzach dyskowych. Treści dokumentów z takich nośników jak dyskietki i płyty CD-ROM nie były przenoszone ani na nośniki nowszych generacji, ani na serwery. Tego typu pomysły były wysuwane przez pracowników Biblioteki Jagiellońskiej, jednak przeszkodą okazał się brak narzędzi emulujących.

W badanych instytucjach nie dokonywano oceny i selekcji dokumentów w celu identyfikacji tych, które mają ponadczasową wartość, istotne znaczenie dla nauki i kultury oraz powinny podlegać szczególnej ochronie. Oznacza to, że w planach dotyczących odczytu treści dokumentu i procesów odświeżania bądź zmiany generacji nośnika, ewentualnie oddzielenia treści dokumentu od nośników

oryginalnych i umieszczenie ich na serwerze archiwalnym, instytucje powinny uwzględnić potrzebę oceny wartości treści materiałów cyfrowych. Prawdopodobne jest, że ze względów finansowych, na serwerach archiwalnych, nawet tych w archiwach i bibliotekach narodowych, przechowywane będą tylko dokumenty o szczególnym znaczeniu i wartości ponadczasowej. W jednej z ankiet zwrotnych zamieszczono komentarz powołujący się na zalecenie IBM oraz Digital Center, z którego wynika, że dużym ułatwieniem procesu długoterminowej archiwizacji jest „ucieczka" od fizycznych nośników przenośnych. W piśmiennictwie przedmiotu, głównie angielskojęzycznym, niejednokrotnie prezentowano podejście w archiwizacji polegające właśnie na oddzielaniu treści dokumentów od oryginalnych nośników i ich archiwizowanie w cyfrowych systemach depozytowych.

Do polskich instytucji pamięci problematykę długoterminowej archiwizacji zasobów cyfrowych wprowadzono z około dziesięcioletnim opóźnieniem w stosunku do krajów zachodnich. Jest to oczywiście zjawisko niepokojące i prowokujące pytania, czy polskim bibliotekarzom i archiwistom uda się ochronić użyteczność najstarszych polskich dokumentów cyfrowych oraz czy polscy użytkownicy bezpowrotnie utracili możliwość ich odczytu i interpretacji. Długoterminowa użyteczność polskich zasobów cyfrowych, głównie tych najstarszych, jest niewątpliwie zależna od świadomości przedstawicieli polskich instytucji pamięci dotyczącej pilnej potrzeby organizacji procesów archiwizacyjnych zgodnie z zasadami charakterystycznymi dla trzeciej dojrzałej fazy jej rozwoju.

Wyniki omawianych badań oraz obserwacja podejmowanych aktualnie inicjatyw potwierdzają, że w polskich instytucjach pamięci są zatrudnieni ludzie z wykształceniem i predyspozycjami umożliwiającymi organizowanie prac w zakresie długoterminowej archiwizacji zasobów cyfrowych. W diagnozowanych instytucjach deklarowano również posiadanie otoczenia sprzętowo-programowego umożliwiającego odczyt i prezentację większości publikacji elektronicznych. Zarówno w Bibliotece Jagiellońskiej, jak i Bibliotece Narodowej przyznano, że poważny problem może stanowić odczyt najstarszych dokumentów. W obu instytucjach zgodzono się z opinią, że jest to wynik nieświadomości potrzeby podjęcia działań archiwizacyjnych we właściwym czasie. W bibliotekach tych nie zadziałał – ponieważ nigdy nie został uruchomiony – mechanizm *technology watch*, którego zadaniem jest przypomnienie we właściwym czasie o potrzebie odświeżenia lub zmiany generacji nośnika, zastosowania emulacji bądź migracji. Stało się tak, pomimo że wszyscy badani respondenci zgodzili się z opinią, iż materiał cyfrowy wymaga szczególnych warunków i metod przechowywania, w przeciwnym razie treści dokumentów zapisanych w cyfrowej postaci zostaną utracone. Istotne znaczenie ma dokładny opis dokumentu, w którym m.in. zawiera się informacja o typie nośnika, jego trwałości oraz planowej dacie testowego odczytu.

Jeden z respondentów przyznał wówczas, że w zasobach instytucji, w której pracuje, jest z pewnością sporo dokumentów cyfrowych, zwłaszcza starszych na dyskietkach 5,25 i 3,5 cala, z których odczytem z różnych powodów mogą być problemy i że należałoby przystąpić do ich diagnozy, jednak dopóki nie robią tego instytucje centralne i nie ma odgórnych wytycznych, takie działania prawdopodobnie nie zostaną podjęte. Tłumaczy się to brakiem środków. Tymczasem autor wypowiedzi zapewniał, że „diagnozę należałoby przeprowadzić nawet, jeśli nie ma na to środków".

W odniesieniu do systemu zarządzania dokumentami cyfrowymi nieliczni respondenci zwracali uwagę, że najważniejszą funkcją zarządzania jest zapewnienie odczytu zgromadzonych dokumentów i ich udostępnienie, najlepiej poprzez sieć. Dodatkowo zwracano uwagę na potrzebę stworzenia centralnego katalogu zasobów cyfrowych w skali globalnej i udostępnienie go online. Równie istotne znaczenie przypisano dokładnemu ujednoliconemu opisowi dokumentów cyfrowych z uwzględnieniem metadanych bibliograficznych, technicznych i administracyjnych oraz sporządzaniu kopii bezpieczeństwa.

W wypowiedziach pojawił się wątek potrzeby przechowywania oryginałów, nazywanych plikami macierzystymi, oraz ich kopii zapasowych w bezpiecznych repozytoriach cyfrowych. Zasugerowano możliwość skorzystania z oferty zewnętrznych usługodawców, którzy przejęliby zadania archiwizacji zasobów cyfrowych.

Częściej preferowano wówczas tworzenie lokalnych archiwów polskich zasobów cyfrowych zorganizowanych i działających na podstawie ujednoliconych zasad. Zdecydowanie rzadziej występowały opinie o utworzeniu centralnego narodowego archiwum gromadzącego kompletny zasób cyfrowy Polski. Argumentowano, że aby zbudować profesjonalny system archiwalny, zgodny ze światowymi standardami, potrzebna jest jednolitość procedur; w przypadku archiwów lokalnych nierzadko identyfikowano rozbieżności, które utrudniały funkcjonowanie jednorodnego, ale rozproszonego archiwum. Z badań wynikało zatem, że nie ma zgodności co do modelu archiwum polskiego dziedzictwa cyfrowego. Niewątpliwie już wtedy potrzebne było współdziałanie polskich instytucji pamięci, choćby właśnie po to, aby opracować i powołać powszechnie akceptowaną formę organizacji i funkcjonowania polskiego archiwum bez względu na jego model. Prace takie, na podstawie zgromadzonych wówczas odpowiedzi respondentów, powinna inicjować i koordynować w przypadku polskiego zasobu bibliotecznego Biblioteka Narodowa. W większości wypowiedzi deklarowano chęć przyłączenia się do grupy opracowującej strategię archiwizacji cyfrowego dziedzictwa Polski.

Próbowano zatem uzyskać opinię respondenta z Biblioteki Narodowej w kwestii ewentualnego podjęcia przez BN roli koordynatora działań archiwizacyjnych

w Polsce. Za bardziej kompetentne przedstawiciel BN uznał Ministerstwo Kultury i Dziedzictwa Narodowego, argumentując: „Polskie dziedzictwo cyfrowe obejmuje zbiory archiwalne, biblioteczne i muzealne, archiwa audiowizualne oraz zbiory należące do fundacji, stowarzyszeń, osób prywatnych, instytucji kościelnych. Koordynacja archiwizacji tak różnorodnego materiału wymaga znajomości potrzeb różnych instytucji pamięci w Polsce, a jednocześnie powinna dawać możliwości zharmonizowanego finansowania prac". Nasunął się zatem wniosek, że Biblioteka Narodowa z dużym prawdopodobieństwem byłaby właściwym w Polsce koordynatorem prac na rzecz długoterminowej archiwizacji elektronicznych zasobów bibliotecznych i takiej roli by się podjęła. Natomiast w odniesieniu do kompletnego dziedzictwa cyfrowego Polski należałoby powołać – co sugerował jeden z uczestników badania – „odrębną jednostkę, zrzeszającą pracowników instytucji gromadzących dokumenty elektroniczne (bibliotekarzy, archiwistów, muzealników) oraz informatyków". Nie jest wykluczone, że takie zrzeszenie działające przy właściwym ministerstwie i finansowane przez rząd mogłoby skutecznie koordynować działania archiwizacyjne.

Ciekawy i potwierdzający tezę o utożsamianiu w Polsce przez długi czas procesów długoterminowej archiwizacji głównie z budową polskich bibliotek cyfrowych okazał się wynik badania, dotyczący znajomości polskich inicjatyw na rzecz opracowania projektu długoterminowej archiwizacji polskiego dziedzictwa cyfrowego. Jako przykłady projektów archiwizacji polskich zbiorów cyfrowych zgodnie wymieniane były: Federacja Bibliotek Cyfrowych, kilka polskich bibliotek cyfrowych oraz Narodowe Archiwum Cyfrowe. Jeden z respondentów zwrócił uwagę na działalność Ministerstwa Kultury i Dziedzictwa Narodowego oraz Zespołu ds. Digitalizacji.

Prowadzone badanie miało też służyć poznaniu opinii pracowników polskich instytucji gromadzących i udostępniających zasoby cyfrowe, z jakich powodów długoterminowa archiwizacja została określona mianem wyzwania oraz na czym, ich zdaniem, polega trudność organizacji działań archiwizacyjnych w Polsce. Respondenci wymieniali najczęściej: brak środków finansowych, brak wiedzy i umiejętności odnośnie do reagowania na zmiany technologiczne oraz brak planu działania, w znaczeniu braku wzorców oraz zaniedbań organizacyjnych. Nieco rzadziej wymieniano: szybkość zmian technologicznych, różnorodność zbiorów, a także brak sprzętu i oprogramowania potrzebnych dla celów długoterminowej ochrony zbiorów cyfrowych.

Zauważono, że realizacja zadań związanych z archiwizacją zasobów elektronicznych wymaga zatrudnienia dodatkowej kadry, która mogłaby skupić się tylko na tych zadaniach. Rozumiano jednak, że jest to zależne od ograniczonych zwykle możliwości finansowych instytucji pamięci. Stąd zapewne na pierwszym miejscu

w prawie wszystkich odpowiedziach wymieniono brak środków finansowych jako czynnik utrudniający bądź uniemożliwiający podjęcie działań archiwizacyjnych w polskich instytucjach bibliotecznych i archiwalnych.

Wydawcy a trwała ochrona zasobów cyfrowych

W badaniu z 2007 r. pytano o ewentualną rolę wydawców publikacji cyfrowych w procesach długoterminowej archiwizacji. Respondenci bez wyjątku odpowiadali, że wydawcy mają lub mogą mieć wpływ na realizację procesów archiwizacji swoich produktów. W ich opinii wydawcy powinni publikować z uwzględnieniem standardów dotyczących stosowanych formatów, a także sprzętu i oprogramowania potrzebnego do odczytu publikacji. W jednej odpowiedzi proponowano, aby zobligować wydawców do dołączania oprogramowania niezbędnego do odczytu i prezentacji treści niestandardowych publikacji. Zwrócono również uwagę, że nie wszystkie dokumenty cyfrowe można odczytać po wykonaniu kopii zapasowej (dotyczy to zwłaszcza programów). Głównie bibliotekarze zgłaszali problem identyfikacji niektórych informacji potrzebnych do sporządzenia metadanych dla publikacji elektronicznych, stąd sugestia respondentów dotycząca większej staranności w tym zakresie, najlepiej ujednolicenia i konsekwentnego stosowania formy i miejsca umieszczania na publikacjach podstawowych metadanych. Proponowano także, aby wydawcy odsyłali publikacje do instytucji archiwalnych wraz z kompletem metadanych. Próby takie w niektórych bibliotekach na świecie podjęto, przy czym z inicjatywą wychodziły jednak instytucje biblioteczne, precyzując swoje oczekiwania wobec wydawców. Głównie dotyczy to elektronicznych publikacji sieciowych zgłaszanych przez wydawców do bibliotek na specjalnym formularzu – zgłoszenie wymaga uzupełnienia przez wydawcę określonych metadanych. W przypadku publikacji na nośnikach fizycznych pewnym rozwiązaniem i pomocą dla bibliotekarzy mogłyby być katalogi wydawnicze, w których wydawcy prezentowaliby nowo opublikowany dokument zgodnie z konwencją zaproponowaną przez instytucje archiwizujące.

Proponowano współpracę pomiędzy wydawcami oraz bibliotekarzami i archiwistami w zakresie przechowywania u wydawców tzw. matrycy („matki") publikacji elektronicznych, co dawałoby dodatkowe zabezpieczenie dokumentów. Deklarowano potrzebę współpracy na rzecz standaryzacji procesów wydawniczych, jednak z sugestią, że będzie to bardzo trudne w przypadku wydawnictw małych, działających nieprofesjonalnie.

Wg ustaleń z 2017 r., temat nadal jest aktualny. Problematyczne jednak okazują się koszty wytwarzania treści w specjalnych formatach dla archiwizacji długoterminowej. Wydawcy ponoszą już koszty wytwarzania dokumentów w określonych for-

matach charakterystycznych dla ich własnych systemów archiwizacji i dystrybucji. Nierozstrzygnięte pozostaje zatem pytanie, kto poniesie koszt dostosowania przez wydawców publikacji do wymogów archiwizacyjnych instytucji pamięci.

Wydawcy są świadomi dobrej jakości usługi archiwizacyjnej Biblioteki Narodowej i odpowiedniego zabezpieczenia ich produktów. Zgadzają się na archiwizację cyfrową i są gotowi przekazać egzemplarz obowiązkowy Bibliotece Narodowej w postaci elektronicznej dla celów archiwizacji. Uważają, że egzemplarz obowiązkowy powinien służyć tylko archiwizacji. Niechętnie natomiast odnosi się środowisko wydawców do oczekiwań licznych instytucji bibliotecznych dotyczących elektronicznych wypożyczeń w postaci wysyłania treści cyfrowej do czytelników. Ich zdaniem, „w Polsce niedostatecznie chroni się książki przed nadmiernym rozpowszechnianiem kopii i skanów – już poza zasięg użytku własnego". Z racji rozbieżności interesów instytucji wydawniczych z instytucjami rozpowszechniającymi publikacje, w tej sprawie trwa ciągły spór. Wprowadzony Public Lending Right i opłaty reprograficzne w znikomy sposób rekompensują autorom i wydawcom rozpowszechnianie ich treści bez opłat[3].

3.4.2. Rola twórców zasobów cyfrowych w procesie długoterminowej archiwizacji

Twórca dokumentów cyfrowych to każda osoba i/lub instytucja uczestnicząca w projektowaniu, tworzeniu oraz rozpowszechnianiu dokumentu cyfrowego, zanim zostanie on objęty ochroną w instytucji archiwizującej. Do kategorii twórców zalicza się głównie wydawców, ale także autorów, redaktorów, programistów, autorów projektów konwersji materiałów analogowych na format cyfrowy, a tym samym instytucje, w których takie projekty są realizowane [National Library of Australia, 2003, s. 89; Neuroth i in., 2009]. W modelu OAIS występuje dodatkowo termin *producent*, którym określa się osobę lub instytucję, zgłaszającą i transferującą dokument cyfrowy do systemu archiwalnego [OAIS, 2002].

Z doświadczeń niemieckiej biblioteki narodowej w zakresie gromadzenia i archiwizacji cyfrowych rozpraw doktorskich i habilitacyjnych oraz elektronicznych czasopism naukowych wynika, że niebagatelne znaczenie dla procesów gromadzenia i efektywnego zarządzania tymi dokumentami jako obiektami archiwalnymi ma współpraca twórców z pracownikami instytucji archiwizujących. Na mocy

[3] Na podstawie rozmów telefonicznych oraz wymiany korespondencji z przedstawicielem Polskiej Izby Książki oraz dyrektorem redakcji publikacji elektronicznych jednej z dużych polskich firm wydawniczych na temat współpracy i relacji pomiędzy środowiskiem instytucji pamięci i instytucji wydawniczych. Materiał pochodzi z 2017 r.

porozumień między biblioteką narodową a rozmaitymi instytucjami tworzącymi i publikującymi zasoby cyfrowe biblioteka – jako centralna instytucja archiwizująca – rejestruje i gromadzi niemieckie dziedzictwo cyfrowe oraz, poprzez stały kontakt z twórcami, dysponuje informacjami, które znacznie ułatwiają zarządzanie publikacjami cyfrowymi. Biblioteka niemiecka wywiera także wpływ na standardy tworzenia i publikowania cyfrowych rozpraw doktorskich i habilitacyjnych. Publikuje poradniki oraz organizuje szkolenia z zakresu publikowania elektronicznego [Dissonline.de, 2009].

Potrzeba współpracy motywowana jest przez bibliotekarzy i archiwistów przede wszystkim faktem, że twórcy nie zawsze biorą pod uwagę potencjalną przynależność powstającego dokumentu do dziedzictwa cyfrowego, a także potrzebę jego dostępności i użytkowania w przyszłości. Twórcom może też zabraknąć wiedzy i narzędzi, których zastosowanie mogłoby ułatwić zabezpieczenie i ochronę danych [National Library of Australia, 2003, s. 89]. Ponadto celem współpracy miałaby być redukcja równoległego występowania różnych pod względem formy obiektów cyfrowych w środowiskach edukacyjnych i naukowych, a także w bibliotekach i archiwach. Instytucje biblioteczne i archiwalne inicjują współpracę z twórcami, aby uświadomić im wartość i potrzebę ochrony tworzonych przez nich dokumentów oraz wskazać, jak istotny jest ich udział w tym procesie, głównie poprzez stosowanie standardowych, powszechnie znanych i dostępnych technik tworzenia materiałów cyfrowych.

Ze względu na fakt, że instytucje pamięci rozmaicie radzą sobie z identyfikacją dokumentów cyfrowych (zwłaszcza sieciowych), współpraca powinna służyć procesom gromadzenia i tworzenia możliwie kompletnych kolekcji, przynajmniej do czasu, kiedy powstaną i zaczną obowiązywać odpowiednie przepisy prawne, na mocy których wszyscy twórcy będą zobligowani do zgłaszania i odsyłania swoich materiałów do instytucji archiwizujących.

Należy pamiętać, że współpraca z twórcami wiąże się jednak z szeregiem barier. Są to osoby rozmaitych profesji, różniące się podejściem do pracy, skalą działalności, możliwościami organizacyjnymi i technicznymi, przede wszystkim zaś zainteresowaniem długoterminową użytecznością swoich produktów. W tak heterogenicznej grupie rozmaite bywa nastawienie do prób „przejęcia" wyników pracy przez programy ochrony [National Library of Australia, 2003, s. 90, 92]. Rozmaicie jest traktowana propozycja akceptacji i stosowania procedur tworzenia oraz zapisu publikacji cyfrowych – zwykle niechętnie. Niezależnie jednak od barier zaleca się, aby wszelkie programy ochrony dziedzictwa cyfrowego dążyły do możliwie wczesnego wywierania wpływu na praktyki tworzenia zasobów cyfrowych i zarządzania nimi. Podejmowanie takiej inicjatywy w momencie pojawienia się pierwszych problemów z użytkowaniem dokumentów może okazać się kosztowne i prowadzić nawet do ich utraty [National Library of Australia, 2003, s. 90].

W badaniach prowadzonych w latach 2007-2009 próbowano poznać stanowisko polskich wydawców publikacji elektronicznych dotyczące trwałej ochrony ich użyteczności [Januszko-Szakiel, 2011a, s. 21-46]. Przedstawicieli instytucji pamięci pytano, czy ich zdaniem twórcy dokumentów cyfrowych mogą być w jakikolwiek sposób pomocni, mają lub mogą mieć wpływ na realizację procesów długoterminowej archiwizacji swoich produktów. Nieliczni respondenci stali na stanowisku, że twórcy nie mają wpływu na realizację procesów długoterminowej archiwizacji dokumentów cyfrowych, bowiem procesy te raczej zależne są od instytucji gromadzących i przechowujących. Większość ankietowanych potwierdziła zasadność udziału twórców zasobów cyfrowych w procesach archiwizacji, a udział ten mógłby realizować się w następujących obszarach:

- wytwarzanie obiektów cyfrowych w specjalnej wersji do archiwizacji i przekazywanie ich instytucjom pamięci (takim jak muzea, archiwa i biblioteki); pracownicy instytucji pamięci nie musieliby konwertować tych obiektów, co wiązałoby się z oszczędnościami;
- odprowadzanie elektronicznego egzemplarza obowiązkowego oraz współpraca wydawców z bibliotekarzami przy budowaniu bibliotek cyfrowych;
- umożliwienie nabywcom dokumentów cyfrowych ich kopiowania na inne nośniki (dla potrzeb własnych) oraz bezpłatnie lub za minimalną opłatą dostarczanie użytkownikom nowych wersji dokumentów;
- okresowe aktualizowanie dokumentów cyfrowych zgodnie ze zmianami technologicznymi.

Dodatkowe informacje na temat stanowiska wydawców uzyskano wówczas również od przedstawicieli Polskiego Towarzystwa Wydawców Książek oraz Polskiej Izby Książki. W wyniku przeprowadzonych rozmów okazało się, że temat długoterminowej ochrony publikacji elektronicznych nie był w tych instytucjach, do momentu podjęcia omawianych badań, wywoływany. Stanowił wówczas dla rozmówców zagadnienie nowe. Nie spotkano się również z głosami wydawców, które świadczyłyby, że trwała archiwizacja ich elektronicznych produktów powinna stać się przedmiotem dyskusji.

W Polsce przeważało wówczas publikowanie tradycyjne, tj. w formie druku na papierze. Przeprowadzono 58 rozmów z wydawcami. Tylko 12 odpowiedziało jednoznacznie twierdząco, że uważają się za wydawców publikacji elektronicznych. Jednak nawet w największych firmach wydawniczych, które miały rozbudowane działy publikacji elektronicznych i publikowały dokumenty na nośnikach przenośnych oraz w sieci, publikowanie elektroniczne było działalnością marginalną, aczkolwiek traktowaną poważnie. Procesy publikowania elektronicznego występowały sporadycznie i najczęściej nie były to publikacje samoistne wydawniczo, lecz materiały uzupełniające do publikacji drukowanych.

Z rozmów wynikało, że wydawnictwa uczelni wyższych w zasadzie nie publikują w wersji cyfrowej. W dwóch wydawnictwach odnotowano wówczas potrzebę dołączenia do publikacji drukowanej pliku na płycie CD-ROM. Były to rozprawy doktorskie z dużym objętościowo materiałem, który stanowił część aneksu. Znacznie częściej to biblioteki uczelniane podejmowały próby publikowania elektronicznego w związku z pracami digitalizacyjnymi oraz tworzonymi repozytoriami cyfrowych materiałów dydaktycznych i publikacji pracowników nauki. Interesujące okazało się podejście wydawców podręczników i materiałów elektronicznych, głównie w postaci systemów multimedialnych do nauczania języków obcych. Wydawcy tłumaczyli, że nie zaliczają siebie do wydawców publikacji elektronicznych, gdyż publikują podręczniki głównie w wersji drukowanej, a dodatki w postaci płyt CD-ROM do podręczników bądź samoistne programy multimedialne do nauczania zamawiają u zewnętrznych producentów – oni pełnią tylko rolę dystrybutora.

Bardzo podobnym podejściem charakteryzowało się wielu wydawców wyszczególnionych w bazie BN. Twierdzili, że opublikowanie kilku zaledwie dokumentów w postaci elektronicznej, głównie na płytach CD, nie czyni ich wydawcami publikacji elektronicznych.

Wydawcy nie byli też świadomi istnienia bazy BN, w której byli ujęci. Obecnie sytuacja zmieniła się. Środowisko wydawnicze deklaruje dobrą komunikację z BN.

Niektóre wydawnictwa deklarowały, że opublikowały więcej elektronicznych dokumentów niż odnotowano w bazie BN. Mogło to wynikać z zaniedbania obowiązku wysyłania do bibliotek egzemplarza obowiązkowego bądź z polityki BN dotyczącej prowadzenia bazy wydawnictw elektronicznych. Z rozmowy przeprowadzonej z pracownikami BN, zajmującymi się wówczas bazą wydawnictw elektronicznych, wynikało, że baza była uaktualniana na bieżąco, na podstawie dostarczanych egzemplarzy obowiązkowych.

Respondent jednego z analizowanych wówczas wydawnictw zdecydowanie zgodził się z opinią, że zastosowanie przez wydawców odpowiedniego nośnika zapisu publikacji może mieć wpływ na efektywność procesu długoterminowej archiwizacji: „wybór odpowiedniego nośnika jest niezwykle istotny w czasach bardzo dynamicznych zmian w dziedzinie technologii cyfrowych". Decydując się na zapis dokumentu w odpowiednim formacie i na nośniku, wydawnictwo kierowało się głównie dwoma kryteriami: „ceną i powszechnością korzystania z konkretnego nośnika".

Z badań wynikało, że jakiekolwiek formy współpracy instytucji wydawniczych z bibliotekami i archiwami w zakresie archiwizacji publikacji elektronicznych nie były podejmowane. Brakowało jednoznacznej deklaracji chęci współpracy. Zdaniem jednego z respondentów „na pewno istotnym czynnikiem decydującym o podjęciu takiej współpracy byłyby ewentualne koszty dostosowania procesu wydawniczego do nowych standardów". Istotny jest jednak fakt, że wydawca takiej

współpracy nie wykluczył. (W rozmowie z 2017 r. prowadzonej w środowisku wydawców podreślono wielokrotnie, że poza Biblioteką Narodową nadal brakuje gotowości współpracy wydawców z bibliotekarzami i archiwistami).

Jedna z badanych firm wydawniczych opublikowała pierwsze dokumenty elektroniczne znacznie wcześniej, bo w 1990 r. Publikacje ukazały się na dyskietkach 5,25 cala i przyznano, że nie jest możliwy odczyt ich treści. Wydawca zgodził się również z opinią, że opublikowanie dokumentu w odpowiednim formacie i na odpowiednim nośniku może mieć wpływ na zachowanie jego użyteczności w dłuższym czasie, a decydując się na format i nośnik zapisu publikacji, podstawowym kryterium jest dostępność technologii.

W odpowiedzi na pytanie dotyczące formy współpracy z instytucjami bibliotecznymi i archiwalnymi firma deklarowała, że: „przekazuje egzemplarze nowo wydanych tytułów do Biblioteki Narodowej i Biblioteki Jagiellońskiej zgodnie z ustawą o przekazywaniu egzemplarzy obowiązkowych". Bez odpowiedzi natomiast pozostało pytanie dotyczące możliwości dostosowania procesu wydawniczego do wytycznych bibliotekarzy i archiwistów, w celu nadania publikacjom parametrów umożliwiających ich długoterminową archiwizację.

Respondent deklarował, że wydawcy mogliby tworzyć wirtualne serwery dla publikacji elektronicznych. Pomysł niewątpliwie interesujący, jednak pozostał bliżej nieokreślony.

Podobną koncepcję archiwizacji publikacji elektronicznych proponował przedstawiciel firmy wydawniczej: „wspaniałym archiwum dla publikacji elektronicznych wydaje się obecnie przestrzeń Internetu. Oczywiście jest to rozwiązanie niedoskonałe pod wieloma względami, choćby z uwagi na brak kompletności, wiarygodności, systematyki. Podstawowymi zaletami Internetu jest ogólnodostępność, brak ograniczeń co do objętości, niski koszt umieszczania w Internecie publikacji i najważniejsza zaleta – publikacja przechowywana jest «wiecznie». Co dzisiaj trafi do sieci, jest przechowywane jako kopia, nawet po usunięciu pierwotnego pliku na wielu serwerach (choćby wyszukiwarka google). Pozostaje oczywiście kwestia kompletności i autentyczności tak «przechowywanych» dokumentów. Do rozwiązania byłby więc problem, jak przy wykorzystaniu opisanej wyżej cechy Internetu, zbudować system dostępu do danych z jakimś systemem oceny wiarygodności źródła". Pomimo wielokrotnie kierowanych próśb do wydawcy, nie udało się uzyskać dookreślenia tego pomysłu.

Na podstawie badań wysunięto wniosek, że w polskich firmach wydawniczych temat ochrony publikacji elektronicznych do 2009 r. nie był podejmowany, a jedyna forma współpracy z instytucjami bibliotecznymi polegała na przekazywaniu egzemplarza obowiązkowego na podstawie ustawy o obowiązkowych egzemplarzach bibliotecznych. Wydawcy deklarowali jednak gotowość zapoznania się z koncepcją

instytucji archiwizujących w tej sprawie i przyłączenia się do działań na rzecz opracowania strategii archiwizacji. Z rozmów prowadzonych w 2017 r. wynika, że omawiany stan rzeczy w zasadzie nie zmienił się. Rozmówcy zwrócili wprawdzie uwagę, że od niedawna zaczynają obowiązywać pewne standardy cyfrowych treści (ePub itp.). Wydawcy archiwizują indywidualnie, a dzieje się tak z racji braku zaufania do instytucji pamięci. „Brak poszanowania prawa i własności intelektualnej jest w Polsce powszechny i bardzo, bardzo zakorzeniony. Stąd brak zaufania…". Przyznano, że wciąż brakuje w środowisku wydawców debaty na temat długoterminowej archiwizacji zasobów cyfrowych. Inne problemy były podejmowane; w centrum uwagi środowiska książki w ostatnich latach znalazły się m.in. kwestie kryzysu w dystrybucji książki, piractwa, zanikania księgarń, wojen rabatowych, zbyt niskich marż.

Współpraca wydawców i pracowników instytucji pamięci prawdopodobnie miałaby w naszym kraju szanse powodzenia, jednak nie ulega wątpliwości, że z inicjatywą powinny wyjść instytucje pamięci. Twórcy dokumentów cyfrowych prawdopodobnie nie dostosują procesów wydawniczych do potrzeb instytucji pamięci, jeśli te nie będą zgłaszać takiej potrzeby oraz nie ustalą wytycznych (parametrów dokumentów), które twórcy mieliby w swej działalności uwzględnić. Nade wszystko rozstrzygnięcia wymaga temat kosztów takiego dostosowania.

Przykładem powodzenia współpracy instytucji archiwizującej z twórcami i dystrybutorami zasobów cyfrowych jest przytaczany już program niemieckiej biblioteki narodowej. Początkowo reakcja niemieckich wydawców na impuls biblioteki, który miał ich zachęcić do współpracy i uświadomić wagę problemu, była bardzo słaba. Jednak z czasem biblioteka narodowa przekonała twórców, głównie wydawców elektronicznych czasopism naukowych, o zasadności przekazania cyfrowych zasobów pod fachową opiekę. Biblioteka przejęła na siebie zadanie uświadomienia wydawcom, co oznacza i z jakimi organizacyjnymi, technicznymi oraz finansowymi trudnościami wiąże się długoterminowe przechowanie zasobów cyfrowych. Poprzez specjalne pisma do wydawców, rozmowy, negocjacje, zaproszenia do udziału w organizowanych przez bibliotekę konferencjach czy warsztatach Biblioteka Narodowa pracowała na rzecz przekonania twórców, że ich udział w procesie gromadzenia i archiwizacji niemieckich zasobów cyfrowych jest niezbędny [Fakten + Zalen, 2008].

W związku ze zmianami w branży wydawniczej wywołanymi upowszechnieniem się publikowania elektronicznego i oczekiwaniami wobec twórców zasobów cyfrowych należy mieć na uwadze następujące obszary procesów gromadzenia i archiwizacji, w których uczestnictwo twórców zasobów cyfrowych jest istotne [Müller, 1998, passim]:

- twórcy stanowią wiarygodne źródło kompletnej i aktualnej informacji o cyfrowych nowościach wydawniczych; z ich udziałem instytucje pamięci są w stanie tworzyć i dysponować kompletnymi wykazami publikacji cyfrowych, np. dla

potrzeb tworzenia bibliografii narodowych. Szczególne znaczenie miałoby to w przypadku zasobów sieciowych, gdyż te pozostają poza kontrolą instytucji archiwizujących, przynajmniej obecnie w naszym kraju. Twórcy mogliby również stanowić źródło ważnych informacji o wartości, znaczeniu poszczególnych dokumentów dla nauki i edukacji, o potencjalnej przynależności dokumentu do narodowego dziedzictwa nauki i kultury, usprawniając tym samym procesy oceny i selekcji, czyli tworzenia kolekcji archiwalnych;

- twórcy stanowią kompetentne źródło informacji o dokumentach (formatach, nośnikach, kodach źródłowych, autorach, współtwórcach, celu powstania, oczekiwanej funkcji dokumentu czy jego dotychczasowej historii), które w istotny sposób usprawniają procesy archiwizacji i powinny być zapisane jako element metadanych;
- twórcy mogą pełnić istotną funkcję dostawcy dokumentu do instytucji archiwizujących;
- twórcy mogą generować oraz dostarczać metadane do instytucji archiwizujących; w przypadku gdy z różnych powodów twórca nie przekazuje dokumentu do instytucji archiwizującej, może udostępnić dane o jego istnieniu uzupełnione o abstrakt (typowy przykład bazy bibliograficzno-abstraktowej), jednak wówczas twórca powinien samodzielnie zatroszczyć się o ochronę swoich zasobów cyfrowych;
- twórcy, podczas wznawiania publikacji cyfrowych czy publikacji drukowanych zawierających dodatki w postaci dokumentów cyfrowych, mogą uaktualniać formaty i nośniki, umożliwiając ich odczyt przy użyciu powszechnego w danym czasie sprzętu i oprogramowania.

Wymienione funkcje prawdopodobnie nie wyczerpują obszaru aktywności wydawców w procesach archiwizacji materiałów cyfrowych, ale nawet w takim ujęciu ich pomoc okazałaby się z pewnością przydatna.

Bez względu na zakres współpracy instytucje archiwalne i tworzące dokumenty powinny ustalić procedury jej funkcjonowania. Przykładowe kwestie porozumienia to [Coy, 2006; Kriterienkatalog, 2006]:

- bieżąca komunikacja – zaleca się korespondencję elektroniczną (rzadziej tradycyjną), rozmowy telefoniczne, okresowe spotkania;
- terminowość, rodzaj i sposób dostarczania zasobów do archiwum (przenośne nośniki danych, sieciowy transfer danych, upload, download), a także ich metadanych. Najpoważniejszy problem w tym zakresie dotyczy sposobu transferu zasobów sieciowych z instytucji tworzącej do archiwizującej. W piśmiennictwie przedmiotu wymieniane są dwa możliwe rozwiązania: *push* i *pull*. *Push* oznacza, że twórca przekazuje dokument na serwer archiwum, *pull* natomiast, że archiwum pobiera dokument z serwera twórcy. Z technicznego

punktu widzenia w zasadzie nie ma specjalnych różnic pomiędzy tymi metodami i każda z nich może być z powodzeniem stosowana. Ważne jest to, aby po dotarciu dokumentu do archiwum, dokonana została kontrola jakości zgodnie z założeniami przyjętymi dla pakietów przychodzących. Metoda *pull* wiąże się z większym zaangażowaniem archiwum w procesy gromadzenia, metoda *push* z kolei bardziej aktywizuje w tych procesach twórców. Preferencje twórców zazwyczaj nakierowane są na metodę *pull*, która minimalizuje ich nakład we współpracy z archiwami [Fülle i Ott, 2006];

- przejęcie obowiązku ochrony przez instytucję archiwizującą (określenie momentu rozpoczęcia i czasu trwania ochrony zasobów w archiwum, zastosowanych metod archiwizacji oraz innych znaczących kwestii, np. prawa do wykonywania czynności konserwatorskich);
- kompatybilność stosowanych formatów, sprzętu i oprogramowania, zapisu w sieci, transfer metadanych, a także techniczne i prawne możliwości przekopiowywania oraz przekodowywania.

Dla efektywności procesów archiwizacji istotne jest zatem nakłonienie producentów wyposażenia i oprogramowania oraz twórców, wydawców i dystrybutorów materiałów cyfrowych do podjęcia współpracy na rzecz ochrony dziedzictwa cyfrowego z bibliotekami, archiwami, muzeami i innymi instytucjami pamięci. Z uwagi na fakt, że nie tylko wydawcy, ale również autorzy oraz instytucje naukowe i badawcze samodzielnie publikują i rozpowszechniają dokumenty cyfrowe, pojawia się potrzeba, aby zarówno brali oni odpowiedzialność za jakość merytoryczną publikowanych treści, jak i dbali o formę ich zapisu, która zapewniałaby możliwość ich długoterminowej archiwizacji w depozytowych systemach archiwalno-bibliotecznych. Rudymentarne znaczenie ma ustalenie standardów tworzenia pierwotnych dokumentów cyfrowych oraz digitalizatów, następnie szkolenia i formy kontroli zgodności produktów cyfrowych z przyjętymi normami [Coy, 2006].

Warto zwrócić też uwagę na podjętą przez wydawców polskich inicjatywę dotyczącą obowiązkowego egzemplarza elektronicznego. Pierwsze posiedzenie powołanej tego samego dnia Sekcji Publikacji Elektronicznych Polskiej Izby Książki (SPE PIK)[4] odbyło się 20 kwietnia 2009 r. w Biurze Polskiej Izby Książki w Warszawie. Sekcję utworzyły następujące firmy wydawnicze: Dom Wydawniczy MEDIUM, Young Digital Planet, SW Czytelnik, Key TExt, Biblioteka Analiz, LexisNexis Polska, Wydawnictwa Szkolne i Pedagogiczne, Wydawnictwo Na-

[4] W 2016 r. SPE PIK została zastąpiona przez Sekcję Dystrybucji Treści, koncentrując swoją uwagę i prace na zagadnieniach doskonalenia rynku dystrybucji książki w Polsce. Szczegóły na stronie PIK, pod adresem: http://pik.org.pl/pages/90/sekcja-dystrybucji-tresci [Dostęp: 29.08.2017].

ukowe PWN, Novae Res – Wydawnictwo Innowacyjne, ZamKor P. Sagnowski i Wspólnicy, Woters Kluwer Polska, Porozumienie Wydawców, IPS International Publishing Service, NetPress Digital.

Celem utworzenia Sekcji było podjęcie działań dla ochrony interesów wydawców publikacji elektronicznych, w tym powołanie stowarzyszenia, które powinno posiadać prawa organizacji zbiorowego zarządu prawami autorskimi w odniesieniu do publikacji elektronicznych. Planowano również cykl szkoleń dla członków PIK w dziedzinach związanych z publikacjami elektronicznymi.

Na posiedzeniu Sekcji Publikacji Elektronicznych PIK 17 lutego 2010 r. jej członkowie podjęli dyskusję na temat projektu zastąpienia dotychczasowo ustawowego przepisu o egzemplarzu obowiązkowym (zgodnie z którym każdy wydawca powinien nieodpłatnie przekazywać 17 egzemplarzy każdej swojej publikacji wyznaczonym w ustawie bibliotekom) przepisem, w myśl którego liczba egzemplarzy przekazywanych nieodpłatnie do bibliotek zostałaby znacznie zmniejszona, a równocześnie wprowadzony zostałby obowiązek przekazywania Bibliotece Narodowej elektronicznego egzemplarza obowiązkowego każdej publikacji. Członkowie Sekcji odnieśli się krytycznie do tego projektu i poparli oświadczenie Sekcji opublikowane w formie Komunikatu PIK z 5 lutego 2010 r., którego podstawowe stwierdzenie brzmi: „Sekcja Publikacji Elektronicznych Polskiej Izby Książki stanowczo sprzeciwia się wprowadzeniu obowiązkowego egzemplarza elektronicznego w miejsce dotychczasowego egzemplarza papierowego". Oświadczenie Sekcji zostało skierowane m.in. do Bogdana Zdrojewskiego, ówczesnego ministra kultury i dziedzictwa narodowego. Swoje stanowisko Sekcja miała okazję przedstawić 15 marca 2010 r. w Bibliotece Narodowej. Podczas spotkania w BN zostało zaprezentowane stanowisko zarówno bibliotekarzy, jak i wydawców w sprawie zmian w ustawie o obowiązkowych egzemplarzach z 1996 r. Temat był kontynuowany. W 2013 r. zamieszczono na stronie WWW Polskiej Izby Książki informacje świadczące jednoznacznie o zrozumieniu tematu zachowania dziedzictwa nauki i kultury w środowisku wydawców oraz potrzeby współpracy w tym zakresie z instytucjami pamięci [Mendruń, 2013]. Z dokumentów udostępnionych przez Polską Izbę Książki wynika, że w 2014 r. odbyło się spotkanie przedstawicieli MKiDN, MNiSW, przedstawicieli wydawców, przedstawicieli Konferencji Dyrektorów Bibliotek Akademickich Szkół Polskich, Konferencji Dyrektorów Wojewódzkich Bibliotek Publicznych, Stowarzyszenia Bibliotekarzy Polskich, Krajowej Rady Bibliotecznej. Podczas spotkania przyjęto ustalenia, wg których „egzemplarz obowiązkowy służy do archiwizacji dorobku wydawniczego kraju, nie służy natomiast do budowania zbiorów bibliotek". Elektroniczny Egzemplarz Obowiązkowy jest takim samym produktem jak egzemplarz obowiązkowy papierowy. Nie jest bezkosztowy. Dlatego jedyną aprobowaną przez wydawców formą jego udostępniania jest udostępnianie w systemie

Academica (Cyfrowa Wypożyczalnia Publikacji Naukowych Academica), zapewniającym odpowiednie standardy bezpieczeństwa (w tym dedykowane komputery).

Z corocznych sprawozdań Biblioteki Narodowej wynika, że polscy wydawcy odsyłają publikacje cyfrowe do Repozytorium Dokumentów Elektronicznych, które zostało uruchomione przez Zakład Technologii Informatycznych BN w 2009 r. Repozytorium BN „składa się z dwóch przystosowanych systemów DSpace: zewnętrznego dostępnego dla wydawców i wewnętrznego przechowującego dane. W BN gromadzone są za jego pomocą wydawnictwa zwarte i ciągłe, publikowane wyłącznie w wersji cyfrowej i udostępniane przez Internet. Dokumenty elektroniczne nadsyłane są dobrowolnie przez wydawców w ramach egzemplarza obowiązkowego. Zgodnie z Rozporządzeniem Ministra Kultury i Sztuki z dnia 6 marca 1997 r. w sprawie wykazu bibliotek uprawnionych do otrzymywania egzemplarzy obowiązkowych poszczególnych rodzajów publikacji oraz zasad ich przekazywania [Ustawa, 1997], w Polsce prawo do egzemplarza obowiązkowego dokumentów elektronicznych przysługuje dwóm bibliotekom: Bibliotece Jagiellońskiej i Bibliotece Narodowej. Dla każdego wydawnictwa zgłaszającego się do Repozytorium zostaje utworzone konto z odseparowanym zasobem, na które wydawcy samodzielnie przekazują publikacje. Bibliotekarze cyfrowi dokonują korekty wprowadzanych metadanych dokumentów i ewentualnie je uzupełniają. Dla każdego zarchiwizowanego tytułu opracowywany jest rekord bibliograficzny w katalogu komputerowym wraz z adnotacją o lokalizacji dokumentu. W 2016 r. w Repozytorium zamieszczono 10 302 obiekty. Zintensyfikowano działalność informacyjną dla wydawców związaną z przeniesieniem kont do nowego Repozytorium Cyfrowego BN. W 2016 r. w Repozytorium założono 921 nowych kont dla wydawców. W maju została stworzona wstępna wersja instrukcji obsługi Repozytorium Cyfrowego BN dla wydawców. Rozwijana jest współpraca z wydawcami, którzy zobligowani są do przekazywania do BN egzemplarza obowiązkowego dokumentów elektronicznych (pomoc techniczna, udzielanie informacji, monitorowanie wpływu). Wszystkie publikacje są obecnie dostępne za pośrednictwem Repozytorium, a cały bieżący wpływ wprowadzany bezpośrednio przez wydawców do Repozytorium" [Sprawozdanie, 2016, s. 81-82].

3.5. Rola kompetencji w realizacji zadań trwałej ochrony zasobów cyfrowych

Bardzo istotne okazuje się przeciwdziałanie zjawisku odsuwania w czasie procesów archiwizacji zbiorów cyfrowych. W wielu krajach wynikało ono z przekonania, że bez kompleksowego programu, odgórnych wzorców i wytycznych nie

można chronić zasobów cyfrowych. Tymczasem punktem wyjścia dla zachowania zasobów okazywało się podjęcie decyzji, przez każdą instytucję indywidualnie, o przyjęciu odpowiedzialności za ochronę [National Library of Australia, 2003, s. 31]. Oznaczało to samodzielność w studiowaniu problemu i poznawaniu istniejących rozwiązań, a także podejmowanie prób ich zastosowania w celu ochrony własnych zasobów. Taka indywidualna strategia w wielu przypadkach miała charakter wyrywkowy i dotyczyła najistotniejszych fragmentów procesu archiwizacji, np. wyboru właściwego formatu zapisu danych cyfrowych lub ustalenia częstotliwości odczytu dokumentów w celu identyfikacji ewentualnych problemów ich użytkowania; nie musiała stanowić kompleksowego i niezawodnego programu.

Zaznaczono wprawdzie, że najbardziej pożądane są programy kompleksowe i niezawodne, lecz – jeśli spełnienie tych warunków nie jest możliwe – lepsze będzie działanie cząstkowe i niegwarantujące pełnej skuteczności niż żadne [National Library of Australia, 2003, s. 31]. Oczekiwanie z podjęciem działań na opracowanie i wdrożenie strategii kompleksowej może oznaczać utratę materiału [National Library of Australia, 2003, s. 34]. Praca metodą „małych kroków" byłaby też w Polsce wskazana z racji ograniczonych zazwyczaj zasobów finansowych na projekty z sektora kultury i nauki. Łatwiej bowiem wyasygnować lub pozyskać fundusze na małe, cząstkowe przedsięwzięcia niż sfinansować implementację całościowego programu. Ponadto łatwiej zapanować nad organizacją wdrożenia pojedynczego modułu i sprawnym jego funkcjonowaniem niż uporać się z całokształtem programu jednocześnie.

Posiadanie wiedzy o tego typu rozwiązaniach i znajomość właściwego podejścia do zadań archiwizacji zasobów cyfrowych mogłyby stanowić poważne przesłanki do podjęcia przez polskich archiwistów i bibliotekarzy odpowiedzialności za długoterminową archiwizację narodowego dziedzictwa cyfrowego. Zdobywanie wiedzy wymaga jednak dostępności jej źródeł oraz wszelkich form propagowania, tj.: szkoleń, studiów, seminariów, konferencji, itp.

Z przywoływanych wcześniej badań jednoznacznie wynikała potrzeba upowszechniania zagadnień związanych z długoterminową archiwizacją zasobów cyfrowych. Zauważono potrzebę rozszerzenia oferty edukacyjnej o moduły przygotowujące specjalistów do zadań związanych z ochroną cyfrowych zasobów. Fakt ten ewidentnie dowodzi, że pracownicy polskich instytucji pamięci byli świadomi ewolucji standardów i praktyk funkcjonowania w swoim zawodzie, a tym samym dostrzegali potrzebę stałego uzupełniania i aktualizacji posiadanej wiedzy. Wychodząc tym potrzebom naprzeciw, w instytucjach kształcących tworzono nowe oraz modelowano istniejące już programy edukacyjne tak, aby dostosować np. studia z zakresu informacji naukowej i bibliotekoznawstwa, także archiwistyki, do zmieniających się realiów funkcjonowania ośrodków informacji i dokumentacji

w świecie. Lista tematów korespondujących z długoterminową archiwizacją materiałów cyfrowych, które powinny zaistnieć w programach studiów z tego zakresu, jest bardzo długa, a są to m.in.: tworzenie i zapis dokumentów cyfrowych, formaty zapisu danych, nośniki danych cyfrowych, formaty metadanych, wiarygodne i certyfikowane archiwa cyfrowe, model OAIS – Open Archival Information System, infrastruktura organizacyjna w procesie długoterminowej archiwizacji, identyfikatory stałe – Persistent Identifier, prawne zagadnienia długoterminowej ochrony obiektów cyfrowych, inicjatywy Open Access, techniczne aspekty długoterminowej archiwizacji, tj. emulacja, migracja, zabezpieczanie danych, tworzenie cyfrowych magazynów, repozytoriów i szereg innych [Neuroth i in., 2008].

Już wiele lat temu w rodzimej literaturze przedmiotu publikowano opinie specjalistów o potrzebie zmian w modelu kształcenia zarówno bibliotekarzy i specjalistów informacji, jak i archiwistów. Powinny w nim zostać uwzględnione treści związane z digitalizacją oraz ochroną zbiorów [Pindlowa, 2004, s. 63-66]. Pisano także o konieczności permanentnej edukacji, niezbędnej dla sprawnego funkcjonowania w społeczeństwie informacyjnym (społeczeństwie wiedzy), także wobec potrzeby dostosowania się do standardów Unii Europejskiej [Kocójowa, 2004, s. 67-77].

Od 2008 r. w polskich ośrodkach kształcenia z zakresu nauki o książce, bibliotece i informacji naukowej, a także z zakresu archiwistyki w programy studiów są wkomponowane przedmioty bezpośrednio lub pośrednio powiązane z zagadnieniami archiwistyki cyfrowej. Wybrane przykłady to: Zarządzanie archiwizowaniem w Internecie, Archiwa elektroniczne i ochrona zasobów archiwalnych, Archiwizacja i zabezpieczenie danych, Przechowywanie i ochrona archiwaliów, Zachowanie dziedzictwa kulturowego: książka, biblioteka, archiwum, Wstęp do archiwizacji i cyfryzacji, Organizacja i zarządzanie zasobami cyfrowymi, Typologia dokumentów elektronicznych, Przygotowanie publikacji cyfrowych, Digitalizacja dziedzictwa narodowego, Archiwa w Internecie, Archiwa, biblioteki, media i muzea w Internecie, Standaryzacja w usługach informacyjnych, Obiekty i kolekcje cyfrowe.

Wszelkie formy zdobywania wiedzy teoretycznej oraz umiejętności praktycznych z zakresu zarządzania i długoterminowej ochrony zasobów cyfrowych są na polskim rynku szkoleniowym potrzebne. W raporcie MKiDN z 2009 r. podkreślono bowiem potrzebę wykształcenia wysokiej klasy specjalistów do spraw digitalizacji i ochrony zasobów cyfrowych. Zwrócono uwagę na konieczność kształcenia kadry w ramach cyklicznych i kilkustopniowych kursów dla pracowników archiwów, bibliotek i muzeów oraz archiwów audiowizualnych, a także w ramach zajęć na uczelniach wyższych w formie kursów specjalizacyjnych bądź studiów uzupełniających dla studentów, którzy potencjalnie stanowić będą kadrę instytucji kultury. W kształceniu szczególny nacisk należy położyć m.in. na przy-

bliżenie idei oraz celowości budowy i przechowywania dziedzictwa narodowego w postaci cyfrowej oraz upowszechniania takich pojęć jak: archiwum cyfrowe, biblioteka cyfrowa, wirtualne muzeum jako integralne części nowoczesnej instytucji kultury. W ministerialnym raporcie ogłoszono potrzebę rozbudowywania działalności edukacyjnej w polskich instytucjach kultury [Program digitalizacji, 2009, s. 37, 50].

Procesy długoterminowej archiwizacji zasobów cyfrowych nabrałyby w Polsce rozmachu dzięki powołaniu stałej konferencji o charakterze fachowych debat, zwoływanej w określonych odstępach czasu. Konferencje takie powinny służyć jako merytoryczne zaplecze dla prac na rzecz opracowania strategii długoterminowej ochrony narodowego dziedzictwa cyfrowego, jej wdrożenia i w końcu monitorowania oraz systematycznej ewaluacji. Podobnie jak proces archiwizacji konferencja powinna mieć charakter długoterminowy, gdyż nawet najefektywniejsza w założeniach strategia wymaga bieżącej rewizji i ewentualnej modyfikacji, tak aby zachowała swoją skuteczność bez względu na zmieniające się okoliczności. Uzasadnione okazuje się publikowanie materiałów z takich spotkań, dokumentujących na bieżąco rozwój sytuacji i informujących opinię publiczną.

Potrzebne byłoby utworzenie polskojęzycznego portalu na temat długoterminowej archiwizacji zasobów cyfrowych. Stanowiłby on platformę informującą o zagranicznych oraz rodzimych inicjatywach, forum wymiany poglądów i idei dla znawców przedmiotu zaangażowanych w ewentualne przedsięwzięcia oraz tych, którzy pozostają bierni w działaniach i chcą tylko śledzić rozwój sytuacji. Korzystanie z tego typu portalu przedmiotowego – oczywiście przy założeniu jego aktualności i poprawności funkcjonowania – byłoby narzędziem bieżącego uzupełniania wiedzy dla zainteresowanych środowisk. W Polsce z powodzeniem funkcjonuje już tego typu tematyczna platforma EBIB. Można by zastanowić się nad jej uzupełnieniem o moduł tematyczny poświęcony długoterminowej archiwizacji polskiego dziedzictwa cyfrowego albo skorzystać z dobrych wzorców[5] i powołać do tych celów odrębny serwis WWW.

Warto zastanowić się nad zróżnicowaniem oferowanych form kształcenia i dokształcania pod względem stopnia specjalizacji i zaawansowania w tematyce – od kursów ogólnokształcących (budzących świadomość problemu

[5] Przykładowe prężnie działające portale prezentujące aktualną wiedzę na temat długoterminowej archiwizacji zasobów cyfrowych to:
Nestor – Kompetenznetzwerk Langzeitarchivierung: http://www.langzeitarchivierung.de/;
Padi – Preserving Access to Digital Information: http://www.nla.gov.au/padi/index.html;
OCLC – The worlds librarie's connected: http://www.oclc.org/services/collection/default.htm [Dostęp: 10.07.2017].

i dostarczających ogólnego zarysu wiedzy) do specjalistycznych szkoleń, przygotowujących do realizacji konkretnych, najbardziej skomplikowanych i specyficznych zadań w procesach długoterminowej archiwizacji. Z doświadczeń niemieckich, australijskich i holenderskich wynika, że do prowadzenia specjalistycznych szkoleń możliwe jest zaproszenie z innych krajów szkoleniowców posiadających wiedzę teoretyczną i doświadczenie praktyczne. W Niemczech np. plany i treści szkoleń z zakresu długoterminowej archiwizacji ustala się każdorazowo indywidualnie, w zależności od konkretnych potrzeb beneficjentów zgłaszających zapotrzebowanie na szkolenie. Odbiorców szkoleń, ludzi w różny sposób powiązanych zawodowo z zadaniami długoterminowej archiwizacji publikacji elektronicznych, przyporządkowuje się do trzech kategorii. Są to: (1) decydenci, (2) pracownicy instytucji kultury i nauki z ogólną znajomością problematyki, (3) pracownicy instytucji pamięci zatrudnieni bezpośrednio przy realizacji zadań związanych z długoterminową archiwizacją. Przynależność do określonej kategorii wiąże się z posiadaniem wiedzy, którą przedstawiono w tabeli 2.

Zaproponowany w tabeli podział problemów składających się na kompetencje w zakresie długoterminowej archiwizacji dokumentów cyfrowych wynika z realnego podejścia do problemu. Nie ma możliwości ani potrzeby, aby wszyscy ludzie związani z funkcjonowaniem instytucji kultury i nauki oraz zarządzaniem ich zasobami byli biegli we wszystkich aspektach tematu długoterminowej archiwizacji. W zależności od obszarów ich działalności oczekuje się od nich określonego zakresu i poziomu wiedzy.

Reasumując – polska teoria i praktyka archiwistyki cyfrowej potrzebuje dyskusji oraz opracowań dostarczających wiedzę na temat długoterminowej archiwizacji zasobów cyfrowych. Ponadto konieczne jest wywołanie przeświadczenia – nie tylko w pracownikach instytucji pamięci, lecz w każdym obywatelu – że nie da się funkcjonować w „digitalnym świecie" bez osobistego komputerowego zaplecza, wiedzy i permanentnego samokształcenia [Nowak, 2007, s. 186]. Wywoływanie takiego przeświadczenia powinno iść w parze z tworzeniem możliwości zdobywania i pogłębiania zarówno teoretycznych, jak i praktycznych umiejętności. Nade wszystko pracownikom polskich instytucji pamięci potrzebne jest przekonanie, że długoterminowa archiwizacja cyfrowych zasobów polskiej nauki i kultury to zadanie, któremu oni mają przewodniczyć w sensie merytorycznym i organizacyjnym, ale przy aprobacie i wparciu technicznym, prawnym oraz finansowym współodpowiedzialnych instytucji. Celem nadrzędnym tych instytucji jest współdziałanie na rzecz zachowania użyteczności cyfrowego dziedzictwa kultury i nauki, aż w Polsce, tak jak w innych krajach, archiwizacją zajmą się certyfikowane archiwa cyfrowe, budowane dla celów długoterminowej archiwizacji narodowego dziedzictwa cyfrowego.

Tabela 2. Zalecenia dotyczące stopnia znajomości poszczególnych zagadnień z zakresu długoterminowej archiwizacji zasobów cyfrowych przez osoby zawodowo związane z tym tematem

Treści z zakresu archiwizacji potrzebne pracownikom wymienionych kategorii	(1)	(2)	(3)
– ogólne rozeznanie w problematyce, – świadomość potrzeby podejmowania działań,	x	x	x
– pogłębiona teoretyczna znajomość rozwiązań w archiwizacji (strategie, infrastruktura, polityka gromadzenia i przechowywania),	x	x	x
– znajomość koncepcji i sposobów realizacji konkretnych zadań w ramach strategii zabezpieczania danych cyfrowych, ich odzysku oraz długoterminowej ochrony,		x	x
– pogłębiona znajomość sposobów realizacji strategii zabezpieczania, odzysku oraz długoterminowej ochrony danych cyfrowych, – znajomość funkcjonowania serwera archiwalnego i przyjętych w nim rozwiązań dotyczących realizacji zadań długoterminowej archiwizacji,	x		x
– pogłębiona znajomość i umiejętność zastosowania standardów obowiązujących w procesach długoterminowej archiwizacji,			x
– umiejętności teoretyczne, praktyczne oraz gotowość do działań w zakresie zarządzania danymi cyfrowymi,		x	x
– pogłębiona znajomość rozwiązań informatycznych i ich zastosowania w procesach długoterminowej archiwizacji,			x
– ogólna znajomość zagadnień prawnych związanych z długoterminową archiwizacją,	x	x	x
– pogłębiona znajomość zagadnień prawnych oraz ich stosowanie w procesach długoterminowej archiwizacji,			x
– znajomość aspektów ekonomicznych długoterminowej archiwizacji.	x	x	x

Źródło: oprac. własne za: [Neuroth i in., 2009]

3.6. Zagadnienia oceny i selekcji w procesie długoterminowej archiwizacji zasobów cyfrowych

Z przeglądu programów działań archiwizacyjnych realizowanych w instytucjach Niemiec, Australii, Wielkiej Brytanii [National Library of Australia, 2003, s. 84; Neuroth i in., 2009; Ravenwood, 2015; Twigge, 2003, s. 132-139] wynika, że jedną z istotnych czynności natury organizacyjnej w procesie długoterminowej archiwizacji zasobów cyfrowych jest opracowanie i zatwierdzenie dokumentu określającego kryteria oceny i selekcji materiałów cyfrowych podlegających długoterminowej ochronie. Potrzebne jest wskazanie zarówno technicznych, jak i merytorycznych parametrów dokumentów kwalifikujących je do kolekcji cyfrowego dziedzictwa nauki i kultury objętej długoterminową ochroną. Tym samym możliwa powinna stać się sprawna eliminacja z procesów długoterminowej archiwizacji materiału nieprzedstawiającego cech ponadczasowej wartości

dla nauki i kultury i z technicznego punktu widzenia niespełniających kryteriów gwarantujących jego zachowanie. Niezbędna jest również bieżąca ewaluacja archiwizowanych już dokumentów odnośnie do zasadności ich przechowywania (pozostawania) w archiwum [Borghoff i in., 2003, s. 99].

Zagadnieniom ochrony zasobów nauki i kultury, w szczególności ustaleniu, które materiały i dlaczego należy poddać szczególnej ochronie poświęca się w polskim piśmiennictwie przedmiotu uwagę okazjonalnie. Pomimo że publikowane wypowiedzi odnoszą się głównie do doboru i technologii ochrony zbiorów tradycyjnych, i nie podejmuje się w nich zagadnień archiwizacji zasobów cyfrowych, to jednak dostarczają one istotnej wiedzy na temat kryteriów tworzenia kolekcji archiwalnych i mogą, a nawet powinny stanowić podstawę dla procesów tworzenia archiwalnych kolekcji cyfrowych [Biliński, 2004; Drewniewska-Idziak, 2002; Sałaciński, 2001; Stachowska-Musiał, 2006].

Pojęcie selekcji należy rozumieć jako dobór poprzez eliminację, a także jako wybór [Dubisz, 2003, s. 1170]. Przy tak ogromnej podaży komunikatów instytucje pamięci nie są w stanie przyjąć jej w całości i dlatego decydują się na wybór materiałów oraz treści według własnych kryteriów [Biliński, 2004, s. 5; Wojciechowski, 2006, s. 13]. Procesy selekcji są przeprowadzane starannie przez wykwalifikowanych pracowników. Mają charakter procesu merytorycznego [Biliński, 2004, s. 5-7].

Proces oceny natomiast – ściśle powiązany z procesami selekcji i wyboru – jest rozumiany jako wydawanie opinii, sądzenie o wartości kogoś lub czegoś [Dubisz, 2003, s. 1101]. Na gruncie praktyki bibliotecznej jest to rozstrzyganie o wartości bibliotecznych zasobów, waloryzowanie i – w ślad za tym – decydowanie, co usunąć, a co włączyć do obiegu bibliotecznego oraz co na dłużej zarchiwizować. Odnosi się to zarówno do zasobów tradycyjnych, jak i elektronicznych. Z uwagi na sukcesywną podaż nowych komunikatów procesy wartościowania mają charakter ciągły. Częściowym zmianom ulegają też kryteria wartościowania. Ostatecznie procesy oceny i selekcji materiałów bibliotecznych służą rozstrzygnięciu, które z materiałów utworzą biblioteczną kolekcję, czyli zbiór zasobów dobranych i skonfigurowanych w sposób niepowtarzalny, bo bazujący na indywidualnej polityce wyboru i gromadzenia konkretnej biblioteki [Wojciechowski, 2006, s. 13-14], a które nie zakwalifikują się do dalszego wykorzystania ze względu na wysoki stopień ich zniszczenia bądź z uwagi na dezaktualizację treści albo z racji braku zainteresowania ze strony użytkowników [Biliński, 2004, s. 7]. Celem selekcji jest również eliminacja pozycji występujących w nadmiernej liczbie egzemplarzy w stosunku do aktualnych i przewidywanych potrzeb [Biliński, 2004, s. 10].

Zasadniczo selekcja jest procesem koniecznym, jednak są grupy materiałów bibliotecznych, które jej nie podlegają. Do takich zaliczane są [Biliński, 2004, s. 10-11]:

- materiały kwalifikujące się do narodowego zasobu bibliotecznego, jednak z wyjątkiem tych, które nie mieszczą się w profilu zbiorów biblioteki i przeznacza się je do wymiany bądź nieodpłatnego przekazania bibliotece specjalizującej się w gromadzeniu materiałów bibliotecznych z określonej dziedziny;
- depozyty czasowe i wieczyste, jeżeli deponent zastrzegł sobie prawo zachowania w całości tego depozytu, a biblioteka przyjmująca depozyt wyraziła na to zgodę;
- dary, ale tylko w przypadku kiedy biblioteka przyjmująca dary podjęła takie zobowiązanie. Na ogół jednak biblioteka stawia darczyńcy warunek, że włączenie daru do zbiorów może nastąpić po jego selekcji;
- materiały mniej wartościowe, jeśli stanowią część ważnej kolekcji w zbiorach biblioteki;
- wydawnictwa (w tym dokumenty życia społecznego) treściowo związane z regionem i środowiskiem, w którym działa biblioteka;
- pozycje zawierające cenne dedykacje bądź zapiski na marginesach, znaki własnościowe o znaczeniu historycznym.

Dotychczasowa wiedza i doświadczenie w zakresie oceny oraz selekcji zasobów bibliotecznych stanowią dużą pomoc w realizacji pierwszego etapu procesu tworzenia cyfrowych kolekcji archiwalnych. Ustalenia jednak wymaga, czy istniejące zasoby biblioteczne powinny być raz jeszcze poddane weryfikacji, czy też w całości będą podlegać długoterminowej ochronie. Choć względy ekonomiczne i racjonalne wskazują, iż niemożliwe jest zachowanie kompletnej cyfrowej kolekcji naukowego piśmiennictwa i dlatego celowo powoływane zespoły specjalistów dziedzinowych określają kryteria selekcji [National Library of Australia, 2003, s. 81-88; Twigge, 2003, s. 132-139], to jednak pytanie o podejście do wyboru materiałów do ochrony jest wciąż otwarte. Trwa dyskusja na temat selektywnego i kompleksowego wyboru dokumentów. Specjaliści popierający podejście kompleksowe twierdzą, że długoterminową wartość mogą mieć wszystkie zasoby, a koszty szczegółowej selekcji mogą przekroczyć wartość utrzymania kompletnej kolekcji. Natomiast rzecznicy wyboru selektywnego argumentują, że selekcja umożliwia przechowanie zbiorów o szczególnej wartości, stwarza gwarancje ich jakości technicznej oraz ułatwia negocjowanie praw dostępu z twórcami [National Library of Australia, 2003, s. 84]. Selekcja jednak jest procesem bardzo skomplikowanym i odpowiedzialnym, wymagającym przemyślenia. Aktywa informacyjne przechodzą przez różne fazy wartości i często odzyskują swoją wartość w przyszłości [Ross, 2003, s. 22-23]. Wyeliminowanie z kolekcji materiału cyfrowego, dlatego że np. osłabło zainteresowanie nim, może okazać się zgubne. Podstawowy problem polega na uzasadnieniu potrzeby długoterminowej ochrony materiału. Tam, gdzie istnieje ekonomiczna korzyść wynikająca z wtórnego skorzystania z informacji lub wymóg prawny jej zachowania, łatwo jest uzasadnić potrzebę archiwizacji. Do zachowania

należy przeznaczyć te cyfrowe zasoby, które mają ewidentną wartość w przypadku spraw sądowych, wartość akademicką, komercyjną lub są cenione za wkład w pamięć zbiorową bądź pamięć narodu [Ross, 2003, s. 16-20]. Kluczowym kryterium selekcji powinna zatem być zawartość treściowa dokumentu, jej znaczenie i przydatność w rozwoju nauki i kultury, ale – z punktu widzenia procesu archiwizacji, a w szczególności jej technicznych aspektów – niemniej istotną rolę odgrywają parametry techniczne. Duże znaczenie ma przegląd i ocena dokumentacji źródłowej, metadanych, a także dostępność zaplecza sprzętowego i programowego potrzebnego do uzyskania dostępu do danych oraz skorzystania z nich. W porównaniu z tradycyjnymi dokumentami papierowymi czas, w którym można podjąć efektywne działania w celu konserwacji dokumentów cyfrowych, jest znacznie krótszy [Twigge, 2003, s. 132-133]. Dzieje się tak z oczywistych – wymienianych już w niniejszej książce – powodów, tj.: z racji tempa zmian technologicznych zachodzących w systemach tworzenia, przechowywania i udostępniania dokumentów, a także nietrwałości nośników, na których przechowywane są treści. Głównym zagrożeniem dla zasobów cyfrowych jest dezaktualizacja technologiczna. Dla uniknięcia ryzyka utraty kompletności i niezawodności dostępu do dokumentów cyfrowych sugeruje się ich okresową ocenę. Należy dokonać jej najpóźniej w ciągu pięciu lat od daty utworzenia najwcześniejszego dokumentu w systemie [Twigge, 2003, s. 132-133]. Optymalnie byłoby, gdyby ocena dokumentów rozpoczynała się równolegle z projektowaniem i wdrażaniem całego systemu. Wcześnie dokonana ocena pozwala na identyfikację i zabezpieczenie tych dokumentów, które powinny być przeznaczone do długiego przechowywania, i jednocześnie umożliwia identyfikację dokumentów o wartości nietrwałej, które nie muszą być długoterminowo archiwizowane (co wiąże się z ograniczaniem kosztów archiwizacji).

W badaniu z 2007 r. pytano o selekcję materiałów rozumianą jako wybór tych, które z racji wartości swej treści zasługują na długoterminowe zachowanie [Januszko-Szakiel, 2011a, s. 21-46]. Respondenci deklarowali tworzenie kolekcji takich dokumentów oraz sporządzanie ich kopii zapasowej. Nie pytano jednak o procedury ich tworzenia, tym samym nie poznano zasad i kryteriów oceny i wyboru materiałów do długoterminowej archiwizacji. Pominięto także kwestię „przeglądu technicznego" zasobów tworzących kolekcję. Z pewnością jednak w każdej instytucji pamięci konieczny jest zabieg okresowej kontroli odczytu i prezentacji treści dokumentów cyfrowych, szczególnie pilny w przypadku dokumentów najstarszych, zapisanych w formatach i na nośnikach wychodzących z użytku lub już niestosowanych.

Ustalenie procedur oceny i wyboru materiałów cyfrowych stanowiących przedmiot długoterminowej ochrony wiąże się z szeregiem trudności do pokonania, wśród których wymienia się [National Library of Australia, 2003, s. 82]:

- mnogość i heterogeniczność materiału cyfrowego;

- różną jakość materiałów wynikającą z powszechnej dostępności różnorodnych środków ich wytwarzania i rozpowszechniania;
- presję związaną z zachowaniem wszelkich komunikatów, także internetowych, bez względu na ich jakość;
- konieczność dokonywania szybkich wyborów z racji wczesnej utraty użyteczności przez niektóre publikacje; szybka „samodyskwalifikacja" materiału cyfrowego z racji merytorycznych bądź technicznych parametrów może spowodować, iż zabraknie czasu na przekonanie się o (potwierdzenie) jego trwałej wartości i powzięcie decyzji o zachowaniu;
- klasyfikowanie (strukturyzowanie [Twigge, 2003, s. 132-133]) dokumentów cyfrowych składających się z powiązanych części lub występujących w wielu różnych wersjach;
- ustalenie pochodzenia i własności praw do niektórych materiałów cyfrowych w celu negocjowania warunków realizacji programu ochrony.

Oprócz wymienionych pojawiają się inne jeszcze problemy, z którymi instytucje pamięci muszą sobie radzić w procesie selekcji. Najważniejsze jest jednak dążenie do wypracowania jednoznacznych kryteriów doboru. Ponadto proces selekcji powinien być starannie przemyślany i konsekwentny. Procedury wyboru materiałów cyfrowych powinny odzwierciedlać potrzeby zainteresowanych środowisk oraz cele instytucji podejmujących się ich ochrony [National Library of Australia, 2003, s. 82].

Takie samo podejście reprezentują niemieckie instytucje archiwizujące. W ustalaniu kryteriów doboru dokumentów do kolekcji archiwalnych kierują się w pierwszej kolejności celami i ustawowymi zadaniami instytucji archiwizującej oraz zawartością treściową dokumentu, pertynentną wobec zapytań użytkowników. Niemcy wyróżniają kryterium merytoryczne oraz formalno-techniczne [Hänger i in., 2008]. W kryterium merytorycznym sugerują się głównie zakresem kompetencji archiwum, czyli w przypadku każdego obiektu rozstrzygają, czy archiwum jest zobowiązane – z racji nałożonych na nie zadań – do przyjęcia i archiwizowania obiektu oraz czy dokument z racji swej treści, formy estetycznej, formy wyrazu bądź swego znaczącego pochodzenia i historii warty jest długoterminowej ochrony w archiwum. Natomiast jako kryterium formalno-techniczne rozpatrywana jest kwestia rodzaju (typu) dokumentu, a wiodące pytanie dotyczy tego, czy archiwum jest w posiadaniu wiedzy i narzędzi do opracowania i zarządzania określonym typem dokumentu, a także czy ma możliwość dokument odczytać i zaprezentować w formie zrozumiałej dla użytkownika.

Jeszcze inne kryterium, prezentowane w niemieckiej literaturze przedmiotu [Hänger i in., 2008], dotyczy wyboru publikacji sieciowych. Oprócz tego, że w ich przypadku zastosowanie mają wymienione kryteria merytoryczne i formalno-techniczne, dodatkowo uwzględnia się inne parametry. Przede wszystkim dzieli

się wszystkie publikacje sieciowe na dwie kategorie: te, które posiadają odpowiednik w wersji drukowanej, i te, które go nie posiadają, czyli tzw. dokumenty typowo sieciowe.

Wśród dokumentów sieciowych posiadających odpowiednik drukowany wyróżnia się publikacje, które są dokładnym odzwierciedleniem publikacji drukowanej, czyli możliwie wierną cyfrową prezentacją „look and feel" wersji drukowanej, z uwzględnieniem układu i formatu strony tytułowej, stylu i wielkości czcionki itd., oraz publikacje sieciowe, które w swej formie nie są podobne do wersji drukowanej, jednak z racji typu charakterystycznego dla publikacji drukowanych mogą być do nich przyporządkowane, np. leksykon w formacie HTML. Przy doborze tychże publikacji do kolekcji archiwalnych należy rozstrzygnąć, czy długoterminowo w archiwum będzie przechowywana tylko wersja sieciowa, czy tylko drukowana, czy też obie.

Inne kryteria wymagają wypracowania w przypadku publikacji typowo sieciowych. Choć archiwizacja zasobów sieciowych jest tematem podejmowanym bardzo często, brak dotychczas wiążących zaleceń. Z racji połączenia publikacji sieciowych z innymi sieciowymi zasobami odpowiedzi wymagają pytania dotyczące granic publikacji sieciowych jako obiektów archiwalnych. Nie wiadomo, czy obiektem archiwalnym powinien być np. sam blog internetowy, czy także wszelkie strony, do których blog odsyła [Hänger i in., 2008]. Być może dobrym rozwiązaniem byłoby potraktowanie przykładowego blogu jako obiektu archiwalnego, natomiast informacje o obiektach, do których blog odsyła, można by zamieścić w metadanych do obiektu. W ten sposób możliwe byłoby zalecane uchwycenie kontekstu obiektu archiwalnego. Wydaje się, że kluczowym elementem takich metadanych powinny być identyfikatory trwałe obiektów, do których odsyła blog.

Przy doborze dokumentów sieciowych do kolekcji archiwalnych można kierować się założeniami dotyczącymi tworzonej kolekcji. Mogą to być dokumenty określonej domeny lub określonego zakresu tematycznego, także publikowane przez określony typ twórców [Hänger i in., 2008; PADI, b.d.]. Problematyczny natomiast jest sposób radzenia sobie z częstością ich zmian. Procesy uzupełniania, aktualizacji mają w zasadzie charakter ciągły. Ustalono już, że niezbędne jest ich gromadzenie w określonych odstępach czasu. Natomiast nie wiadomo, jaki interwał czasu przyjąć [Hänger i in., 2008].

Zaleca się publikowanie procedur, a w szczególności informowanie zainteresowanych środowisk o tym, co zostało zakwalifikowane, a co pominięte w tworzeniu archiwalnych kolekcji. W polityce selekcji najważniejsze są: odpowiedzialność, obiektywizm i transparentność [National Library of Australia, 2003, s. 82]. Publikacja zasad oceny i selekcji jest traktowana jako poważny czynnik budujący wiarygodność instytucji archiwizujących [Hänger i in., 2008].

Instytucje pamięci powinny liczyć się z tym, że niełatwe jest zaproponowanie ogólnych kryteriów tworzenia cyfrowych kolekcji archiwalnych [National Library of Australia, 2003, s. 83]. W tej kwestii powinny raczej stawiać na strategie indywidualne. W zależności od typu instytucji i realizowanych w niej celów zmieniać się będą opinie o wartości materiału i potrzebie jego długoterminowego zachowania. W przypadku archiwów instytucji naukowych podstawowym kryterium doboru wydaje się być ponadczasowa przydatność dokumentów w procesach edukacji i rozwoju nauki, w szczególności zaś wartość informacji naukowych w nich zawartych. W zależności od specyfiki instytucji oraz zwyczajów kraju, w którym funkcjonuje, dopuszcza się przyjmowanie różnych zasad wyboru materiałów, choć kryterium najważniejszym winna być ich istota i trwała wartość kulturowa, naukowa czy dokumentalna. Decyzje – i ich ewentualne przyszłe korekty – powinny wynikać ze zdefiniowanych zasad, polityki, procedur i standardów [National Library of Australia, 2003, s. 25].

Dodatkowo w procesach selekcji zaleca się – niełatwe – zestawienie wartości merytorycznej materiału cyfrowego z prawdopodobnymi kosztami i trudnościami jego długotrwałej ochrony [National Library of Australia, 2003, s. 83]. Przy założeniu, że kwalifikowanie obiektów do kolekcji archiwalnej jest wynikiem starannej, przemyślanej konfrontacji znaczenia dokumentów dla nauki i kultury wraz z finansowymi i technicznymi możliwościami systemu archiwalnego do ich długoterminowej ochrony, wzrasta efektywność procesu archiwizacji i tym samym wiarygodność archiwum cyfrowego. Uwzględnienia wymaga też kwestia praw własności zasobów cyfrowych. Długoterminowa, niekiedy skomplikowana i kosztowna ochrona materiału, którego udostępnienie w przyszłości – z racji obwarowań prawnych – nie będzie możliwe, byłaby raczej nieracjonalna [National Library of Australia, 2003, s. 83].

Pod uwagę trzeba też wziąć fakt, że koszty i możliwości programów ochrony zasobów cyfrowych będą z czasem ulegać zmianom. Z racji upowszechniania rozwiązań koszty prawdopodobnie obniżą się, a wydajność narzędzi archiwizacyjnych będzie wzrastać wraz z wciąż postępującym rozwojem technologicznym. W związku z tym zalecana jest ostrożność przy eliminacji materiału, którego utrzymanie obecnie wydaje się trudne i kosztowne [National Library of Australia, 2003, s. 83]. W przypadku publikacji zasługujących na ochronę, ale jednocześnie przerastających teraźniejsze możliwości systemu archiwalnego, warte przemyślenia jest opracowanie strategii tymczasowej. Prawdopodobnie krótkoterminowa ochrona dokumentów pozwoliłaby im przetrwać do czasu wdrożenia ich do profesjonalnego archiwum cyfrowego. Na tym właśnie polega zalecana ostrożność w procesach selekcji. Słuszne podejście polega na zdecydowaniu: które materiały powinny zostać zachowane na pewno (i na jak długo), które nie wymagają ochrony, a które

należy zachować tymczasowo (w oczekiwaniu na podjęcie ostatecznej decyzji). Rezygnacja z zachowania materiału cyfrowego powinna mieć charakter ostateczny.

W przypadku dokumentów pochodzących z procesów digitalizacji selekcja nie stanowi aż tak poważnego problemu; ich ucyfrowienie świadczy poniekąd o ich szczególnej wartości kulturowej, potrzebie w nauce, edukacji, sztuce. Ustalenie listy tytułów poddawanych digitalizacji też jest zadaniem trudnym, zwykle wykonywanym przez zespół osób współpracujących ze specjalistami dziedzinowymi [Daszewski, 2004]. Istotną rolę w wyborze obiektów do kolekcji archiwalnej mogą odegrać ich twórcy. Dysponują oni istotnymi danymi o formie i treści dokumentu, o przeznaczeniu, wartości, relacjach z innymi obiektami itp. W przypadku wielu obiektów cyfrowych jeśli dane te nie zostaną przejęte od twórcy, późniejsze ich zrekonstruowanie bywa trudne lub niemożliwe [National Library of Australia, 2003, s. 84].

Jedną z cech procesu selekcji zasobów cyfrowych powinna być jego cykliczność. Decyzje dotyczące wyboru materiałów cyfrowych wchodzących do archiwalnych kolekcji raczej nie powinny mieć charakteru ostatecznego. W przypadku wątpliwości wskazany jest okresowy przegląd w celu sprawdzenia, czy wartość materiału wciąż uzasadnia koszty jego ochrony. W piśmiennictwie zaleca się unikanie tzw. weryfikacji negatywnej polegającej na eliminacji z kolekcji materiałów, które jednak zasługują na ochronę (a w dodatku bezterminową). Uwzględnienia wymaga także fakt, że proces selekcji – z ekonomicznych względów – powinien być powtarzany jak najrzadziej. Decyzje o wyborze publikacji do kolekcji archiwalnej powinny być uzasadnione na piśmie wraz z opisem atrybutów materiału, z racji których podlega on długoterminowej ochronie [National Library of Australia, 2003, s. 83-84].

Powyższe rozważania wskazują na obszar problemów, które należy sobie uzmysłowić i obszar zadań, które należy podjąć w związku z doborem materiałów cyfrowych podlegających długoterminowej ochronie. Szczególnego przemyślenia wymaga sposób podejścia do tworzenia kolekcji archiwalnej. Możliwe jest zastosowanie metody selektywnej, której słabością jest to, że zwykle bazuje na subiektywnej ocenie wartości naukowej i kulturowej materiałów, więc wiele z tych, które mogą okazać się cenne w przyszłości, może zostać pominiętych, a długoterminowej ochronie mogą zostać poddane zasoby, których wartość okaże się ulotna. Natomiast kompleksowa metoda archiwizacji zasobów cyfrowych, choć faktycznie kosztowna, zapewniłaby dostępność wszelkich materiałów, tym samym oszczędziłaby czasochłonnej i odpowiedzialnej pracy. Istotne wydaje się, aby w przypadku podejścia selektywnego instytucje pamięci poddały ocenie możliwie wszystkie zgromadzone zasoby cyfrowe. Przeglądem merytorycznym oraz technicznym powinny zostać objęte dokumenty zapisane na nośnikach przenośnych, dokumenty

umieszczone na serwerach instytucji, a także zasoby internetowe. Najpoważniejsze wyzwanie związane jest prawdopodobnie z zasobami sieciowymi. Procesy tworzenia i archiwizacji kolekcji zasobów sieciowych to wielki obszar zadań. Jednak – jak wynika z obserwacji poczynań wielu światowych instytucji pamięci – można sobie z nim poradzić [PADI, b.d.; Phillips, 2005, s. 57-71; Ślaska, 2009].

3.7. Polskie wybrane inicjatywy na rzecz trwałej ochrony polskiego dziedzictwa cyfrowego

W australijskich zaleceniach dotyczących ochrony dziedzictwa cyfrowego słusznie zwraca się uwagę, iż „dziedzictwo cyfrowe nie podlega ograniczeniom czasowym, geograficznym ani kulturowym. Jest ono związane z kulturą, w której powstało, lecz pozostaje dostępne wszystkim mieszkańcom globu. Dziedzictwo cyfrowe wszystkich regionów, krajów i społeczności należy zachowywać i udostępniać, by umożliwić powstanie w nim zrównoważonej i sprawiedliwej reprezentacji wszystkich ludów, narodów, kultur i języków". Zatem każdy kraj powinien zatroszczyć się o zachowanie własnego dziedzictwa nauki i kultury oraz o udostępnienie go zainteresowanym użytkownikom innych krajów. Samodzielne opracowanie strategii archiwizacji oraz zaprojektowanie infrastruktury archiwum cyfrowego może być zbyt kosztowne i właśnie to powinno stanowić silny bodziec do poszukiwania możliwości wspólnego korzystania z doświadczeń i zasobów organizacyjnych, technicznych, finansowych, a także pomysłów na rozwiązania prawne. Wymieniane korzyści współpracy to: dostęp do szerszych zasobów wiedzy i doświadczeń, wspólne pokrycie kosztów rozwoju, szersze spektrum chronionych materiałów, uniknięcie powielanych działań, większa siła przebicia w negocjacjach z twórcami materiałów cyfrowych oraz nakłonienia wpływowych ludzi do poważnego traktowania problemu. Ponadto połączenie sił powoduje większy wpływ na badania i większe możliwości uzyskania pomocy finansowej. Współpraca umożliwi sformułowanie wspólnych standardów działań, zapewniając kompatybilność rozwiązań, wypracowanie jednolitego stanowiska i jego prezentację w kręgach oraz w inicjatywach popularyzujących tematykę ochrony dziedzictwa cyfrowego [National Library of Australia, 2003, s. 26, 72].

W strategiach narodowych celem nadrzędnym jest objęcie ochroną cyfrowych zasobów opublikowanych w określonym kraju i w ojczystym języku określonego kraju bądź poza jego granicami i w językach obcych, ale treściowo jego dotyczących. Bez względu na to, czy dąży się do utworzenia jednego centralnego archiwum dla narodowego zasobu cyfrowego, czy też przyjmuje się model archiwum jednorodnego rozproszonego, propaguje się przyjęcie i stosowanie

jednakowych, istniejących już standardowych praktyk organizacji archiwów oraz współodpowiedzialność [Dobratz i Tappenberg, 2002, s. 257-261].

W 2007 r. przedstawicieli polskich instytucji pamięci pytano, czy w naszym kraju powinno powstać jedno centralne archiwum dla polskiego zasobu cyfrowego (jedna instytucja archiwizująca kompletny zasób cyfrowy Polski), czy raczej należałoby zakładać lokalne archiwa cyfrowe polskich zasobów cyfrowych zorganizowane i działające w instytucjach pamięci na podstawie ujednoliconych, ogólnie przyjętych zasad (tzw. archiwum rozproszone jednorodne), czy też powinny powstawać lokalne archiwa polskich zasobów cyfrowych zorganizowane i działające w instytucjach pamięci na podstawie ich indywidualnych zasobów wiedzy i możliwości (tzw. archiwum rozproszone niejednorodne). Respondenci opowiedzieli się za modelem archiwum rozproszonego jednorodnego, co oznaczałoby, że przy polskich instytucjach pamięci powinny powstawać lokalne archiwa zasobów cyfrowych na ujednoliconych w skali kraju zasadach ich organizacji i funkcjonowania. Jedna odpowiedź sugerowała, że przy takim modelu archiwum warta przemyślenia jest opcja utworzenia tzw. narodowej kopii bezpieczeństwa cyfrowych zasobów i przechowywania jej w „krajowym magazynie danych". W niektórych odpowiedziach autorzy zakładali utworzenie centralnego narodowego archiwum cyfrowego. Uzupełnienie jednej z tych odpowiedzi stanowiła sugestia – bardzo słuszna – o lokalizacyjnym oddaleniu co najmniej dwóch kopii archiwalnej kolekcji. Autor odpowiedzi nawiązał prawdopodobnie do praktyki stosowanej w krajach zaawansowanych w działaniach archiwizacyjnych, polegającej na tworzeniu tzw. lustrzanych kopii kolekcji archiwalnych i przechowywaniu ich w oddalonych miejscach bądź oddawanie ich pod opiekę instytucjom partnerskim w zamian za ochronę kopii ich narodowej kolekcji cyfrowej. Na takie rozwiązanie zdecydowały się biblioteki narodowe Niemiec i Holandii. Na mocy umowy partnerskiej kraje wzajemnie przechowują kopie narodowych kolekcji cyfrowych.

Zarówno utworzenie centralnego archiwum narodowego, jak i lokalnych, rozproszonych archiwów jednorodnych wymaga współpracy na rzecz opracowania polityki postępowania, powszechnie akceptowanej przez zainteresowane środowiska, tzw. narodowej strategii długoterminowej ochrony zasobów cyfrowych. W tym celu potrzebna jest inicjatywa oraz koordynacja działań i przede wszystkim chęć współpracy. Omawiane badanie służyło też rozpoznaniu, kto – jaka instytucja czy organizacja – powinien przejąć rolę narodowego koordynatora działań oraz czy pracownicy polskich instytucji pamięci byliby skłonni przyłączyć się do współpracy na rzecz opracowania narodowej strategii archiwizacji.

W kilku odpowiedziach wskazano na usługę dedykowaną ochronie polskich zasobów o nazwie Krajowy Magazyn Danych (KMD), oferowaną przez Poznańskie Centrum Superkomputerowo-Sieciowe.

W materiale badawczym z lat 2007-2009 respondenci odnieśli się do rodzimych inicjatyw związanych z zarządzaniem i ochroną polskich zasobów cyfrowych. Zwrócono uwagę na opracowaną przez MKiDN *Strategię digitalizacji i budowania zasobów cyfrowych dla instytucji kultury i nauki na lata 2007-2013*. Wymienione zostały działania polskiej Biblioteki Narodowej w zakresie ochrony zasobów cyfrowych – głównie w odniesieniu do Cyfrowej Biblioteki Polona. Zwrócono też uwagę na projekt związany z utworzeniem Narodowego Repozytorium Dokumentów Elektronicznych NRDE BN (obecnie Repozytorium Cyfrowe BN).

Uczestnicy badania wspomnieli też o inicjatywie Sekcji Archiwów Naukowych Szkół Wyższych, w ramach której organizowano szkolenia przygotowujące specjalistów do zadań długoterminowej archiwizacji oraz o działaniach Archiwum Akt Nowych. Nie podano dodatkowych informacji, ale prawdopodobnie rzecz dotyczyła przedsięwzięcia związanego z utworzeniem Narodowego Archiwum Cyfrowego (NAC).

Z analizy odpowiedzi ankietowych wynika, że bez wyjątku wszyscy respondenci wyrazili zainteresowanie i chęć przyłączenia się do prac nad narodową strategią archiwizacji dziedzictwa cyfrowego Polski. Z kolei w odpowiedziach dotyczących funkcji koordynatora działań w tym zakresie wymieniano Bibliotekę Narodową, Poznańskie Centrum Superkomputerowo-Sieciowe, Ministerstwo Kultury i Dziedzictwa Narodowego, a także Naczelną Dyrekcję Archiwów Państwowych.

Zaangażowanie przedstawicieli świata polityki w opracowywanie strategii długoterminowej ochrony polskiego dziedzictwa cyfrowego wydaje się bardzo zasadne, przede wszystkim z racji potrzeby zrozumienia przez rządzących wagi problemu, pilnej potrzeby działań oraz ich finansowania. Bez finansowego wsparcia, tego typu przedsięwzięć nie można realizować i, co ważniejsze, nie chodzi o finansowanie jednorazowe bądź krótkoterminowe. Długoterminowa, wieczysta archiwizacja zasobów cyfrowych wymaga nakładów permanentnych. Z wypowiedzi respondentów badania wynika jednoznacznie, że jednym z wyzwań, z którymi polskie instytucje pamięci muszą uporać się w procesie długoterminowej archiwizacji, jest właśnie pozyskanie środków finansowych. Wśród pozostałych wyzwań respondenci jednogłośnie wymienili brak planu działania i zaniedbania organizacyjne. Jako wyzwanie postrzegano zmiany technologiczne oraz brak specjalistów, a tym samym umiejętności reagowania na zmiany technologiczne oraz brak profesjonalnego sprzętu i oprogramowania potrzebnych dla celów długoterminowej ochrony zbiorów cyfrowych. Wskazano też na wyzwanie natury prawnej – potrzebę nowelizacji przepisów dotyczących procesów archiwizacji cyfrowych zasobów nauki i kultury.

3.7.1. Aktywność Rządu RP w obszarze trwałej ochrony polskiego zasobu cyfrowego

Z doświadczeń krajów zaawansowanych w działania archiwizacyjne wynika, że w procesie tworzenia narodowych strategii archiwizacji powinny uczestniczyć przede wszystkim instytucje pamięci, ale ze wsparciem instytucji tworzących zasoby cyfrowe, specjalistów branży IT, ekonomistów, prawników i przedstawicieli rządu sprawujących nadzór nad sprawami kultury oraz ochrony dziedzictwa narodowego. Programy działań archiwizacyjnych chociażby Australii, Holandii bądź Niemiec wskazują, że zespoły narodowe powinny przyłączać się do inicjatyw międzynarodowych, gdyż tylko w ten sposób poszczególne kraje mają szansę opracować strategie zunifikowane, dające możliwości dostępu do zasobów dziedzictwa w skali globalnej, a także ułatwiające wzajemne przechowywanie kopii kolekcji narodowych. Zwłaszcza kraje planujące albo przystępujące do działań powinny odstąpić od tworzenia programów indywidualnych na rzecz nawiązania współpracy z programami, instytucjami i organizacjami, które dysponują już określonym doświadczeniem i mogą zaoferować pomoc [Dobratz i Tappenberg, 2002, s. 257-261; National Library of Australia, 2003, s. 34].

Polskie instytucje pamięci powinny ustalić, czy polski zasób cyfrowy będzie archiwizowany w lokalnych archiwach cyfrowych, czy też w jednym depozycie centralnym. W tej kwestii istniała w latach 2007-2009 i istnieje nadal potrzeba podjęcia współpracy i stworzenia powszechnie akceptowanego narodowego programu działań archiwizacyjnych. Bez względu na to, czy będzie to model archiwum rozproszonego, czy centralnego, potrzebne jest wyznaczenie instytucji bądź organizacji, która przyjmie rolę inicjatora i koordynatora działań archiwizacyjnych w Polsce.

Brakuje wiążących deklaracji o archiwizacji polskiego dziedzictwa cyfrowego. Zasadniczą zmianę i duży krok naprzód stanowi przywoływany już raport MKiDN z 2009 r., w którym stwierdzono, iż bezpieczna i długoterminowa archiwizacja polskich zasobów cyfrowych jest jednym z zadań najpilniejszych [Program digitalizacji, 2009]. Kolejnym ważnym wydarzeniem w rodzimych działaniach na rzecz trwałej ochrony zasobów cyfrowych było ustanowienie przez MKiDN merytorycznych centrów kompetencji [Duńczyk-Szulc, 2012]. Centra kompetencji zostały utworzone przy następujących instytucjach:

- Biblioteka Narodowa (2010 r.),
- Narodowy Instytut Dziedzictwa (2010 r.),
- Narodowy Instytut Audiowizualny (2010 r.),
- Narodowe Archiwum Cyfrowe (2010 r.),
- Narodowy Instytut Muzealnictwa i Ochrony Zbiorów (2013 r.).

Zakres kompetencji wymienionych instytucji to:

- koordynacja działań digitalizacyjnych, edukacja kadr instytucji kultury;
- zarządzanie zasobami cyfrowymi, udostępnianie zasobów cyfrowych;
- wypracowanie standardów digitalizacji różnego rodzaju obiektów;
- wdrożenie zmian technologicznych w zakresie digitalizacji i przechowywania danych cyfrowych.

Zgodnie z założeniem MKiDN poszczególne centra kompetencji przejmują odpowiedzialność za zarządzanie procesami digitalizacji, ochrony i udostępniania następujących typów materiałów:

- Biblioteka Narodowa – materiały biblioteczne;
- Narodowy Instytut Dziedzictwa – zabytki;
- Narodowy Instytut Audiowizualny – materiały audiowizualne i audialne;
- Narodowe Archiwum Cyfrowe – materiały archiwalne;
- Narodowy Instytut Muzealnictwa i Ochrony Zbiorów – muzealia.

Udostępniono kilka instrumentów finansowania zadań z zakresu digitalizacji i ochrony polskiego dziedzictwa kultury i nauki:

- 2007-2009 r. Program MKiDN Dziedzictwo kulturowe, priorytet Tworzenie zasobów cyfrowych dziedzictwa kulturowego;
- 2010 r. Program MKiDN Zasoby cyfrowe;
- 2010-2015 r. Wieloletni Program KULTURA+, priorytet Digitalizacja;
- 2011-2012 r. Program własny Narodowego Instytutu Audiowizualnego Dziedzictwo cyfrowe;
- 2013-2015 r. Program MKiDN Dziedzictwo kulturowe, priorytet Ochrona i cyfryzacja dziedzictwa kulturowego;
- 2016 r. do chwili obecnej Program MKiDN Kultura cyfrowa.

Jednym z celów szczegółowych Programu KULTURA+ było stworzenie sieci profesjonalnych repozytoriów cyfrowych w centrach kompetencji i innych dużych ośrodkach umożliwiających właściwe przechowywanie zdigitalizowanych zbiorów. Z analiz prowadzonych w latach 2016-2017 wynika, że wybrane instytucje pamięci zdołały stworzyć warunki informatyczno-techniczne dla składowania cyfrowych zasobów. Jednak postulowana jest potrzeba utworzenia centralnego depozytu dla dziedzictwa cyfrowego bądź sieci magazynów cyfrowych, ale zarządzanych i utrzymywanych centralnie. Odpowiedzią na te postulaty jest prawdopodobnie najnowsza inicjatywa rządowa – projekt KRONIK@ – Krajowe Repozytorium Obiektów Nauki i Kultury. Z inicjatywą opracowania i uruchomienia projektu wystąpiło Ministerstwo Cyfryzacji w 2017 r.

Projekt zakłada realizację dwóch kluczowych celów: stworzenie wspólnej, zintegrowanej infrastruktury, służącej do przechowywania cyfrowych zasobów nauki i kultury oraz udostępnienie wszystkich zgromadzonych obiektów poprzez jedną

platformę. Spodziewana korzyść to zmniejszenie kosztów budowy i utrzymania infrastruktury IT oraz poprawa jakości i dostępności cyfrowych zasobów kultury i nauki. Projekt ma integrować środowiska i jednostki, które są w posiadaniu zasobów nauki i kultury [Kronik@, 2017].

3.7.2. Aktywność Naczelnej Dyrekcji Archiwów Państwowych w obszarze trwałej ochrony polskiego zasobu cyfrowego

Na wniosek Naczelnego Dyrektora Archiwów Państwowych, Minister Kultury i Dziedzictwa Narodowego przekształcił Archiwum Dokumentacji Mechanicznej w Narodowe Archiwum Cyfrowe (NAC). Oficjalnie Narodowe Archiwum Cyfrowe zostało powołane 8 marca 2008 r. Celem NAC jest zapewnienie bezpiecznej i długookresowej archiwizacji materiałów cyfrowych stanowiących narodowy zasób archiwalny [Narodowe Archiwum Cyfrowe, b.d.; Program digitalizacji, 2009, s. 21-22]. Obecnie NAC administruje systemem o nazwie Archiwum Dokumentów Elektronicznych (ADE), który jest prototypem narzędzia służącego do zarządzania archiwalnymi dokumentami elektronicznymi wytworzonymi przez administrację publiczną. Prototyp został wykonany w latach 2005-2006 przez Naukową i Akademicką Sieć Komputerową na podstawie decyzji Ministra Edukacji i Nauki oraz na zlecenie Naczelnej Dyrekcji Archiwów Państwowych. Po serii testów przeprowadzonych w 2006 r. prototyp uznano za gotowy do przyjęcia dokumentów elektronicznych uporządkowanych zgodnie z przepisami prawa. Z informacji zamieszczonych na witrynie internetowej projektu ADE wynika, że nadal trwają prace nad wprowadzaniem poprawek i udogodnień, zasugerowanych w fazie testowania [NAC, b.d.]. W 2015 r. opublikowano informacje o tym, że Projekt Archiwum Dokumentów Elektronicznych, przygotowany przez Naczelną Dyrekcję Archiwów Państwowych, otrzymał rekomendację do dofinansowania w ramach unijnego Programu Operacyjnego Polska Cyfrowa dla działania 2.1 Wysoka dostępność i jakość e-usług publicznych [NDAP, 2015b]. NDAP przystępuje do realizacji projektu z założeniem, że „Misją archiwów państwowych jest trwałe zachowanie świadectw przeszłości i zapewnienie do nich powszechnego dostępu w celu wspierania rozwoju państwa i społeczeństwa obywatelskiego". „W podmiotach realizujących zadania publiczne zaczęła powstawać dokumentacja, która od początku do końca swojego cyklu życia posiada postać elektroniczną". „Osiągnięcie gotowości do zachowania w długim czasie dokumentów elektronicznych powstających współcześnie w administracji publicznej jest wyzwaniem nie tylko dla archiwów państwowych". „Wszystkie projekty, które prowadzą do powstania i rozwoju e-administracji powinny mieć zapewnione miejsce bezpiecznego przechowywania wytworzonych materiałów archiwalnych w postaci elektronicznej, które zapewni możliwość udostępniania ich w przyszłości" [NDAP, 2015a].

3.7.3. Aktywność Biblioteki Narodowej w obszarze trwałej ochrony polskiego zasobu cyfrowego

Biblioteka Narodowa jako Centrum Kompetencji w zakresie digitalizacji materiałów bibliotecznych wdraża zmiany technologiczne dotyczące digitalizacji i przechowywania zasobów cyfrowych oraz koordynuje działania w zakresie ich gromadzenia i przechowywania. Ponadto, jak wynika ze sprawozdania BN za 2016 r., książnica prowadzi szkolenia i konsultacje eksperckie z zakresu digitalizacji, udziela informacji dotyczących przygotowania kopii wzorcowych oraz metadanych przekazywanych do przechowywania wieczystego w Repozytorium Cyfrowym BN [Sprawozdanie, 2016, s. 80].

Od 2009 r. przy BN jest rozbudowywane repozytorium cyfrowe. Są w nim składowane dokumenty zarówno będące wynikiem digitalizacji, jak i *born digital*. Repozytorium jest wyposażone w Centrum Podstawowe, mieszczące się na terenie siedziby głównej BN, wyposażone i prowadzone zgodnie z najlepszymi standardami oraz Centrum Zapasowe będące własnością BN, zabezpieczające ciągłość działania podstawowych funkcji systemu na wypadek poważnych awarii lub katastrof w Centrum Podstawowym. „Repozytorium Cyfrowe BN jest archiwum wieczystym, które zapewnia ciągłość przechowywania informacji, dlatego zastosowane przy jego projektowaniu technologie odznaczają się elastycznością, otwartością i skalowalnością. Wykorzystane mechanizmy umożliwią migrację danych w ramach Repozytorium pomiędzy kolejnymi generacjami sprzętu lub oprogramowania. Repozytorium jest przygotowane do przechowywania zarówno zasobów cyfrowych wytworzonych lub pozyskanych przez Bibliotekę Narodową, jak i przekazanych przez inne instytucje" [Sprawozdanie, 2016, s. 82-83].

Biblioteka Narodowa przystąpiła również do realizacji dużych przedsięwzięć: Patrimonium – digitalizacja i udostępnienie polskiego dziedzictwa narodowego ze zbiorów Biblioteki Narodowej oraz Biblioteki Jagiellońskiej oraz E-USŁUGA OMNIS [Sprawozdanie, 2016, s. 82-85].

W 2016 r. decyzją Międzynarodowej Federacji Stowarzyszeń i Instytucji Bibliotekarskich (IFLA) powierzono Bibliotece Narodowej zadanie organizacji Ośrodka Ochrony i Konserwacji IFLA (Preservation and Conservation Centre). Ośrodek „zajmie się zagadnieniami przechowywania zasobów cyfrowych, w tym kwestią wyboru obiektów do digitalizacji, jej standardów oraz warunków długotrwałego przechowywania zasobów cyfrowych. Będzie wspierał działania bibliotek Europy Środkowej i Wschodniej w zakresie zabezpieczenia cyfrowego dziedzictwa kulturowego" [Sprawozdanie, 2016, s. 11].

3.7.4. Aktywność Poznańskiego Centrum Superkomputerowo-Sieciowego w obszarze trwałej ochrony polskiego zasobu cyfrowego

Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS), afiliowane przy Instytucie Chemii Bioorganicznej PAN, działa od 1993 r. i świadczy głównie usługi związane z Internetem dla środowiska naukowego. W 1996 r. PCSS rozpoczęło prace badawcze, które miały na celu stworzenie oprogramowania do budowy bibliotek cyfrowych. W rezultacie w 2002 r. powstał system dLibra jako pierwszy zastosowany do budowy Wielkopolskiej Biblioteki Cyfrowej i z czasem do kolejnych polskich bibliotek cyfrowych współtworzących obecnie Federację Bibliotek Cyfrowych.

Oprócz projektów dotyczących oprogramowania, m.in.: dLibra, dLab, dMuseion, dArceo, do budowy i ochrony polskich zasobów cyfrowych, PCSS angażuje się w europejskie projekty, np.: ENRICH, EuropeanaLocal, Europeana, Europeana Cloud, LoCloud, SCAPE, Centrum Kompetencji IMPACT w Zakresie Digitalizcji, SYNAT, Succeed, DCH-RP.

W wywiadzie przeprowadzonym w maju 2008 r. w Poznańskim Centrum Superkomputerowo-Sieciowym pracownicy PCSS wyrazili zainteresowanie udziałem w pracach nad ustaleniem strategii długoterminowej archiwizacji zasobów cyfrowych. W toku rozmowy ustalono, że tworząc system archiwalny dla polskich zasobów cyfrowych, należałoby założyć ich składowanie na serwerze archiwalnym, tj. profesjonalnym archiwizerze. Treści dokumentów zapisane na nośnikach fizycznych najsensowniej byłoby oddzielić od nośników i również umieścić na serwerze. Jeśli natomiast program ochrony zakładałby archiwizację treści na nośnikach oryginalnych, należałoby przynajmniej co pięć lat dokonać zmiany nośnika. Pracownicy PCSS podkreślali znaczenie stosowania otwartych standardów zapisu.

W rozmowie pytano również o wskazania dotyczące modelu archiwum dla polskiego dziedzictwa cyfrowego. Z odpowiedzi wynikało, że łatwiejsze i sprawniejsze organizacyjnie wydaje się utworzenie jednego centralnego archiwum. Należy jednak rozważyć, na ile wadą, a na ile zaletą jest złożenie funduszy i odpowiedzialności „w jedne ręce". Z związku z tym prawdopodobnie lepszy byłby model jednorodnego archiwum rozproszonego z rozproszoną odpowiedzialnością. Archiwum rozproszone wydaje się też lepiej spełniać wymagania bezpieczeństwa – np. w przypadku awarii, zagrożenie uszkodzenia bądź utraty zasobów dotyczy fragmentu kolekcji, a nie całości jak w przypadku archiwum centralnego.

W wywiadzie ustalono, że ochrona zasobów polskich bibliotek cyfrowych odbywała się wówczas w sposób powszechnie przyjęty w systemach informatycznych, tzn. wykonywane były sumy kontrolne oraz kopie bezpieczeństwa obiektów cyfrowych. Zasoby bibliotek cyfrowych nie były objęte programem archiwizacji

długoterminowej; nie było w architekturze i funkcjach systemu dLibra komponentu odpowiedzialnego za długoterminową użyteczność zasobów cyfrowych. Pomysł jego utworzenia bardzo spodobał się pracownikom PCSS i wyrazili opinię, że warto uzupełnić architekturę dLibry. Interesująca wydała się rozmówcom idea selekcjonowania zasobów i utworzenia w systemie modułu zajmującego się przechowywaniem treści ważnych. Prawdopodobnie na tej podstawie zostało opracowane oprogramowanie dArceo, które stanowi komponent niezależny od dLibry. Jest to system zapewniający długoterminowe przechowywanie danych cyfrowych, zgodny z modelem OAIS [Mazurek i in., 2013, s. 101-111].

W rozmowie pracownicy PCSS zwrócili uwagę na projekt o nazwie Krajowy Magazyn Danych, którego efektem jest rozproszony system trwałego i bezpiecznego przechowywania danych o zasięgu krajowym, z przeznaczeniem głównie dla instytucji akademickich i edukacyjnych [KMD, b.d.]. KMD jako projekt o charakterze badawczym przyczynił się do rozwoju wiedzy i dostarczył cennych doświadczeń w dziedzinie zarządzania zasobami cyfrowymi. Równolegle dostępna jest też Usługa Powszechnej Archiwizacji PLATON U4 [Brzeźniak i in., 2009].

Z informacji uzyskanych ze środowiska PCSS w 2016 r. wynika, że Usługa PLATON Powszechna Archiwizacja została zrealizowana w ramach projektu PLATON – Platforma Obsługi Nauki (2009-2011). Obecnie jest utrzymywana i świadczona przez centra KMD i operatorów sieci MAN w ramach konsorcjum sieci naukowej PIONIER – Polski Internet Optyczny. Jej koordynatorem jest Poznańskie Centrum Superkomputerowo-Sieciowe. W ramach usługi partnerzy Konsorcjum oferują instytucjom z sektora nauki i kultury możliwość długoterminowego przechowywania danych archiwalnych oraz zapasowych kopii danych. „Łączna przestrzeń przechowywania danych w systemie, to ok. 12.5 PB na 5 bibliotekach taśmowych w technologii LTO5 oraz około 2 PB na macierzach dyskowych i serwerach plików. Pamięci taśmowe zapewniają możliwość długoterminowego i bezpiecznego oraz efektywnego ekonomicznie przechowywania dużych wolumenów danych. Natomiast pamięci dyskowe stanowią bufor przyspieszający składowanie i dostęp do danych. Systemy przechowywania są rozmieszczone w 10 lokalizacjach: w 5 jednostkach KMD zainstalowane są biblioteki taśmowe i macierze dyskowe, w 5 jednostkach MAN – serwery plików. Umożliwia to składowanie danych blisko klienta końcowego, co ma duże znaczenie dla wydajności transmisji dużych wolumenów danych (terabajty na dobę)". Usługodawca podkreśla, że „powyższe zasoby można w miarę wzrostu potrzeb rozbudowywać do większych pojemności i to z gwarancją takiego samego, wysokiego poziomu jakości i dostępności usługi". „Rozproszona infrastruktura sprzętowo-programowa gwarantuje również geograficzną replikację danych. Funkcje te zapewnia dedykowane oprogramowanie – zaprojektowane, zoptymalizowane i skonfigurowane dla tej usługi, opracowane przez PCSS i partnerów

w ramach projektu Krajowy Magazyn Danych (2007-2009 oraz 2011-2013). Replikacja zapewnia długoterminowe bezpieczeństwo danych, a także ich dostępność pomimo lokalnych awarii sieci lub systemów przechowywania. Rozproszone geograficznie i redundantne środowisko zapewnia ciągłą możliwość składowania i dostępu do usługi Powszechnej Archiwizacji. Istotnym czynnikiem dla zapewnienia niezawodności usługi jest infrastruktura IT będąca na wyposażeniu 10 centrów danych. Obejmuje ona niezbędne do bezpiecznego funkcjonowania systemy, włączając w to klimatyzację, redundantne zasilanie, ochronę dostępu oraz odpowiednie przyłącza sieciowe: minimum o przepływności 10Gbit/s, a w praktyce wielokrotność tej przepustowości. Uzupełnieniem infrastruktury są wdrożone mechanizmy oraz procedury wsparcia i obsługi użytkowników. Obejmują one m.in.: portal informacyjny (umożliwiający rejestrację do usługi), zgłaszanie pytań i problemów, a także lokalne punkty obsługi użytkowników. Do zadań tych punktów należą: obsługa rejestracji użytkowników, opracowanie konfiguracji usługi zgodnie z wymaganiami poszczególnych instytucji (np. tryb replikacji, zarządzanie prawami dostępu) oraz obsługa lokalnych zgłoszeń dotyczących bieżącej obsługi. Dzięki rozproszeniu geograficznemu operatorów infrastruktury oraz hierarchicznej strukturze systemu wsparcia użytkowników, możliwa jest płynna obsługa ponad 250 instytucji klienckich. W powiązaniu z usługą Powszechnej Archiwizacji funkcjonuje również aplikacja dArceo, służąca do długoterminowego przechowywania informacji na poziomie obiektów cyfrowych w formie ustandaryzowanego archiwalnego pakietu informacji AIP (ang. *Archival Information Package*). System dArceo jest zgodny z międzynarodowym modelem OAIS (ang. *Open Archival Information System*), a jego funkcje pozwalają na realizację dowolnego scenariusza przeformatowania danych źródłowych. Przykładowo dArceo umożliwia ciągłe przetwarzanie danych źródłowych do formatów udostępniania w bibliotekach cyfrowych. Jednocześnie wybrane dane mogą być w innym scenariuszu przetwarzane pod kątem wykorzystania ich do celów badawczych lub edukacyjnych. Aktualnie w Polsce system dArceo wykorzystywany jest przez kilkadziesiąt instytucji z sektora nauki i kultury”[6].

Wobec powyższego PCSS byłoby właściwym partnerem w działaniach archiwizacyjnych polskich instytucji pamięci. Od wielu lat pracownicy Centrum są skupieni na pracach związanych z tworzeniem i zarządzaniem polskimi bibliotekami cyfrowymi, z rozwojem projektu Federacja Bibliotek Cyfrowych, prezentowaniem polskiej kolekcji cyfrowej w zbiorach Europeany. Znają potrzeby środowisk odpowiedzialnych za zgromadzenie i przechowanie dziedzictwa nauki i kultury. Są również świadomi potrzeby działań na rzecz długoterminowej archiwizacji zasobów cyfrowych, posiadają wiedzę i doświadczenie potrzebne w tych procesach.

[6] Materiał przygotowany i udostępniony przez PCSS w 2016 r.

Prawdopodobnie byliby cennymi doradcami w zakresie technicznych zagadnień długoterminowej archiwizacji i tworzenia narodowego systemu depozytowego.

Usługi Poznańskiego Centrum Superkomputerowo-Sieciowego oraz inne podobne inicjatywy będą miały w Polsce spore szanse rozwoju. W cytowanym wielokrotnie już raporcie MKiDN z września 2009 r., dużą uwagę przywiązywano do budowy bezpiecznych repozytoriów i magazynów danych dla polskich zasobów cyfrowych. Z raportu nie wynika wprawdzie, jakie dokładnie metody, narzędzia i środki, a także wzorce, normy i standardy będą stanowić podstawę polskiej strategii archiwizacji długoterminowej. Jednak nie ulega wątpliwości, że w koncepcję dotyczącą ochrony polskich zasobów cyfrowych wpisana została potrzeba zbudowania sieci bezpiecznych repozytoriów i magazynów danych. Uwzględniono w niej również potrzebę odniesienia się w pracach nad tworzeniem polskiego programu ochrony do programów już istniejących [Program digitalizacji, 2009, s. 30].

3.8. Koncepcja programu ochrony polskich zasobów cyfrowych

Zgodnie z treściami Programu digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020 [Program digitalizacji, 2009] oraz Zaleceniami Komisji Europejskiej [Zalecenie, 2011] publikacje cyfrowe zgromadzone w polskich instytucjach nauki i kultury powinny zostać objęte programem trwałej ochrony, zapewniającym bezpieczeństwo ich nienaruszalności oraz przechowanie w formatach, które umożliwią korzystanie z ich treści w długiej perspektywie czasowej. W związku z tym została opracowana i opublikowana propozycja programu trwałej ochrony zasobów cyfrowych. W pierwotnej wersji, z 2010 r., proponowany program dotyczył organizacji zadań ochrony zbiorów cyfrowych polskich bibliotek, pod przewodnictwem Biblioteki Narodowej [Januszko-Szakiel, 2011b, s. 211-230]. Dwa lata później przedłożono jego wersję drugą, w ujęciu rozszerzonym, uwzględniającym również zasoby cyfrowe archiwów, muzeów, instytucji gromadzących i przechowujących dokumenty audiowizualne, instytutów badawczych i rozwojowych oraz innych typów podmiotów, które w toku swej działalności tworzą świadectwa polskiej nauki i kultury, zasługujące na trwałą ochronę z myślą o przyszłych użytkownikach. W hipotetycznym scenariuszu rolę koordynatora działań na rzecz opracowania narodowej strategii trwałej ochrony polskiego dziedzictwa nauki i kultury w postaci cyfrowej powierzono Ministerstwu Kultury i Dziedzictwa Narodowego [Januszko-Szakiel, 2013, s. 173-199].

W prezentowanej w niniejszej książce wersji trzeciej programu wprowadza się kolejne zmiany. W najnowszej koncepcji zakłada się, że podmioty, których zasoby mogą tworzyć zasób cyfrowy, podlegający trwałej ochronie, to głównie

wydawnictwa, biblioteki, archiwa, muzea, instytuty audiowizualne, instytucje kształcące, instytuty naukowe, instytuty badawczo-rozwojowe. Jednak potrzeba profesjonalnego zarządzania materiałami cyfrowymi i ich trwałej ochrony jest coraz częściej zgłaszana również przez podmioty sektora administracji i biznesu. Dlatego zaproponowany program odnosi się do zadań ochrony materiału cyfrowego nie tylko sektora nauki i kultury, ale również biznesu i administracji. Uwzględnia się fakt, że nie wszystkie zasoby cyfrowe występujące w obiegu obu sektorów mają wartość ponadczasową. Dyskusyjny zatem może być pomysł włączenia ich wszystkich do centralnego archiwum Polski.

W projektowanej koncepcji zasoby wygenerowane bądź zgromadzone w podmiotach pamięci, kultury, nauki, administracji i biznesu współtworzą **Polski Centralny Depozyt Cyfrowy** (PCDC). PCDC to archiwum Polski, integrujące i organizujące dostęp poprzez metadane do możliwie kompletnego polskiego zasobu cyfrowego, umożliwiające wyszukanie i udostępnienie rozmaitych treści cyfrowych z jednego miejsca (za pomocą centralnej multiwyszukiwarki).

PCDC ma pełnić różne zadania. Podstawowe to zabezpieczenie zasobów, które podlegają ochronie wieczystej i stanowią o polskim dziedzictwie nauki i kultury, są eksponowane w europejskich i światowych kolekcjach cyfrowych. Kolejne ważne zadanie to zagregowanie i centralne udostępnienie informacji o zasobach cyfrowych innych typów podmiotów, głównie publicznych, ważnych dla społeczeństwa, dla operacyjnych i strategicznych celów podmiotów społecznych, gospodarczych, ekonomicznych, finansowych. Nie wyklucza się włączenia w projekt zasobów cyfrowych podmiotów prywatnych (na podstawie odpowiednich założeń finansowych świadczenia usług archiwizacyjnych dla sektora prywatnego). Ważne założenie projektowanego programu to dobrowolność i przyjmowanie w depozyt, w zależności od potrzeb i preferencji podmiotów współpracujących i deponujących zasoby w PCDC albo tylko metadanych dokumentów cyfrowych i ich kopii bezpieczeństwa, albo dokumentów w wersji zarówno prezentacyjnej (użytkowej), jak i wersji wzorcowej (tzw. plików master wysokiej jakości stanowiących tworzywo obiektu archiwalnego trwale chronionego). Podmioty generujące i gromadzące zasoby cyfrowe mogą tworzyć samodzielnie archiwa, repozytoria, depozyty cyfrowe (prywatne i publiczne, indywidualne, instytucjonalne, lokalne, regionalne, etc.) i trwale nimi zarządzać, odsyłając do PCDC tylko metadane (w celu ich włączenia do wyszukiwarki centralnej) oraz kopie zapasowe dokumentów (w celu ochrony). W innym trybie współpracy podmioty mogą odsyłać do PCDC wygenerowany bądź zgromadzony materiał cyfrowy wprost, nie tworząc własnych platform jego prezentacji, udostępniania oraz ochrony. Kolejne ważne założenie programu to organizacja przy PCDC centralnej pracowni cyfryzacji, z myślą o instytucjach, które przechowują zasoby warte przekształcenia do postaci cyfrowej i zaprezentowania ich w centralnej kolekcji cyfrowej Polski,

jednak z różnych powodów (np. z braku kompetencji, nieadekwatnej infrastruktury, zbyt małego zbioru dokumentów, aby zasadne było tworzenie pracowni cyfryzacji) nie decydują się na realizację zadań digitalizacyjnych, chcą je powierzyć podmiotom profesjonalnym, odpowiednio przygotowanym do tych prac.

Program odnosi się do różnych typów treści cyfrowych (danych, informacji, dokumentów urzędowych, raportów, analiz, czasopism, książek, obiektów muzealnych, nagrań, filmów, etc.; do dokumentów zdigitalizowanych oraz born digital, zapisanych (w pierwotnej wersji) na przenośnych mediach fizycznych typu: dyskietki, CD, DVD oraz w innych technologiach zapisu danych cyfrowych). Założono jednak, że z czasem, w toku zabiegów archiwizacyjnych, treść dokumentów cyfrowych będzie odłączana od oryginalnych mediów przenośnych i umieszczana na serwerach archiwalnych (depozytowych) oraz udostępniana w sieci Internet lub w sieciach wewnętrznych.

Założono, że w programie zostaną uwzględnione komptencje funkcjonujących już w naszym kraju podmiotów i realizowanych przedsięwzięć w obszarze ochrony zasobów cyfrowych, np. merytoryczne centra kompetencji, ustanowione przez MKiDN, Federacja Bibliotek Cyfrowych i PCSS, Archiwum Dokumentów Elektronicznych i NDAP, zapowiadany projekt KRONIK@ i MC. Identyfikacji i uwzględnienia wymagają też ewentualne funkcjonujące już przedsięwzięcia zarządzania zasobami cyfrowymi inicjowane na przykład przez ministerstwa ds. finansów, administracji, rozwoju.

Z uwagi na to, że projektowany PCDC ma objąć zasoby podlegające nadzorowi wielu resortów, zaproponowano ich współpracę i utworzenie międzyresortowego Komitetu Sterującego ds. organizacji i funkcjonowania PCDC, podlegającego i doradzającego bezpośrednio koordynatorowi PCDC powołanemu przez szefa rządu RP. Wobec zapowiedzi działań dotyczących utworzenia Krajowego Repozytorium Obiektów Nauki i Kultury KRONIK@ warte rozważenia wydaje się powołanie Ministerstwa Cyfryzacji do pełnienia roli koordynatora PCDC.

Proponowany program zawiera siedemnaście postulatów, przyporządkowanych do trzech grup:

3.8.1. Przyjęcie roli koordynatora działań archiwizacyjnych w Polsce

1. Rząd Polski (RP) powierza koordynatorowi działania na rzecz opracowania narodowej strategii trwałej ochrony polskich zasobów cyfrowych oraz utworzenie centralnego depozytu cyfrowego dla Polski.

Do zadań koordynatora należy wytyczanie oraz kierowanie zadaniami związanymi z ochroną, zabezpieczaniem i udostępnianiem zbiorów cyfrowych Polski.

W celu zabezpieczenia kompletnej kolekcji polskiego zasobu cyfrowego koordynator powołuje Komitet Sterujący ze struktur co najmniej: Ministerstwa Kultury

i Dziedzictwa Narodowego, Ministerstwa Nauki i Szkolnictwa Wyższego, Ministerstwa Cyfryzacji, Ministerstwa Rozwoju, Ministerstwa Finansów, Ministerstwa Spraw Wewnętrznych i Administracji, Polskiej Izby Książki, Poznańskiego Centrum Superkomputerowo-Sieciowego, uwzględniając doświadczenia i pełnione funkcje Biblioteki Narodowej, Narodowego Instytutu Dziedzictwa, Narodowego Instytutu Audiowizualnego, Narodowego Archiwum Cyfrowego, Narodowego Instytutu Muzealnictwa i Ochrony Zbiorów.

2. Koordynator ustala ekonomiczne podstawy działań na rzecz trwałej ochrony polskich zasobów cyfrowych.

Koordynator tworzy solidne, pochodzące z wiarygodnych źródeł, podstawy długoterminowego finansowania podjętych działań archiwizacyjnych. Oprócz środków pochodzących z budżetu państwa, koordynator oraz Komitet Sterujący poszukują źródeł finansowania w innych instytucjach rządowych, pozarządowych, państwowych i prywatnych.

3. Koordynator formułuje cele i założenia podejmowanego przedsięwzięcia.

W celu zapewnienia zgodności działań archiwizacyjnych z zasadami działalności archiwów wiarygodnych, koordynator definiuje cel i założenia podjętych działań, a także zasady, którymi będzie kierować się realizując nowe zadania.

Celem koordynatora jest zorganizowanie i scentralizowanie sieci wiarygodnych długoterminowych archiwów polskich zasobów cyfrowych.

W ogólnym założeniu, długoterminowe archiwum cyfrowe ma zapewnić użyteczność, czyli dostępność, autentyczność, integralność oraz poufność zdeponowanych w nim materiałów cyfrowych z myślą o potrzebach obecnych i przyszłych użytkowników.

Koordynator zakłada współodpowiedzialność i współdziałanie polskich instytucji kultury, nauki, administracji i biznesu na rzecz opracowania strategii długoterminowej archiwizacji polskich zasobów cyfrowych.

Koordynator zamierza uwzględnić dotychczasowe ustalenia międzynarodowych organizacji w zakresie trwałej ochrony zasobów cyfrowych i przyłączyć się do dyskusji toczącej się wśród organizatorów działań archiwizacyjnych różnych instytucji na świecie.

Koordynator zamierza kierować się w swej działalności następującymi zasadami:

- dokumentowanie wszelkich pomysłów, planów, rozwiązań organizacyjnych, wdrożeniowych, podstaw prawnych oraz ekonomicznych związanych z trwałą ochroną polskich zasobów cyfrowych;
- przejrzystość działań dotyczących organizowanej sieci archiwów; koordynator przekazuje do publicznej wiadomości dokumentację działań oraz zakłada staranną analizę i uwzględnienie wszelkich opinii wewnętrznych oraz zewnętrznych;

- adekwatność, czyli skrupulatna ocena przydatności i możliwości zastosowania we własnych warunkach, przyjętych w świecie rozwiązań, standardów i norm;
- ewaluacja rozwoju przedsięwzięcia, czyli poddawanie wewnętrznemu oraz zewnętrznemu opiniowaniu, jak koordynator radzi sobie z realizacją wytyczonych celów.

Ponadto koordynator zakłada, że inicjowane przedsięwzięcie będzie mieć charakter otwarty, co oznacza, że zarówno w procesach merytorycznych, jak i decyzyjnych, opiniodawczych oraz wykonawczych mogą udzielać się osoby nie tylko zaproszone, ale wszyscy zainteresowani tematyką, mogący i chcący pomóc.

4. Koordynator publikuje informacje o podejmowanym przedsięwzięciu.

W celu zapewnienia przejrzystości działań archiwizacyjnych koordynator publikuje informacje o podjętym przedsięwzięciu. Wszystkie instytucje, organizacje, podmioty w Polsce, w szczególności te, których zbiory tworzą narodowe dziedzictwo cyfrowe, podlegające szczególnej ochronie, otrzymują pełną informację o przedsięwzięciu RP.

5. Koordynator gromadzi wiedzę na temat trwałej ochrony zasobów cyfrowych.

Zgodnie z zaleceniami prezentowanymi w piśmiennictwie przedmiotu, koordynator rozpoczyna wszelkie prace związane z planowaniem działań archiwizacyjnych od starannej analizy opracowań przedmiotu oraz informacji z wszelkich dostępnych źródeł na temat zaleceń, standardów, norm, opinii, wreszcie od konsultacji z fachowcami odnośnie do zamierzeń i sposobów ich realizacji. Koordynator zapoznaje się ze sposobem organizowania prac archiwizacyjnych w instytucjach nauki i kultury, administracji i biznesu innych krajów, zaawansowanych i mogących pełnić rolę doradcy. Koordynator zapoznaje się też z działalnością i zaleceniami międzynarodowych organizacji skupionych wokół zadań długoterminowej archiwizacji.

Z różnych źródeł i doświadczeń koordynator tworzy zarówno zasób wiedzy, jak i bazę wiedzy na temat organizowania działań archiwizacyjnych w Polsce.

6. Koordynator identyfikuje ewentualne rodzime projekty oraz plany programów ochrony dziedzictwa cyfrowego.

Z uwagi na zalecenia dotyczące unikania dublowania prac i środków potrzebnych na działania archiwizacyjne, koordynator stara się ustalić, czy w polskich instytucjach nauki i kultury, administracji i biznesu realizuje się, bądź planuje realizację programów długoterminowej ochrony dokumentów cyfrowych. W przypadku istnienia takich programów lub planów, koordynator konfrontuje ich cele, założenia i zakres z własnymi zamiarami. Koordynator rozważa połączenie projektów i kooperację.

3.8.2. Organizacja pracy w zakresie trwałej ochrony polskich zasobów cyfrowych

7. Koordynator zwołuje stałą konferencję na temat długoterminowej archiwizacji polskich zasobów cyfrowych.

Koordynator organizuje ogólnopolską konferencję na temat trwałej ochrony polskich zasobów cyfrowych oraz zaprasza do udziału przedstawicieli polskich wydawnictw, instytucji nauki i kultury, biznesu i administracji.

Koordynator deklaruje, że zwołana konferencja będzie mieć charakter spotkań organizowanych cyklicznie, najlepiej w ustalonych odstępach czasu i określonym miejscu.

Na pierwszej konferencji koordynator przestawia problematykę ochrony zasobów cyfrowych, podkreśla rangę problemu, podaje przykłady inicjatyw światowych oraz informuje o rozpoczętym przedsięwzięciu, przytaczając cele, założenia i zasady działania.

Koordynator zgłasza potrzebę ukonstytuowania ogólnopolskiej grupy roboczej do spraw opracowania strategii ochrony polskich zasobów cyfrowych i zaprasza w jej szeregi osoby, które chcą współpracować, posiadają wiedzę w zakresie organizacyjnych, technicznych, prawnych oraz ekonomicznych zagadnień tworzenia strategii archiwizacji zasobów cyfrowych. Celem grupy roboczej ma być ścisła współpraca z Komitetem Sterującym w zakresie decyzji i działań dotyczących utworzenia i funkcjonowania PCDC, rozumianego jako rozproszony system depozytowy polskich zasobów cyfrowych, w postaci sieci trwałych, wiarygodnych archiwów cyfrowych, pracujących na zunifikowanych zasadach, ściśle współpracujących.

8. Koordynator powołuje Ogólnopolską Grupę Roboczą do spraw trwałej ochrony polskich zasobów cyfrowych.

Koordynator wraz z Komitetem Sterującym formuje i powołuje Ogólnopolską Grupę Roboczą do zadań długoterminowej archiwizacji polskich zasobów cyfrowych, ustala strukturę organizacyjną Grupy z zespołem zarządzającym na czele.

Integralną częścią zespołu zarządzającego jest zespół ds. obsługi merytorycznej, złożony z ekspertów, znawców organizacyjnych, technicznych, prawnych oraz ekonomicznych aspektów długoterminowej archiwizacji zasobów cyfrowych. W skład tego zespołu wchodzą również doradcy z branży wydawniczej. Zespół ds. obsługi merytorycznej tworzą głównie eksperci krajowi, ale przewidywane jest również zaproszenie do współpracy fachowców zagranicznych, z krajów i instytucji zaawansowanych w pracach nad strategią długoterminowej ochrony zasobów cyfrowych.

Zespół ds. obsługi merytorycznej identyfikuje metody prac, rozwiązania, standardy, normy, akty prawne, etc. dotyczące archiwizacji zasobów cyfrowych, stosowane na świecie. Opracowuje plany działania i poddaje je ocenie wewnętrznej

przez zespół zarządzający oraz zewnętrznej przez opinię publiczną. Wszelkie sugestie uwzględnia modyfikując proponowany plan działania.

Przy zespole zarządzającym działa również zespół ds. zarządzania jakością i dokumentacji, który na bieżąco kontroluje poprawność i terminowość wykonywanych zadań w poszczególnych zespołach oraz gromadzi i zarządza dokumentacją z działalności grupy roboczej; obejmuje kontrolą oraz ewaluacją wszystkie procesy realizowane w ramach działalności archiwizacyjnej.

W strukturze organizacyjnej, obok zespołu zarządzającego, koordynator powołuje cztery następujące zespoły:

- Zespół ds. obsługi organizacyjnej działań archiwizacyjnych, składający się z archiwistów, bibliotekarzy, bibliotekoznawców, informatologów, muzealników, menedżerów kultury i oświaty, menadżerów zasobów cyfrowych uczelni wyższych, instytucji badawczych, badawczo-rozwojowych, menedżerów kolekcji dokumentów audiowizualnych, urzędników państwowych, przedsiębiorców;
- Zespół ds. obsługi technicznej działań archiwizacyjnych, składający się z fachowców branży IT, ściśle współpracujących z zespołem ds. obsługi organizacyjnej;
- Zespół ds. obsługi prawnej działań archiwizacyjnych, składający się z prawników, ściśle współpracujących z zespołem ds. obsługi organizacyjnej;
- Zespół ds. obsługi ekonomicznej działań archiwizacyjnych, tworzony przez ekonomistów, ściśle współpracujących z zespołem ds. obsługi organizacyjnej.

Wymienione zespoły mają charakter zadaniowy (wykonawczy). Ich zadania definiuje zespół zarządzający. Poszczególne zespoły zadaniowe mają liderów, którzy organizują i odpowiadają za terminową i właściwą pracę ich zespołów przed zespołem zarządzającym. Liderzy dokumentują na bieżąco działania zespołów, a dokumentację przekazują do podzespołu ds. zarządzania jakością i dokumentacji.

Przewiduje się ścisłą współpracę osób przynależących do różnych zespołów. W skład poszczególnych zespołów wchodzą osoby z różnych polskich instytucji, głównie bibliotek, muzeów, archiwów, instytutów audiowizualnych, centrów informacji naukowej, instytutów naukowych, instytutów badawczo-rozwojowych, uczelni wyższych, firm informatycznych, kancelarii prawnych, instytucji sektora administracji, biznesu, finansów, rozwoju.

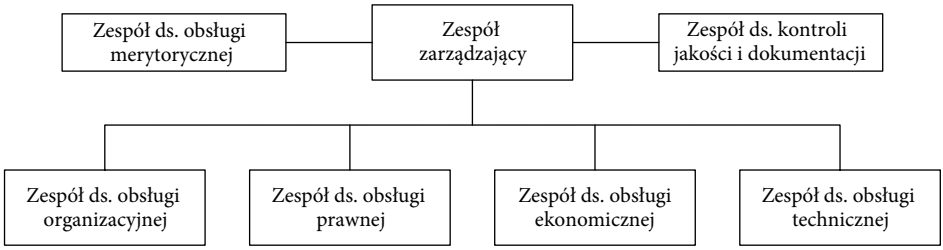
Praca w zespołach odbywa się w zależności od potrzeb, tradycyjnie bądź w trybie zdalnym.

Zespół zarządzający ustala sposób komunikowania się w obrębie poszczególnych zespołów oraz grupy roboczej.

Zespół zarządzający ustala częstotliwość i miejsce spotkań – w zależności od potrzeby – całej grupy roboczej, bądź liderów poszczególnych zespołów.

Proponowaną strukturę organizacyjną Ogólnopolskiej Grupy Roboczej ds. długoterminowej archiwizacji polskich zasobów cyfrowych przedstawiono na schemacie 1.

Schemat 1: **Struktura organizacyjna Ogólnopolskiej Grupy Roboczej ds. długoterminowej archiwizacji polskich zasobów cyfrowych**



Źródło: oprac. własne.

9. Koordynator definiuje ogólne cele Ogólnopolskiej Grupy Roboczej i poszczególnych zespołów.

- Cel(e) Ogólnopolskiej Grupy Roboczej ds. archiwizacji polskich zasobów cyfrowych:
  - utworzenie długoterminowego, wiarygodnego i stabilnego systemu depozytowego dla polskich zasobów cyfrowych;
  - zapewnienie długoterminowej użyteczności zasobów cyfrowych zgromadzonych w polskich instytucjach kultury i nauki, biznesu i administracji, z myślą o obecnych i przyszłych użytkownikach;
  - wyeksponowanie wyselekcjonowanej kolekcji polskich zasobów cyfrowych w kolekcji dziedzictwa światowego.
- Cel(e) zespołu zarządzającego:
  - organizowanie i koordynowanie pracy grupy roboczej.
- Cel(e) podzespołu ds. obsługi merytorycznej:
  - projektowanie polskich działań archiwizacyjnych na podstawie zgromadzonej wiedzy o organizowaniu długoterminowych i wiarygodnych archiwów cyfrowych.
- Cel(e) podzespołu ds. kontroli jakości i dokumentacji:
  - dbałość o dokumentację i przejrzystość polskich działań archiwizacyjnych;
  - ocena efektywności podejmowanych działań archiwizacyjnych.
- Cel(e) zespołu ds. obsługi organizacyjnej:
  - zapewnienie w długim czasie organizacyjnej sprawności i płynności polskich działań archiwizacyjnych.
- Cel(e) zespołu ds. obsługi technicznej:
  - zapewnienie w długim czasie technicznej ochrony i stabilności działania systemu depozytowego.
- Cel(e) zespołu ds. obsługi prawnej:
  - ukonstytuowanie długoterminowo obowiązującej polskiej *preservation policy*.

- Cel(e) zespołu ds. obsługi ekonomicznej:
  - zapewnienie źródeł stabilnego i długoterminowego finansowania polskich działań archiwizacyjnych;
  - zarządzanie finansami przedsięwzięcia dotyczącego utworzenia systemu depozytowego polskich zasobów cyfrowych.

10. Koordynator zakłada i prowadzi portal internetowy na temat trwałej ochrony polskich zasobów cyfrowych.

Koordynator zakłada portal WWW dedykowany sprawom długoterminowej archiwizacji polskich zasobów cyfrowych. Portal pełni funkcję wirtualnego informatorium o archiwizacji zasobów cyfrowych w Polsce; jest miejscem dyskusji, wymiany poglądów, publikowania informacji na temat rozwoju przedsięwzięcia.

Istnienie i sprawne funkcjonowanie portalu jest świadectwem przejrzystości działań archiwizacyjnych, tym samym czynnikiem budującym wiarygodność PCDC.

Elementem towarzyszącym serwisowi WWW jest elektroniczny biuletyn na temat archiwizacji, rozsyłany w formie newslettera do zainteresowanych i współpracujących podmiotów, w celu bieżącego informowania np. o terminach spotkań oraz rezultatach działań.

Koordynator stara się uzyskać opinię i akceptację zainteresowanych środowisk dla projektowanych działań archiwizacyjnych w Polsce. Tworzy rejestr instytucji i osób zainteresowanych współpracą i otrzymywaniem bieżących informacji na temat rozwoju przedsięwzięcia, a także rejestr osób i instytucji obserwujących i opiniujących rodzimą działalność archiwizacyjną.

3.8.3. Planowanie szczegółowych zadań Ogólnopolskiej Grupy Roboczej ds. trwałej ochrony polskich zasobów cyfrowych

11. Koordynator definiuje zadania szczegółowe zespołu zarządzającego.

Zespół zarządzający:

- wraz z zespołem ds. obsługi merytorycznej wytycza i planuje w czasie poszczególne zadania dla grup roboczych, z uwzględnieniem ich kolejności podyktowanej wynikaniem kolejnych zadań z wcześniejszych;
- zleca zadania poszczególnym zespołom, wraz z określeniem czasu ich wykonania i formą prezentacji wyników, a także sugestią dotyczącą sposobu wykonania zadań;
- organizuje spotkania z liderami poszczególnych zespołów, w celu zapoznawania się z postępami prac;
- wraz z zespołem ds. kontroli jakości i dokumentacji, analizuje wyniki i wnioski prac zespołów oraz przetwarza je do postaci spójnego dokumentu na temat kształtowania polityki długoterminowej ochrony polskich zasobów cyfrowych;
- prezentuje działania rodzime na forum krajowym i międzynarodowym.

12. Koordynator definiuje zadania szczegółowe zespołu ds. obsługi merytorycznej.

Zespół ds. obsługi merytorycznej:

- pełni funkcję merytorycznego zaplecza (źródła wiedzy) dla całości projektu;
- gromadzi wiedzę o działaniach archiwizacyjnych w instytucjach innych krajów, w różnych sektorach;
- identyfikuje i rejestruje istniejące rozwiązania, zalecenia, wzorce, normy, standardy, etc. dotyczące długoterminowej archiwizacji;
- szuka doradców w kraju i za granicą dla własnych rozwiązań;
- proponuje procedury i rozwiązania dotyczące działań archiwizacyjnych w Polsce;
- tworzy strategię długoterminowej archiwizacji polskich zasobów cyfrowych.

13. Koordynator definiuje zadania szczegółowe zespołu ds. kontroli jakości i dokumentacji.

Zespół ds. kontroli jakości i dokumentacji:

- kontroluje poprawność i terminowość wykonania zadań i procesów w ramach działalności archiwizacyjnej;
- opracowuje wytyczne dotyczące sporządzania dokumentacji z wszelkich czynności wykonywanych w poszczególnych zespołach grupy roboczej;
- sprawuje kontrolę nad terminowością, poprawnością i kompletnością sporządzania dokumentacji dotyczącej wszystkich zadań i procesów;
- gromadzi dokumentację z działalności grupy roboczej;
- zarządza dokumentacją z działalności grupy roboczej;
- organizuje dostęp do dokumentacji z działalności grupy roboczej;
- opracowuje i publikuje komunikaty na temat działalności archiwizacyjnej na stronach portalu i biuletynu;
- kontroluje prace i sposób zarządzania finansami przeznaczonymi na działania archiwizacyjne, przez zespół ds. obsługi ekonomicznej.

14. Koordynator definiuje zadania szczegółowe zespołu ds. obsługi organizacyjnej.

Zespół ds. obsługi organizacyjnej:

- organizuje badania dotyczące identyfikacji materiałów cyfrowych przechowywanych w polskich instytucjach, w szczególności w sektorze nauki i kultury;
- opracowuje i publikuje wytyczne dla twórców publikacji cyfrowych;
- organizuje szkolenia z zakresu publikowania elektronicznego;
- organizuje szkolenia z zakresu digitalizacji materiałów bibliotecznych, archiwalnych, muzealnych, etc.;
- opracowuje procedury oceny i selekcji materiałów cyfrowych, czyli zasady tworzenia kolekcji archiwalnych;

- organizuje współpracę wydawców z instytucjami archiwizującymi; zapoznaje wydawców z koncepcją i możliwymi podejściami do długoterminowej archiwizacji publikacji elektronicznych;
- wraz z zespołem ds. obsługi technicznej opracowuje wytyczne dotyczące stosowania standardowych formatów zapisu dokumentów cyfrowych;
- wraz z zespołem ds. obsługi technicznej oraz prawnej opracowuje politykę tworzenia metadanych; definiuje format, zakres oraz poziom szczegółowości metadanych archiwizowanych dokumentów;
- wraz z zespołem ds. obsługi technicznej ustala wytyczne dotyczące zgłaszania i przekazywania dokumentów do instytucji archiwizującej;
- organizuje szkolenia z zakresu długoterminowej archiwizacji dokumentów cyfrowych;
- organizuje stałą konferencję na temat długoterminowej archiwizacji polskich zasobów cyfrowych.

Zgodnie z zaleceniami i istniejącymi wzorcami postępowania, priorytetowe zadanie programów ochrony to identyfikacja najbardziej zagrożonych materiałów cyfrowych. Dlatego w proponowanym programie zakłada się organizację i przeprowadzenie badań, polegających na ustaleniu ilości zgromadzonych dokumentów cyfrowych, daty ich opublikowania, zastosowanego formatu, zastosowanego nośnika, zdefiniowaniu platformy sprzętowo-programowej potrzebnej do odczytu i prezentacji treści publikacji, dokonaniu krótkiej charakterystyki (oceny) ich wartości merytorycznej oraz określeniu stopnia zapotrzebowania zgłaszanego przez użytkowników.

W tym celu koordynator przygotowuje zalecenie dotyczące przeprowadzenia badania oraz tworzy rejestr instytucji objętych badaniem. Badanie jest skierowane przede wszystkim do instytucji, których zbiory tworzą narodowe dziedzictwo. Do badania przystępują także instytucje, które, według własnej opinii, posiadają w swych zbiorach materiały o szczególnej wartości, zasługujące na włączenie do PCDC i długoterminową ochronę.

Poszczególne instytucje, na mocy zaleceń koordynatora, powołują wewnętrzne zespoły do spraw ochrony zasobów cyfrowych. Ich zadaniem jest diagnoza zgromadzonych zasobów cyfrowych, w szczególności ich ocena i selekcja oraz utworzenie kolekcji materiałów, które mają ponadczasową wartość i zasługują na status dziedzictwa narodowego.

W zaleceniu dotyczącym badania zostaje określony cel badań, termin ich wykonania oraz wytyczne dotyczące sporządzenia dokumentacji z badań. Do rozporządzenia zostaje załączony specjalnie przygotowany formularz badania, służący do odnotowania informacji o badanych parametrach oraz ułatwiający sporządzenie dokumentacji z badania.

Zaleca się, aby potrzebne i wartościowe materiały cyfrowe, o dużym znaczeniu dla rozwoju Polski, wytypować do określonych zabiegów konserwatorskich. Instytucje podejmują próbę odczytu i prezentacji treści dokumentów cyfrowych, aby oddzielić zasoby nieużyteczne od użytecznych. Instytucje podejmują również próbę oceny wartości treści materiałów nieużytecznych i stopnia zapotrzebowania na nie zgłaszanego przez użytkowników. Instytucje ustalają, czy nieużyteczne dokumenty cyfrowe posiadają odpowiednik analogowy. W przypadku dokumentów o szczególnej wartości i dużym zapotrzebowaniu, które nie mają substytutów analogowych, podejmowane są wszelkie możliwe działania, umożliwiające odtworzenie treści dokumentu.

Opisy dokumentów, których próby odczytu i prezentacji treści z różnych powodów nie mogły zostać podjęte, bądź nie powiodły się, są odnotowywane w specjalnym rejestrze zasobów nieużytecznych. Zakłada się, że nieużyteczne materiały unikatowe i szczególnie wartościowe zostaną poddane bardziej skomplikowanym zabiegom, w celu odtworzenia ich treści. Zabiegi takie mogą być podejmowane przez poszczególne instytucje, w ramach ich możliwości finansowych, technicznych oraz merytorycznych, bądź zlecane firmom zewnętrznym. Zakłada się również organizację przy PCDC jednostki oferującej wsparcie przy pracach nad odtworzeniem treści dokumentów szczególnie ważnych.

Koordynator organizuje szkolenia dotyczące metod archiwizacji, tak by dokumenty cyfrowe, w zależności od potrzeby, zostały poddane operacjom odświeżenia nośnika, zmiany generacji nośnika, migracji bądź emulacji.

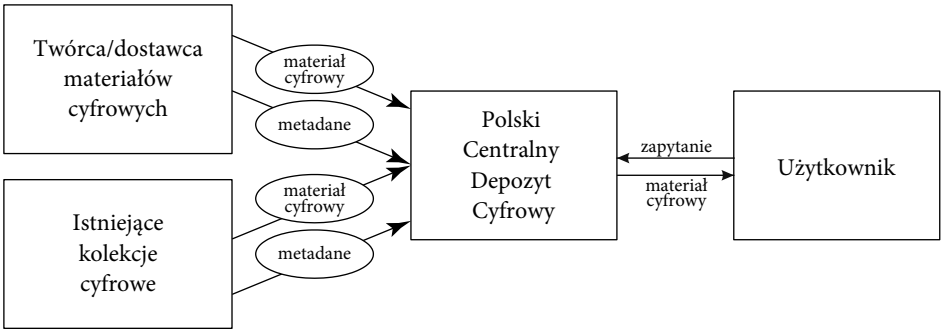
W programie przyjmuje się, że deponenci materiałów cyfrowych (wydawcy i miejsca wydawnicze, twórcy, szeroko pojęci dostawcy zasobów z różnych sektorów) decydują się w umowie z PCDC na określony model współpracy w ramach działalności archiwizacyjnej.

Pierwszy proponowany model współpracy (Schem. 2) zakłada, że materiał cyfrowy zostaje zgłoszony oraz odesłany do instytucji archiwizującej, w celu umieszczenia w systemie depozytowym. W metadanych użytkowych zostają określone warunki jego udostępniania. W modelu tym pełną odpowiedzialność za długoterminową ochronę treści cyfrowych ponosi instytucja archiwizująca. Rola dostawców polega na wygenerowaniu, następnie zgłoszeniu i przekazaniu materiału, wraz z odpowiednimi metadanymi, do instytucji archiwizującej, zgodnie z wytycznymi załączonymi do umowy.

W modelu drugim natomiast (Schem. 3) dostawcy samodzielnie realizują proces archiwizacji materiałów cyfrowych, decydując się na tzw. model *self archiving*, jednak również w ścisłej współpracy z instytucją archiwizującą.

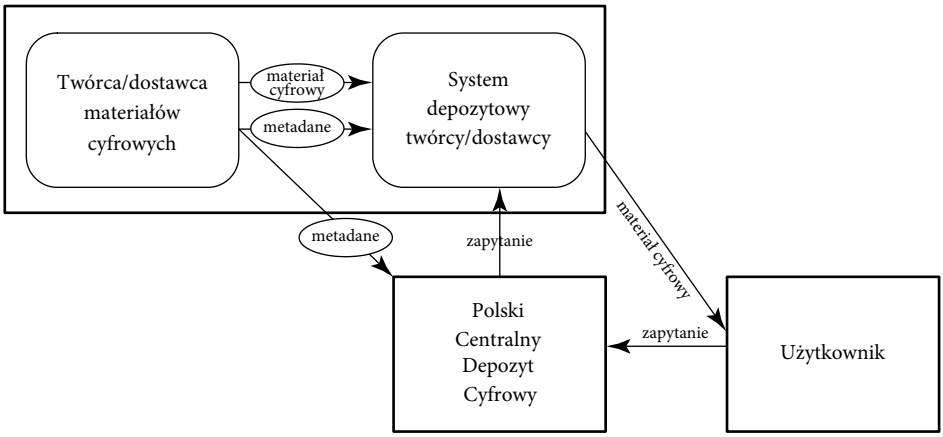
W drugim proponowanym podejściu podmioty tworzące, publikujące, zarządzające dokumentami cyfrowymi ponoszą koszty związane z organizowaniem procesu

Schemat 2: **Model współpracy deponentów (dostawców materiałów cyfrowych) z instytucją archiwizującą w zakresie długoterminowej archiwizacji zasobów cyfrowych**



Źródło: oprac. własne.

Schemat 3: **Model współpracy deponentów (dostawców metadanych) z instytucją archiwizującą w zakresie długoterminowej archiwizacji zasobów cyfrowych, tzw. model *self archiving***



Źródło: oprac. własne.

długoterminowej archiwizacji oraz dostępu do archiwizowanych zasobów. Ponoszą również odpowiedzialność za utrzymanie użyteczności deponowanych materiałów.

W omawianym modelu, do systemu depozytowego instytucji archiwizującej trafia jedynie zgłoszenie faktu utworzenia dokumentu oraz przekazanie jego metadanych. Udostępnianie treści dokumentu użytkownikom odbywa się bądź bezpośrednio z systemu depozytowego określonego podmiotu, bądź za pośrednictwem instytucji archiwizującej.

W modelu *self archiving* istotna jest świadomość dotycząca złożoności i odpowiedzialności związanej z procesem archiwizacji zasobów. Konieczne jest również

ustalenie procedur, według których poszczególne podmioty miałyby trwale chronić zdeponowane u siebie zasoby.

W obu modelach bardzo istotne jest zapewnienie technicznej obsługi, związanej z ochroną użyteczności i organizacją dostępu do archiwizowanych zasobów.

W programie przyjęto, że istotne znaczenie ma przygotowanie oferty szkoleń z zakresu tworzenia i archiwizacji dokumentów cyfrowych. Instytucja zarządzająca tworzy wykaz tematów szkoleń oraz poszukuje instruktorów w kraju i za granicą do ich prowadzenia. Proponowane przedmioty szkoleń to:

- publikowanie elektroniczne;
- digitalizacja zbiorów;
- formaty zapisu dokumentów cyfrowych, z uwzględnieniem formatów dla różnych typów dokumentów;
- tworzenie metadanych dokumentów cyfrowych i formaty ich zapisu;
- systemy trwałego identyfikowania dokumentów cyfrowych;
- długoterminowa archiwizacja dokumentów cyfrowych – strategie, metody, techniki, narzędzia;
- systemy depozytowe zasobów cyfrowych i ich bezpieczeństwo;
- aspekty prawne długoterminowej archiwizacji zasobów cyfrowych;
- aspekty ekonomiczne długoterminowej archiwizacji zasobów cyfrowych.

W zależności od zgłoszonego zapotrzebowania na określone szkolenie, instytucja zarządzająca organizuje szkolenie i ustala miejsce jego przeprowadzenia.

Zakłada się, iż proponowany wykaz tematów szkoleń stanowi ofertę wstępną. Z czasem powinna być rozszerzana o szkolenia dotyczące zagadnień bardziej szczegółowych, wskazujących konkretne rozwiązania poszczególnych zadań organizacyjnych, technicznych, prawnych oraz ekonomicznych procesu długoterminowej archiwizacji zasobów cyfrowych.

Dodatkowo, proponuje się organizację cyklicznej międzynarodowej konferencji na temat długoterminowej archiwizacji zasobów cyfrowych. W założeniu konferencja jest podzielona na dwie sesje. Celem sesji pierwszej jest raportowanie o stanie badań, najnowszych osiągnięciach, trendach i kierunkach rozwoju długoterminowej archiwizacji na podstawie efektów pracy krajów zaawansowanych w działaniach archiwizacyjnych. Do wystąpienia w tej sesji zaprasza się prelegentów z instytucji zagranicznych, których osiągnięcia i doświadczenia są szczególnie interesujące i mogą usprawnić organizowanie działań archiwizacyjnych w Polsce.

Przedmiotem sesji drugiej natomiast jest prezentowanie celu i założeń działań rodzimych, informowanie o stopniu ich zaawansowania oraz najpilniejszych zadaniach i planach. Ważną częścią tej sesji jest dyskusja, ukierunkowana na gromadzenie opinii i pomysłów odnośnie do archiwizacji polskich zasobów cyfrowych.

Konferencja ma być okazją do nawiązania kontaktów z fachowcami, praktykami z zakresu archiwizacji, którzy mogą pełnić rolę instruktorów dla polskich kadr odpowiedzialnych za zadania archiwizacji. Ponadto mogą recenzować polskie plany i działania archiwizacyjne, a także doradzać instytucji zarządzającej. Konferencja powinna również dostarczyć sposobność ustalenia, z którą instytucją można nawiązać współpracę w zakresie przekazania jej pod ochronę kopii polskiej kolekcji archiwalnej.

15. Koordynator definiuje zadania szczegółowe zespołu ds. obsługi prawnej. Zespół ds. obsługi prawnej:

- poddaje ekspertyzie obowiązujące akty prawne odnoszące się do zagadnień gromadzenia, archiwizowania oraz udostępniania wszelkich typów i form dokumentów cyfrowych oraz określa potrzebę i zakres ich uzupełnienia;
- konstytuuje polską *preservation policy*, czyli opracowuje propozycje wszelkich przepisów, tworzących podstawy prawne dla przyjęcia odpowiedzialności, podjęcia i prowadzenia działań na rzecz archiwizacji polskich zasobów cyfrowych, w tym:
  - pracuje nad nowelizacją ustawy o bibliotekach i obowiązkowych egzemplarzach bibliotecznych: celem noweli prawa ma być umożliwienie zgromadzenia w systemie depozytowym wszelkich form publikacji elektronicznych, także sieciowych; nowe ustawodawstwo ma zobligować twórców do zgłaszania i odsyłania wszystkich publikacji do depozytu oraz bibliotekę narodową – do ich wieczystej archiwizacji,
  - pracuje nad nowelizacją ustawy o prawie autorskim i prawach pokrewnych: celem noweli prawa ma być nadanie instytucji archiwizującej prawa do działań konserwatorskich na publikacjach elektronicznych, z uwzględnieniem ewentualnych zmian treści i formy publikacji, wywołanych tymi pracami,
  - przygotowuje wzory umów dotyczących współpracy z wydawcami;
- zapewnia obsługę prawną współpracy instytucji archiwizującej z instytucją partnerską i następczą;
- negocjuje prawa do prac konserwatorskich na publikacjach elektronicznych z właścicielami praw (do czasu wejścia w życie noweli prawa regulującego tę kwestię);
- zapewnia obsługę prawną współpracy deponentów/dostawców z instytucją archiwizującą;
- wraz z zespołem ds. obsługi organizacyjnej oraz technicznej ustala procedury tworzenia metadanych użytkowych dla archiwizowanych obiektów.

W programie zakłada się, że procesy gromadzenia zasobów cyfrowych w PCDC i działalności archiwizacyjnej odbywają się na podstawie dobrowolnych umów deponentów i instytucji archiwizującej.

Do wszystkich wydawców, miejsc wydawniczych, jednostek organizacyjnych i osób fizycznych nie będących wydawcami, ale prowadzących działalność polegającą na tworzeniu materiałów cyfrowych, zostaje skierowany pakiet informacji przybliżający cele i założenia projektu PCDC i zachęcający do zgłaszania i nadsyłania materiałów cyfrowych w celach ich archiwizacji. Zostaje również sporządzony dokument, określający warunki współpracy z twórcami deklarującymi udział w projekcie oraz wzór umowy o współpracy. W warunkach współpracy określa się:

- parametry dotyczące formatu zapisu i nośnika dokumentu cyfrowego;
- parametry dotyczące utworzenia podstawowego pakietu metadanych opisowych, technicznych oraz użytkowych;
- parametry dotyczące jakości oraz kompletności tzw. pakietów zgłoszeniowych; oprócz dokumentu i metadanych utworzonych według wytycznych zdefiniowanych przez instytucję archiwizującą, dopuszcza się określenie w indywidualnych umowach z podmiotami zainteresowanymi współpracą, w zależności od typu i specyfiki zgłaszanych materiałów elektronicznych, dodatkowych parametrów dla pakietów zgłoszeniowych;
- czas od daty opublikowania dokumentu do zgłoszenia i przekazania pakietu zgłoszeniowego do instytucji archiwizującej;
- wytyczne dotyczące sposobu przekazania pakietu zgłoszeniowego do instytucji archiwizującej;
- założenia dotyczące karencji pierwszego udostępnienia archiwizowanej publikacji;
- wytyczne dotyczące dezaktywacji mechanizmów kontroli użytkowania dokumentów, np. DRM;
- założenia dotyczące przeprowadzania niezbędnych prac konserwatorskich na obiektach archiwalnych, w celu zapewnienia ich długoterminowej użyteczności.

16. Koordynator definiuje zadania szczegółowe zespołu ds. obsługi ekonomicznej.

Zespół ds. obsługi ekonomicznej:

- ustala realne źródła długoterminowego finansowania polskich działań archiwizacyjnych;
- tworzy „trwały model finansowania" polskich działań archiwizacyjnych;
- ustala plany finansowania poszczególnych zadań, etapów prac, itp.;
- szacuje koszty poszczególnych działań archiwizacyjnych;

- kontroluje wpływy i wydatki instytucji archiwizującej;
- tworzy raporty z procesu zarządzania finansami przeznaczonymi na działania grupy roboczej oraz przekazuje je zespołowi ds. kontroli jakości i dokumentacji.

W proponowanym scenariuszu zakłada się, że podjęcie inicjatywy długoterminowego archiwizowania polskich zasobów cyfrowych i koordynowania polskich działań archiwizacyjnych odbywa się na podstawie finansowego wsparcia ze strony RP.

Z uwagi na długoterminowy charakter przedsięwzięcia, instytucja zarządzająca stara się uzyskać potwierdzenie stabilności finansowania działań archiwizacyjnych w długim czasie.

Rozwiązaniem optymalnym byłoby potwierdzenie pełnego finansowania działalności archiwizacyjnej. Jednak bardziej prawdopodobne wydaje się, iż potrzebne będzie ujmowanie poszczególnych działań archiwizacyjnych w projekty częściowe i wnioskowanie o ich finansowanie do różnych instytucji. W związku z tym tworzy się rejestr wszelkich rządowych, pozarządowych, państwowych i prywatnych instytucji, organizacji, fundacji, stowarzyszeń, które finansują tego typu projekty.

Instytucja zarządzająca od początku działań opracowuje preliminarz dotyczący działalności archiwizacyjnej. Planując poszczególne zadania, szacuje ich koszty, uwzględniając wartość wszelkich środków potrzebnych do ich wykonania. Koszty zakładane konfrontuje następnie z kosztami faktycznymi. Na tej podstawie opracowuje plany finansowania kolejnych zadań archiwizacyjnych, przedstawiając je we wnioskach o finansowanie.

Planowanie działań archiwizacyjnych, szacowanie ich kosztów oraz pozyskiwanie środków na ich finansowanie odbywa się z określonym wyprzedzeniem, tak aby możliwe było zapewnienie ciągłości realizacji zadań. W kosztorysach uwzględnia się koszty nieprzewidziane, aby nie dopuścić do przekroczenia sumy planowanych kosztów, tym samym do sytuacji niewypłacalności, bądź zawieszenia działań.

W celu zagwarantowania przejrzystości działań archiwizacyjnych instytucja zarządzająca prowadzi i w razie potrzeby udostępnia, skrupulatną dokumentację wszelkich wpływów i ich źródeł oraz poniesionych kosztów.

17. Koordynator definiuje zadania szczegółowe zespołu ds. obsługi technicznej.

Zespół ds. obsługi technicznej:

- opracowuje koncepcję organizacji i funkcjonowania systemu depozytowego dla polskich zasobów cyfrowych; istotne jest rozstrzygnięcie, czy zasoby składowane w systemie depozytowym będą udostępniane na bieżąco, czy tylko składowane i chronione z myślą o przyszłych użytkownikach;

- wraz z zespołem ds. obsługi merytorycznej ustala, której firmie informatycznej może zostać powierzone zadanie zaprojektowania, stworzenia i implementacji systemu depozytowego;
- ustala politykę bezpieczeństwa systemu depozytowego;
- ustala techniczne parametry dla wejściowych obiektów cyfrowych, czyli definiuje, jakie cechy charakterystyczne powinny posiadać dokumenty, aby zostały przyjęte do systemu depozytowego i objęte ochroną;
- ustala parametry wejściowych metadanych dla obiektów cyfrowych, następnie uzupełnia je i chroni wraz z obiektem; szczególnie istotne są metadane techniczne (dotyczące technicznych parametrów dokumentów oraz wszelkich elementów do nich przynależących, umożliwiających zarządzanie pracami konserwatorskimi na obiektach) oraz użytkowe (dotyczące praw i warunków udostępniania i użytkowania);
- ustala politykę postępowania z obiektami cyfrowymi, w której najważniejsze elementy to:
  - „wyciągnięcie" treści dokumentów zagrożonych z dotychczasowych środowisk do otoczenia bezpiecznego, czyli umieszczenie ich na serwerze archiwalnym,
  - strategia sporządzania kopii zapasowych kolekcji,
  - nadanie uprawnień dotyczących dostępu do systemu depozytowego i przeprowadzania prac konserwatorskich na archiwizowanych obiektach,
  - ustalenie parametrów wyjściowych obiektów cyfrowych, dotyczących przeszukiwania zasobów, udostępniania oraz zakresu użytkowania,
  - opracowanie systemu trwałego identyfikowania obiektów archiwalnych;
- monitoruje zmiany technologiczne i dostosowuje do nich przyjętą politykę archiwizacji;
- ustala poziom niezawodności procesu ochrony autentyczności i integralności zasobów archiwalnych (definiuje dopuszczalny poziom błędu, odstępstwa od wersji pierwotnej);
- wraz z zespołem ds. obsługi organizacyjnej oraz merytorycznej poszukuje instytucji partnerskiej, która przejmie odpowiedzialność za zabezpieczenie kopii polskiej kolekcji archiwalnej, w zamian za ochronę kolekcji instytucji partnerskiej oraz instytucji następczej.

W programie proponuje się formę archiwum cyfrowego, w którym zasoby są udostępniane na bieżąco. Zakłada się, że podmioty zainteresowane odsyłają swoje materiały cyfrowe do PCDC w celach zagregowania informacji o polskich zasobach cyfrowych w postaci pakietów zgłoszeniowych, zdefiniowanych przez insty-

tucję archiwizującą (za OAIS: Submission Information Package, SIP). W dziale gromadzenia pakiety zgłoszeniowe zostają poddane procesowi kontroli kompletności, opracowania oraz oceny i selekcji, których celem jest identyfikacja cech publikacji, decydujących o ich oznaczeniu w metadanych jako wyselekcjonowanej kolekcji dóbr nauki i kultury, tym samym poddaniu ich procesowi szczególnej ochrony. Następnie zasoby cyfrowe wraz z ich metadanymi, w postaci pakietów archiwalnych (za OAIS: Archive Information Package, AIP) zostają przekazane do systemu depozytowego. Po zamówieniu są udostępniane użytkownikom w postaci pakietów udostępnianych (użytkowych) (za OAIS: Dissemination Information Package).

Istnieje potrzeba opracowania procedur dotyczących formatów zapisu i opisu, a także identyfikowania obiektów cyfrowych, umożliwiających utrzymanie ich autentyczności i integralności w długim czasie oraz gwarantujących ich stabilną dostępność i pełną użyteczność. Na podstawie literatury przedmiotu oraz zaleceń doświadczonych instytucji archiwizujących zakłada się stosowanie formatów, które umożliwią bezpieczne przetrwanie zasobów cyfrowych przez okres przejściowy, do czasu opracowania formatów typowo archiwalnych.

Podobną propozycję instytucja archiwizująca wysuwa w związku z potrzebą opisania zasobów cyfrowych, czyli utworzenia ich metadanych. Rozważa się tworzenie metadanych zgodnie z formatem METS bądź też DC, bardziej rozpowszechnionym w Polsce, stosowanym m.in. w Federacji Bibliotek Cyfrowych. Jednak z uwagi na dostosowywanie polskiej strategii do istniejących wzorów, zaleca się w procesie decyzyjnym, dotyczącym najistotniejszych kwestii technicznych, takich właśnie jak wybór formatu zapisu i opisu dokumentów cyfrowych oraz system ich trwałego identyfikowania, dokonać starannej analizy rozwiązań zastosowanych w instytucjach zaawansowanych, ze szczególnym uwzględnieniem strategii tych instytucji, które typowane są do roli instytucji partnerskiej. Wskazane jest również przeprowadzenie konsultacji z ekspertami pochodzącymi właśnie z tych instytucji.

Utrzymanie ciągłości ochrony zasobów archiwalnych jest nadrzędnym celem archiwum. W programie zakłada się, że przedsięwzięcie PCDC ma charakter długoterminowy, w sensie wieczysty. W takich programach przyjmuje się, iż możliwa stanie się potrzeba przekazania obowiązków ochrony systemu depozytowego. Stąd też należy zawrzeć umowę z instytucją właściwą, w znaczeniu odpowiednio przygotowaną, do przejęcia i kontynuacji zadań ochrony depozytu, a także przygotować i utrzymywać aktualną dokumentację, dotyczącą gabarytu kolekcji archiwalnej, przyjętej strategii jej ochrony, planowanych i wykonanych prac konserwatorskich, wraz z opisem ich przebiegu i uzyskanego efektu.

Zaproponowany program zakłada ścisłą współpracę Ogólnopolskiej Grupy Roboczej, ukonstytuowanej z przedstawicieli wszystkich zainteresowanych środowisk, których dotyczy zagadnienie trwałej ochrony zasobów cyfrowych (głównie z bibliotekarzy, bibliotekoznawców, informatologów, archiwistów, muzealników, menedżerów kultury i oświaty, prawników, informatyków, ekonomistów) w celu opracowania strategii postępowania z cyfrowymi zasobami polskich instytucji nauki i kultury oraz administracji i biznesu i utworzenia centralnego archiwum cyfrowego.

Utworzenie archiwum centralnego nie wyklucza powstawania i funkcjonowania programów cząstkowych, dotyczących kolekcji regionalnych, lokalnych bądź instytucjonalnych.

Należałoby jednak zastanowić się nad zasadnością takiego podejścia; jeśli bowiem centralny depozyt zostanie dobrze opracowany, będzie właściwie zarządzany i chroniony, oraz będzie spełniał zakładane oczekiwania, czyli zapewni przejęcie oraz długoterminową ochronę polskich zasobów cyfrowych, to kolejne archiwa będą generowały dodatkowe koszty.

Aby przedłożony scenariusz miał szansę sprawdzić się w rzeczywistości, należałoby poddać go stosownym analizom, przekształcić do postaci wykonalnego projektu, oszacować koszty jego realizacji.

Z rozmów prowadzonych z przedstawicielami czołowych instytucji nauki i kultury Polski (MKiDN, BN, NAC) w latach 2016-2017 wynika ich wyraźna skłonność do przyjęcia w naszym kraju modelu systemu depozytowego rozproszonego. Bardzo prawdopodobne jest, że cyfrową kolekcję polskiego dziedzictwa nauki i kultury utworzą zasoby składowane i chronione w instytucjach powołanych przez MKiDN do pełnienia roli centrów kompetencji.

Należy jednak mieć świadomość, że rozwój i docelowy kształt depozytu polskiego zasobu cyfrowego będą, prawdopodobnie, wynikać z założeń inicjatorów projektu KRONIK@.

Podsumowanie

Nietrwałość zapisów cyfrowych została niejednokrotnie stwierdzona i poparta dobitnymi przykładami. Podstawowe zagrożenia dla materiałów cyfrowych wynikają z niskiej trwałości, rozpadu, zniszczenia nośnika, starzenia się formatu. Na problemy użytkowania dokumentów cyfrowych mają także wpływ zmiany technologiczne i powiązane z nimi wyjście z powszechnego użycia sprzętu i oprogramowania potrzebnych do odczytu i prezentacji treści zapisanych w cyfrowej postaci. Kolejnym istotnym zagrożeniem, prawdopodobnie poważniejszym niż same zmiany technologiczne, jest brak świadomości odnośnie do potrzeby ich obserwacji i właściwej reakcji. Dopiero utrata ważnych danych zapisanych cyfrowo w wielu instytucjach na świecie uświadomiła osobom odpowiedzialnym za ich przechowanie, na czym polegał popełniony błąd. Zdano sobie sprawę, że nośnika cyfrowego nie można odłożyć na półkę magazynową, by wrócić po niego za kilkanaście bądź kilkadziesiąt lat. Tym samym przekonano się, jak istotna i potrzebna jest ochrona zasobów cyfrowych, szczególnie tych, których treść ma wartość ponadczasową. Bezpowrotna utrata możliwości odczytu pierwszych zapisów cyfrowych dała asumpt do badań w zakresie archiwizacji zasobów cyfrowych.

Jedne z pierwszych udokumentowanych prac badawczych z zakresu archiwizacji materiałów cyfrowych zostały zorganizowane w 1998 r. przez Research Library Group. Polegały one na przeglądzie dokumentów elektronicznych zgromadzonych głównie w bibliotekach, archiwach, muzeach i innych instytucjach sektora nauki i kultury, w celu ustalenia ilości oraz jakości zgromadzonych dokumentów, tj. zastosowanych nośników, formatów, otoczenia sprzętowo-programowego odczytu i prezentacji ich treści w formie zrozumiałej dla użytkownika, a przede wszystkim w celu identyfikacji zasobów, do których treści dostęp utracono. Zarówno z raportów Research Library Group, jak i zaleceń Digital Preservation Coalition wynika, że po udokumentowaniu ilościowego i jakościowego stanu zasobów cyfrowych instytucje pamięci powinny przystąpić do definiowania metod, narzędzi

i technologii umożliwiających długoterminowe utrzymanie użyteczności „czytelnych" dokumentów cyfrowych oraz ewentualne odzyskanie treści dokumentów „nieczytelnych".

Z uwagi na szybko postępujący rozwój technologiczny i ciągłe zmiany technologii zapisu i odczytu dokumentów cyfrowych, instytucje pamięci stoją przed pilnym zadaniem opracowania strategii postępowania z materiałami cyfrowymi, która zagwarantuje zachowanie ich użyteczności, czyli autentyczności, integralności oraz poufności w długim czasie, co pozwoli obecnym oraz przyszłym użytkownikom odczytać i interpretować zapisy cyfrowe stanowiące dziedzictwo nauki i kultury.

Długoterminową archiwizację zasobów cyfrowych postrzega się jako pokaźne rozszerzenie spektrum dotychczasowej działalności instytucji pamięci, przyjęcie kolejnych poważnych obowiązków i realizację wielu nowych zadań. Środowiska odpowiedzialne za ochronę dziedzictwa kultury i nauki rozumieją potrzebę działań archiwizacyjnych i akceptują nowe zadania. Świadczą o tym podejmowane przez biblioteki, archiwa i muzea wielu krajów próby tworzenia programów długoterminowej archiwizacji narodowych zasobów cyfrowych oraz realizowane projekty, w których wyniku powstają prototypy systemów depozytowych zasobów cyfrowych oraz próby określenia, z jakimi organizacyjnymi, technicznymi, prawnymi i ekonomicznymi zadaniami powinny liczyć się instytucje przystępujące do archiwizacji dziedzictwa cyfrowego.

Z uwagi na przyjęte założenia dotyczące łączenia i udostępniania zasobów cyfrowych w skali świata, w planowanych i powstających programach ochrony narodowych kolekcji cyfrowych powinny zostać uwzględnione zalecenia wypracowane w ramach programów zaawansowanych. Wiedza i doświadczenia pochodzące z programów indywidualnych powinny być popularyzowane, dyskutowane, opiniowane na międzynarodowym forum i łączyć się w zunifikowaną światową politykę ochrony dziedzictwa cyfrowego. Należy szukać możliwości ujednolicenia opracowanych, wdrożonych i funkcjonujących programów ochrony oraz zaniechać koncepcji tworzenia strategii oryginalnych. W zamierzeniu polityka ochrony światowego dziedzictwa ma być efektem synergizmu działań wszelkich instytucji na świecie zainteresowanych trwałą archiwizacją i widocznością rodzimych kolekcji cyfrowych w zasobach świata. Systemy zunifikowane mają ułatwić wzajemne udostępnianie zasobów archiwalnych oraz wzajemne przechowywanie kopii kolekcji archiwalnych. Niektóre instytucje archiwizujące wypracowały bowiem koncepcję ochrony narodowych kolekcji, której ważnym punktem jest nawiązanie współpracy i decyzja o partnerstwie z instytucją archiwizującą innego kraju w celu wymiany i wzajemnej ochrony lustrzanych kopii narodowych zasobów archiwalnych. Zunifikowana strategia działań archiwizacyjnych oraz jednolite unormowane zasady organizacji i funkcjonowania narodowych archiwów

cyfrowych mają zwiększyć szanse długoterminowej dostępności i użyteczności cyfrowej pamięci świata. W wielu krajach trwają prace projektowe dotyczące utworzenia archiwum narodowego dziedzictwa cyfrowego. Za podstawę tych projektów przyjmuje się referencyjny model organizacji i funkcjonowania archiwów cyfrowych OAIS, któremu przyznano status normy ISO oraz katalogi kryteriów, według których mają być tworzone wiarygodne i certyfikowane archiwa cyfrowe.

Światowe instytucje pamięci organizując cyfrowe archiwa pracują nad możliwością zapewnienia oczekiwanej przez użytkowników jakości usług oraz uzyskania ich zaufania w kwestii zagwarantowania dostępności i użyteczności zbiorów cyfrowych w długim czasie. Ważne jest, aby przekonać deponentów i użytkowników o trwałości istnienia i sprawności funkcjonowania instytucji, która prowadzi archiwum oraz podawać do publicznej wiadomości informacje o celach i założeniach podejmowanych działań archiwizacyjnych oraz ich efektach. W ten sposób ma być budowane zaufanie do instytucji archiwalnych oraz tworzona możliwość konfrontacji założeń z osiągnięciami, następnie zaś kontrola i ocena sprawności oraz jakości działań archiwizacyjnych.

Długoterminowa archiwizacja zasobów cyfrowych jest procesem złożonym. Jego realizacja wymaga podjęcia wielu czynności natury organizacyjnej, prawnej, technicznej oraz ekonomicznej. W wyniku prowadzonych prac archiwizacyjnych instytucje archiwizujące przekonały się, że jest to proces, który one mają inicjować i organizować oraz któremu mają przewodniczyć, jednak przy założeniu współodpowiedzialności i współpracy wszystkich środowisk zainteresowanych przetrwaniem i użytecznością cyfrowej pamięci. Konieczne jest zaangażowanie organizacji rządowych, przekonanie ich o znaczeniu projektu i potrzebie jego finansowania. Równie istotne jest uświadomienie twórcom i środowiskom wydawniczym ich roli w procesach gromadzenia i archiwizacji wytworzonych materiałów cyfrowych. Instytucje archiwizujące opracowują i publikują wytyczne dotyczące parametrów dokumentów cyfrowych, od których uzależnione jest zachowanie ich autentyczności i integralności w długim czasie oraz starają się o ich uwzględnienie w procesach publikowania elektronicznego oraz digitalizacji.

Poważną trudność nastręcza instytucjom archiwizującym opracowanie kryteriów tworzenia kolekcji archiwalnych. Istnieje zgodność co do tego, że mało realne jest zachowanie całości cyfrowej podaży, dlatego też zasoby instytucji pamięci poddaje się procesom oceny i selekcji, tak aby do kolekcji archiwalnej włączyć materiały o szczególnym znaczeniu, będące świadectwem wybitnej naukowej i kulturowej działalności narodów, które należałoby wyeksponować w kolekcji dziedzictwa światowego. Oprócz oceny merytorycznej wartości treści dokumentów cyfrowych uwzględnia się parametry techniczne ich zapisu, tak aby do kolekcji archiwalnej trafiły zasoby rokujące powodzenie procesu zachowania

długoterminowej użyteczności. Dotychczas procesy tworzenia kolekcji archiwalnych są prowadzone na podstawie indywidualnych założeń instytucji archiwizujących poszczególnych krajów. Trwają wprawdzie prace nad ustaleniem choćby najogólniejszych procedur tworzenia narodowych kolekcji archiwalnych, ale w archiwistyce cyfrowej pozostaje to wciąż zagadnieniem otwartym.

Ponadto w działalności archiwizacyjnej instytucje pamięci poszukują doradców i współpracowników z branży prawniczej oraz informatycznej.

Gromadzenie, archiwizacja oraz udostępnianie zasobów archiwalnych powinno odbywać się na jasno zdefiniowanych zasadach. W związku z trendami gromadzenia i archiwizowania zasobów sieciowych, instytucje archiwizujące wchodzą w obszar działalności nieuregulowanej prawnie. Podkreśla się więc potrzebę przeglądu i uzupełnienia obowiązujących aktów prawnych związanych z gromadzeniem, archiwizacją i udostępnianiem wszelkich typów zasobów. Potrzebne jest również doprecyzowanie warunków przeprowadzania prac konserwatorskich na obiektach archiwalnych. Instytucje archiwizujące stawiają sobie wprawdzie za podstawowy cel przechowanie dokumentów autentycznych i integralnych, zgodnych z oryginałem, jednak w specyfikę archiwistyki cyfrowej są wpisane prace, które mogą wywoływać odstępstwa od pierwotnej formy i treści dokumentu. Tym samym instytucje archiwizujące wchodzą w konflikt z przepisami prawa, które nakazują chronić nienaruszalność treści i formy utworów. Wizerunek wiarygodności, na który starają się zapracować instytucje archiwizujące, wymaga działalności zgodnej z prawem, zatem potrzebne jest jednoznaczne uregulowanie kwestii prac archiwizacyjnych na obiektach cyfrowych. Instytucje archiwizujące realizują zadania prawne samodzielnie – na podstawie własnych zasobów kadrowych i ich wiedzy – bądź przy współudziale ekspertów zewnętrznych.

Bardzo ważnym aspektem działalności archiwizacyjnej jest techniczna ochrona archiwum cyfrowego i jego zasobów. W opracowywanych strategiach zakłada się potrzebę stałej ochrony archiwum cyfrowego z zastosowaniem rozmaitych form jego zabezpieczeń wypracowanych w branży informatycznej. Ochronie i odpowiednim zabiegom konserwatorskim podlegają też zgromadzone obiekty archiwalne. Typowe zabiegi konserwatorskie stosowane i rekomendowane w archiwistyce cyfrowej to odświeżanie oraz zmiana generacji nośnika, migracja do nowych formatów zapisu danych, a w przypadku potrzeby odtworzenia publikacji w oryginalnym środowisku emulacja otoczenia sprzętowo-programowego. Należy również zadbać, aby zasoby archiwalne były zapisane w standardowych otwartych formatach, wyczerpująco opisane w postaci metadanych oraz łatwe do znalezienia i trwale dostępne poprzez przypisane im identyfikatory unikalne. Ponadto zauważa się potrzebę nieustannego obserwowania rozwoju rynku technologicznego oraz właściwego reagowania na zachodzące w nim i na nim zmiany.

W pracach tych niezbędny jest udział wyszkolonych techników i informatyków, w związku z czym instytucje archiwizujące rozbudowują istniejące już w ich strukturze działy IT, szkolą zatrudnionych i nowo przyjętych pracowników do zadań archiwizacyjnych; nawiązują także współpracę z firmami informatycznymi i zlecają im wykonywanie specjalistycznych prac konserwatorskich.

Tylko instytucje pamięci w krajach bogatych realizują nowe, skomplikowane, kosztowne organizacyjnie i prawnie zadania oraz techniczne eksperymenty z zakresu archiwistyki cyfrowej. Ich prace przysługują się przyrostowi doświadczenia, wniosków i wiedzy, na podstawie których tworzone są instrukcje i rekomendacje dla działań archiwizacyjnych podejmowanych na całym świecie. Dają one również podstawy dalszych prac badawczych i wytyczają kierunki rozwoju w dziedzinie archiwistyki cyfrowej. Liderami są narodowe instytucje pamięci Ameryki, Australii, Anglii, Holandii i Niemiec. Wraz z instytucjami wielu innych krajów współtworzą światową platformę działalności archiwizacyjnej.

W Polsce problematyka długoterminowej archiwizacji zasobów cyfrowych jest wciąż mało rozpoznana. Polskie instytucje pamięci gromadzą dokumenty cyfrowe od połowy lat 90. XX stulecia. Zadania długoterminowej archiwizacji są sporadycznie uwzględniane w ich bieżącej działalności. Polscy wydawcy wykazują wprawdzie świadomość potrzeby specjalnej ochrony opublikowanych dokumentów elektronicznych, jednak ich działalność w tym obszarze wymaga aktywizacji.

Dużą zmianę w działalności polskich instytucji pamięci odnośnie do tworzenia i ochrony zasobów cyfrowych zapowiedziało Ministerstwo Kultury i Dziedzictwa Narodowego w opublikowanym we wrześniu 2009 r. *Programie digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020*. Z treści dokumentu wynika, że obecne dziesięciolecie ma być w instytucjach pamięci dedykowane przede wszystkim działaniom digitalizacyjnym i archiwizacyjnym.

Podobnie jak w przypadku zbiorów cyfrowych innych krajów, tak i dorobek polskiej działalności naukowej oraz kulturowej opublikowany w cyfrowej postaci niewątpliwie stanowi polskie i światowe dziedzictwo cyfrowe. Prawo do dostępu i korzystania z niego przysługuje zarówno obecnym, jak i przyszłym użytkownikom. O jego zgromadzenie, bezpieczne przechowywanie oraz prezentację w kraju i za granicą powinny zatroszczyć się instytucje pamięci odpowiedzialne za wieczystą archiwizację dziedzictwa narodowego.

Rolą MKiDN, jako organu sprawującego opiekę nad dziedzictwem narodowym, jest zainicjowanie i koordynowanie polskich działań archiwizacyjnych, jednak w warunkach pełnej współodpowiedzialności oraz współpracy wszystkich środowisk korzystających oraz odpowiedzialnych za ochronę dziedzictwa polskiej nauki i kultury.

Długoterminowa ochrona zasobów cyfrowych jest w instytucjach pamięci całego świata postrzegana jako zadanie złożone, skomplikowane i przede wszystkim niezwykle kosztowne. Biblioteki, archiwa, muzea oraz inne instytucje odpowiedzialne za przechowanie dziedzictwa narodowego powinny wyjść mu naprzeciw, aby wywiązać się ze swych ustawowych obowiązków. Trudność realizacji tego zadania polega głównie na tym, że pomimo wieloletniej dyskusji i aktywności wielu instytucji w tym zakresie, nadal bez odpowiedzi pozostaje pytanie o skuteczny sposób archiwizacji dokumentów cyfrowych, zapewniający ich przetrwanie i użytkowanie w dalekiej przyszłości. Jednak bez względu na to, jak mało precyzyjna i znana jest strategia długoterminowej archiwizacji, instytucje przechowujące dokumenty cyfrowe powinny to zadanie uznać i przystąpić do jego realizacji.

Bibliografia

4C (2013). *4C Collaboration to Clarify the Costs of Curation* [online]. 4C Project. Dostępny w WWW: http://www.4cproject.eu/about-us/ [Dostęp: 10.07.2017].

Adamczewski, Piotr (2005). *Słownik informatyczny. Terminologia informatyczna w pigułce*. Gliwice: Helion.

Adobe (b.d.) *Adobe Flash Player* [online]. Adobe Systems Software Ireland Ltd. Dostępny w WWW: http://get.adobe.com/pl/flashplayer/about/ [Dostęp: 10.07.2017].

Altenhöner, Reinhard; Klaproth, Frank; Stoll, Wilhelm; Ullrich, Dagmar (2008). *Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen – KOPAL* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://indi.langzeitarchivierung.de/allg/detail.php?show=657 [Dostęp: 10.07.2017].

Amse, Anne K. (2003). Zabezpieczanie historycznych zasobów dla przyszłości – archiwum cyfrowe holenderskiej biblioteki narodowej [online]. *Biuletyn EBIB*, nr 2(42). Dostępny w WWW: http://www.ebib.pl/2003/42/amse.php [Dostęp: 10.07.2017].

APARSEN (2013). *Alliance Permanent Access to the Records of Science in Europe Network* [online]. APARSEN. Dostępny w WWW: http://www.alliancepermanentaccess.org/index.php/about-aparsen/aparsen-deliverables/ [Dostęp: 10.07.2017].

Archiwistyka (2007). *Archiwistyka cyfrowa – zarys problematyki* [online]. Archiwistyka.pl. Dostępny w WWW: http://www.archiwistyka.pl/artykuly/informatyka_w_archiwum/Archiwistyka_cyfrowa_zarys_problematyki [Dostęp: 10.07.2017].

Arms, Caroline R. (2000). Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress [online]. *RLG DigiNews*, June 15, t. 4, nr 3. Dostępny w WWW: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file505.html#feature1 [Dostęp: 10.07.2017].

Attributes (2001). *Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources. An RLG-OCLC Report* [online]. The Research Libraries Group. Dostępny w WWW: http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2172.pdf [Dostęp: 10.07.2017].

Barczak, Andrzej; Sydoruk, Tadeusz (2002). *Bezpieczeństwo systemów informatycznych*. Siedlce: Wydawnictwo Akademii Podlaskiej.

Barta, Janusz; Markiewicz, Ryszard (2004). Wirtualne biblioteki a prawo autorskie. [W:] Kocójowa, Maria [red.]. *Przestrzeń informacji i komunikacji społecznej. Księga pamiątkowa ku czci prof. Wandy Pindlowej*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, s. 115-124.

Baza wiedzy (2016). *Różnice pomiędzy typami dysków twardych, macierzami RAID i ich kontrolerami w serwerach Dell PowerEdge i serwerach kasetowych* [online]. Dell. Dostępny w WWW: https://www.dell.com/support/article/pl/pl/pldhs1/sln129581/r%C3%B3%C5%BCnice-pomi%C4%99dzy-typami-dysk%C3%B3w-twardych--macierzami-raid-i-ich-kontrolerami-w-serwerach-dell-poweredge-i-serwerach-kasetowych?lang=pl [Dostęp: 10.10.2017].

Beagrie, Neil; Chruszcz, Julia; Lavoie, Brian (2008). *Keeping Research Data Safe. A cost model and guidance for UK universities* [online]. JISC. Dostępny w WWW: http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf [Dostęp: 10.07.2017].

Beagrie, Neil; Greenstein, Daniel (1998). A Strategic Policy Framework for Creating and Preserving Digital Collections. Version 4.0 (Final Draft). *ELib Supporting Study* P3. London: Library Information Technology Centre, South Bank University.

Bednarek, Grzegorz (2006). *PDF czy DjVu, w którą stronę?* [online]. GB Soft. Dostępny w WWW: http://www.djvu.com.pl/pdfanddjvu/DjVu_czy_pdf.php [Dostęp: 07.07.2017].

Bednarek-Michalska, Bożena (2006). Kujawsko Pomorska Biblioteka Cyfrowa a standardy [online]. *Biuletyn EBIB*, nr 4(74). Dostępny w WWW: http://www.ebib.pl/2006/74/michalska.php [Dostęp: 07.07.2017].

Berthon, Hilary; Howell, Alan (2000). *Preserving Access to Digital Information (PADI) – an Australian experience* [online]. PADI Preserving Access to Digital Information. National Library of Australia. Dostępny w WWW: http://www.alanhowell.com.au/wp-content/uploads/2012/10/BerthonHowell2000_PADI_Paper.pdf [Dostęp: 19.07.2017].

Beucke, Daniel (2010). *Geschaftsmodelle für die digitale Langzeitarchivierung. Das Beispiel Forschungsdaten*. Berlin: Institut für Bibliotheks- und Informationswissenschaft der Humboldt Universität zu Berlin, s. 26-37.

Bide, Mark (2000). *Standards for Electronic Publishing. An Overview*. The Hague: Koninklijke Bibliotheek.

Biliński, Lucjan (2004). *Selekcja materiałów bibliotecznych*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Bilski, Tomasz (2008). *Pamięć. Nośniki i systemy przechowywania danych*. Warszawa: Wydawnictwa Naukowo-Techniczne.

Bojar, Bożenna [oprac.] (2002). *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Bończa-Tomaszewski, Nikodem (2014). *Archiwistyka cyfrowa w Polsce. Sytuacja obecna i wizja przyszłości* [online]. Bartłomiej Kuczynski. Dostępny w WWW: http://prezi.com/2n7uiqeyczzx/archiwistyka-cyfrowa-w-polsce/ [Dostęp: 10.07.2017].

Borawski, Zdzisław (2007). Zasady ochrony dziedzictwa dźwiękowego według IASA. [W:] Woźniak-Kasperek, Jadwiga; Franke Jerzy [red.]. *Biblioteki cyfrowe. Projekty, realizacje, technologie. Praca zbiorowa*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Borghoff, Uwe M. (2005). *Vergleich bestehender Archivierungssysteme* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://files.d-nb.de/nestor/materialien/nestor_mat_03.pdf [Dostęp: 10.07.2017].

Borghoff, Uwe M.; Rödig, Peter; Scheffczyk, Jan; Schmitz, Lothar (2003). *Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente*. Heidelberg: dpunkt.verlag.

Börsenverein (1996). *Die unendliche Bibliothek. Digitale Information in Wissenschaft, Verlag und Bibliothek*. Börsenverein des deutschen Buchhandels. Wiesbaden: Harrassowitz.

Brauner, Detlev J.; Weigert, Martin M.; Raible-Besten, Robert [hrsg.] (1997). *Lexikon des Verlagswesens*. München: Oldenburg.

Brzeźniak, Maciej; Meyer, Norbert; Mikołajczak, Rafał; Stroiński, Maciej (2009). *Usługi przechowywania danych KMD/PLATON-U4 dla bibliotek cyfrowych* [online]. Repozytorium Zespołu Bibliotek Cyfrowych PCSS. Dostępny w WWW: http://lib.psnc.pl/Content/226/04-prezentacja_PLATON-U4_KMD_biblioteki%20%5Btryb%20zgodno%C5%9Bci%5D.pdf [Dostęp: 10.07.2017].

Buczyński, Ludwik (1999). *Komputerowe nośniki informacji – technologie zapisu*. Warszawa: Wydawnictwo Techniczne Grzegorz Safinowski.

Building (2013). *Building a Preservation Policy* [online]. The British Library Board. Dostępny w WWW: http://www.bl.uk/aboutus/stratpolprog/collectioncare/publications/booklets/building_a_preservation_policy.pdf [Dostęp: 10.07.2017].

CAMiLEON (b.d.). *Creative Archiving at Michigen & Leeds: Emulating the Old on the New* [online]. DCC. Dostępny w WWW: http://www.dcc.ac.uk/resources/external/camileon-creative-archiving-michigan-and-leeds-emulating-old-new [Dostęp: 10.07.2017].

CCSDS (b.d.). *CCSDS. Mission Operations and Information Management Services Area* [online]. CCSDS/ASRC Federal Technical Services. Dostępny w WWW: http://public.ccsds.org/publications/MOIMS.aspx [Dostęp: 10.07.2017].

CDLib (2008). *Archival Resource Key* [online]. California Digital Library. Dostępny w WWW: http://www.cdlib.org/inside/diglib/ark [Dostęp: 10.07.2017].

CEDARS (b.d.). *Curl exemplars in digital archives* [online]. UKOLN. Dostępny w WWW: http://www.ukoln.ac.uk/services/elib/projects/cedars/ [Dostęp: 10.07.2017].

Center for Research Libraries; OCLC (2007). *Trustworthy repositories audit & certification (TRAC) criteria and checklist* [online]. The Center for Research Libraries. Online Computer Library Center. Dostępny w WWW: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf [Dostęp: 13.07.2017].

Cieślak, Daniel (2002). Ranking nośników [online]. *PC Word*. Dostępny w WWW: http://www.pcworld.pl/news/33999/Ranking.nosnikow.html [Dostęp: 10.07.2017].

Clavel-Merrin, Genevieve (2000). *The NEDLIB List of Terms*. The Hague: Koninklijke Bibliotheek.

Content Packaging (b.d.) *Content Packaging Specification* [online]. IMS Global Learning Consortium. Dostępny w WWW: http://www.imsproject.org/content/packaging [Dostęp: 10.07.2017].

Coy, Wolfgang (2006). *Perspektiven der Langzeitarchivierung multimedialer Objekte* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://files.d-nb.de/nestor/materialien/nestor_mat_05.pdf [Dostęp: 10.07.2017].

CRL (2010). *Reports on Digital Archives and Repositories* [online]. The Center for Research Libraries. Dostępny w WWW: http://www.crl.edu [Dostęp: 11.07.2017].

Curation Costs Exchange (b.d.) [online]. 4C, Komisja Europejska. Dostępny w WWW: http://www.curationexchange.org/ [Dostęp: 10.07.2017].

Czajkowski, Michał (2002). *Wielka encyklopedia Internetu i nowych technologii*. Kraków: Edition.

Czapnik, Grzegorz (2009). *Książki internetowe w bibliotekach – dostęp komercyjny* [online]. Uniwersytet Śląski. Wydział Filologiczny. Dostępny w WWW: https://www.sbc.org.pl/dlibra/show-content/publication/edition/19449?id=19449 [Dostęp: 30.08.2017].

Czapnik, Grzegorz; Gruszka, Zbigniew [oprac.]; Tadeusiewicz, Hanna [współpr.] (2011). *Podręczny słownik bibliotekarza*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Czermiński, Jurand B. (2002). *Cyfrowe środowisko współczesnej biblioteki*. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego.

Czerni, Sergiusz; Skrzyńska, Maria [red.] (1986). *Słownik naukowo-techniczny angielsko-polski*. Warszawa: Wydawnictwa Naukowo-Techniczne.

Dąbrowska, Ewa (2017). Problem archiwizacji internetu w kontekście egzemplarza obowiązkowego: sytuacja w Polsce i wybranych krajach europejskich [online]. *Biuletyn EBIB*, nr 2(172). Dostępny w WWW: http://open.ebib.pl/ojs/index.php/ebib/article/view/523/682 [Dostęp: 7.07.2017].

Daszewski, Włodzimierz (2004). Konserwacja zbiorów i jej wpływ na zakres informacji naukowej. [W:] Daniłowicz, Czesław [red.]. *Multimedialne i sieciowe systemy informacyjne*. Vol. 1, Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, s. 269-278.

Daszewski, Włodzimierz (2006). Trwałość płyt optycznych CD i DVD. *Przegląd Informacyjno-Dokumentacyjny*, nr 2/293, s. 85-86.

Data Max (b.d.). *Uszkodzenia – typowe objawy* [online]. Data Max Recovery. Dostępny w WWW: http://www.odzyskiwanie-danych.net/uszkodzenia_objawy_typowe.php [Dostęp: 10.10.2017].

Day, Michael (2002). *Mapping between metadata formats* [online]. UKOLN. Dostępny w WWW: http://www.ukoln.ac.uk/metadata/interoperability [Dostęp: 10.07.2017].

Day, Michael (2003). ECDL-2003 Web Archiving [online]. *ARIADNE. Web Magazine for Information Professionals*. Dostępny w WWW: http://www.ariadne.ac.uk/issue37/ecdl-web-archiving-rpt [Dostęp: 10.07.2017].

DCC (2010). *Digital Curation Centre* [online]. DCC. Dostępny w WWW: http://www.dcc.ac.uk [Dostęp: 10.07.2017].

Deja, Marek (2015). Analiza zjawiska elektronicznego samopublikowania: model elektronicznego samopublikowania. *Nowa Biblioteka*, 1(16), s. 7-20 [online]. Repozytorium Uniwersytetu Jagiellońskiego. Dostępny w WWW: http://ruj.uj.edu.pl/xmlui/handle/item/27366 [Dostęp: 30.08.2017].

Derfert-Wolf, Lidia (2012). Archiwizacja Internetu – wprowadzenie i przegląd wybranych inicjatyw [online]. *Biuletyn EBIB*, nr 1(128). Dostępny w WWW: http://www.ebib.pl/images/stories/numery/128/128_derfert.pdf [Dostęp: 10.07.2017].

DFG (2014). *DFG-Projekt: LuKII (LOCKSS und KOPAL Infrastruktur und Interoperabilität) LuKII – LOCKSS und KOPAL Infrastruktur und Interoperabilität* [online]. Humboldt Universität zu Berlin. Dostępny w WWW: https://www.ibi.hu-berlin.de/de/forschung/digibib/forschung/projekte/LuKII/ [Dostęp: 10.07.2017].

Digi CULT Technology Challenges for Digital Culture (b.d.) [online]. Digi Cult Consortium. Dostępny w WWW: http://www.digicult.info/pages/info.php [Dostęp: 10.07.2017].

Digital Moving (2017). *Digital Moving-Picture Exchange (DPX), Version 2.0* [online]. Digital Formats. Dostępny w WWW: https://www.loc.gov/preservation/digital/formats/fdd/fdd000178.shtml [Dostęp: 10.07.2017].

Digital Preservation Europe (b.d.) [online]. National Archives of the Netherlands. Dostępny w WWW: http://en.nationaalarchief.nl/information-management-and-creation-of--archives/sustainable-management-of-digital-archiva-23 [Dostęp: 19.07.2017].

Digitale Erhaltungsstrategien (2008). [W:] Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 1.5*. [online]. Nestor: Göttingen. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf [Dostęp: 10.07.2017].

Dindorf, Marcus; Schrimpf, Sabine (2012). EU-Projekte zur digitalen Langzeitarchivierung. *Dialog mit Bibliotheken*, 24, z. 1, s. 33-35.

Dissonline.de (2009). *Digitale Dissertationen im Internet* [online]. Serwis Dissonline.de. Dostępny w WWW: http://www.dissonline.de [Dostęp: 10.07.2017].

DNB (b.d.). *Deutsche Nationalbibliothek* [online]. Deutsche Nationalbibliothek. Dostępny w WWW: http://www.d-nb.de/ [Dostęp: 10.07.2017].

DNBG (2009). *Gesetz über die Deutsche Nationalbibliothek (DNBG)* [online]. Bundesministerium der Justiz. Dostępny w WWW: http://bundesrecht.juris.de/dnbg/BJNR133800006.html [Dostęp: 10.07.2017].

Dobratz, Susanne; Tappenbeck, Inka (2002). Thesen zur Zukunft der digitalen Langzeitarchivierung in Deutschland [online]. *Bibliothek*, t. 26, nr 3, s. 257-261. Dostępny w WWW: https://www.degruyter.com/downloadpdf/j/bfup.2002.26.issue-3/bfup.2002.257/bfup.2002.257.pdf [Dostęp: 10.07.2017].

DOI System (2017). *The DOI System* [online]. International DOI Foundation. Dostępny w WWW: http://www.doi.org [Dostęp: 11.07.2017].

DPC (2009). *DPC History* [online]. Digital Preservation Coalition. Dostępny w WWW: http://www.dpconline.org/docman/miscellaneous/members/102-dpc-history [Dostęp: 18.07.2017].

DPC (2010a). *Digital Preservation Coalition* [online]. Digital Preservation Coalition. Dostępny w WWW: http://www.dpconline.org/ [Dostęp: 10.07.2017].

DPC (2010b). *DPC/PADI What's new in digital preservation* [online]. PADI, Preserving Access to Digital Information, National Library Of Australia. Dostępny w WWW: http://www.dpconline.org/blog/whats-new-issue-6 [Dostęp: 10.07.2017].

DPC (2016). *Comments on David Rosenthal's "The case for a revision of OAIS"* [online]. DPConline. Dostępny w WWW: http://wiki.dpconline.org/index.php?title=Comments_on_David_Rosenthal%27s_%E2%80%9CThe_case_for_a_revision_of_OAIS%E2%80%9D [Dostęp: 10.07.2017].

Drewniewska-Idziak, Barbara [red.] (2002). *Aktualne tendencje ochrony zbiorów bibliotecznych i archiwalnych: materiały z ogólnopolskich warsztatów konserwatorskich. Warszawa, 13-14 czerwca 2002*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Dubisz, Stanisław [red.] (2003). *Uniwersalny Słownik Języka Polskiego*. T. 2. Warszawa: Wydawnictwo Naukowe PWN.

Dublin Core (2017). *The Dublin Core Metadata Initiative* [online]. Dublin Core Metadata Initiative. Dostępny w WWW: http://dublincore.org [Dostęp: 11.07.2017].

Dudek, Paulina; Kowalska, Anna [red.] (2010). *Narodowe Archiwum Cyfrowe. Wizja, projekt, ludzie* [online]. Narodowe Archiwum Cyfrowe. Dostępny w WWW: http://www.nac.gov.pl/wp-content/uploads/2015/05/NAC_wizja_projekt_ludzie_WWW.pdf [Dostęp: 10.07.2017].

Duńczyk-Szulc, Anna (2012). *Plany MKiDN w zakresie digitalizacji. Materiał prezentowany podczas konferencji* Dziedzictwo w sieci – różne aspekty digitalizacji Kraków, 29.11.2012 r. [online]. Muzeum Fotografii w Krakowie. Dostępny w WWW: http://www.mhf.krakow.pl/files/attachments/20121213080947_Anna%20Dunczyk-Szulc_Strategia%20digitalizacji%20MKiDN.pdf [Dostęp: 10.07.2017].

e-Depot (b.d.). *e-Depot and digital preservation* [online]. Koninklijke Bibliotheek. National Library of the Netherlands. Dostępny w WWW: http://www.kb.nl/hrd/dd/index-en.html [Dostęp: 10.07. 2017].

ERPANET (2004). *Experts Workgroup on the Preservation of Digital Memory* [online]. Electronic Resource Preservation And Access Network, University of Glasgow. Dostępny w WWW: http://www.erpanet.org/events/workgroup/index.php [Dostęp: 10.07.2017].

Fakten+Zahlen 2007 (2008). [W:] Fischer Barbara [red.]. *Deutsche Nationalbibliothek Jahresbericht 2007*. [online]. Deutsche Nationalbibliothek. Dostępny w WWW: http://d-nb.info/995476365/34 [Dostęp: 10.07.2017].

Feather, John; Sturges, Paul [eds.] (2003). *International Encyclopedia of Information and Library Science*. London, New York: Routledge.

Feenstra, Bendert; IBM (2000). *Standards for the Implementation of a Deposit System for Electronic Publications*. The Hague: Koninklijke Bibliotheek.

Filas, Matylda; Wiorogórska, Zuzanna (2010). LOCKSS i Portico – projekty archiwizacji bibliotecznych zasobów elektronicznych. *Przegląd Biblioteczny*, z. 1, s. 33-43.

FITS (2008). *Definition of the Flexible Image Transport System (FITS)* [online]. FITS Working Group Commission 5: Documentation and Astronomical Data International Astronomical Union. Dostępny w WWW: https://fits.gsfc.nasa.gov/standard30/fits_standard30.pdf [Dostęp: 17.07.2017].

Formats (b.d.). *Sustainability of Digital Formats: Planning for Library of Congress Collecion* [online]. Digital Preservation, The Library of Congress. Dostępny w WWW: http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml [Dostęp: 10.07.2017].

Freedman, Alan (2004). *Encyklopedia komputerów*. Gliwice: Wydawnictwo Helion.

Fülle, Gunnar; Ott, Tobias (2006). Langzeiterhaltung digitaler Publikationen. Archivierung elektronischer Zeitschriften (E-Journals) [online]. Pagina GmbH. *Nestor Materialien*, nr 4. Dostępny w WWW: http://files.d-nb.de/nestor/materialien/nestor_mat_04.pdf [Dostęp: 10.07. 2017].

Funk, Stefan E. (2008a). Digitale Objekte und Formate. [W:] Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.5*. [online]. Nestor Göttingen. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf [Dostęp: 10.07.2017].

Funk, Stefan E. (2008b). Emulation. [W:] Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.5*. [online]. Nestor Gottingen. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf [Dostęp: 10.07.2017].

Funk, Stefan E. (2008c). Migration. [W:] Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.5*. [online]. Nestor Gottingen. Do-

stępnyw WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_344.pdf [Dostęp: 10.07.2017].
Giaretta, David (2011). *Advanced Digital Preservation*. Berlin, Heidelberg: Springer Verlag.
Goebel, Jürgen W.; Scheller, Jürgen; Zimmermann, Wolfgang (2004). *Digitale Langzeitarchivierung und Recht* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://files.d-nb.de/nestor/materialien/nestor_mat_01.pdf [Dostęp: 10.07.2017].
Goldenline (2009). *Archiwistyka cyfrowa?* [online]. Blog Kariera. GoldenLine. Dostępny w WWW: http://www.goldenline.pl/grupy/Uczelnie_studia_studenci_absolwenci/archiwistyka/archiwistyka-cyfrowa,1036174/ [Dostęp: 10.07.2017].
Gozdek, Jerzy (2013). *Dane na wieczność – oto najtrwalsze nośniki na świecie* [online]. Chip. Dostępny w WWW: http://www.chip.pl/2013/02/dane-na-wiecznosc-najtrwalsze-nosniki-na-swiecie/ [Dostęp: 10.10.2017].
Grossmann, Wendy (1997). *Leksykon: komputery, multimedia, Internet*. Warszawa: Wydawnictwo RTW.
Grygrowski, Dariusz (2001). *Dokumenty nieksiążkowe w bibliotece*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.
GUID (global unique identifier) (2005) [online]. TechTarget. Dostępny w WWW: http://searchwindowsserver.techtarget.com/definition/GUID-global-unique-identifier [Dostęp: 18.07.2017].
Handle System (2017). *The Handle System* [online]. Corporation for National Research Initiatives. Dostępny w WWW: http://www.handle.net [Dostęp: 11.07.2017].
HDD (2014). *Przyczyny awarii HDD* [online]. Serwis Laptopów. Dostępny w WWW: http://www.computerwelt24.com/przyczyny-awarii-hdd/ [Dostęp: 10.10.2017].
HD-Rosetta. *HD-Rosetta Archival Preservation Technologies and Services* [online]. Norsam Technologies, Inc. Dostępny w WWW: http://www.norsam.com/rosetta.html [Dostęp: 10.07.2017].
Hacker, Rupert (1992). *Bibliothekarisches Grundwissen*. 6. Aufl. München, London, New York: K.G. Saur.
Hägele, Günter (2000). Expertengespräch zur Langzeitarchivierung digitaler Publikationen. *Dialog mit Bibliotheken*, nr 12(1), s. 17-21.
Handle (b.d.). *Handle.Net® Information Services* [online]. Corporation for National Research Initiatives. Dostępny w WWW: https://www.handle.net/ [Dostęp: 10.07.2017].
Hänger, Andrea; Huth, Karsten; Wiesenmüller, Heidrun (2008). Auswahlkriterien. [W:] Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.5*. [online]. Nestor Gottingen. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf [Dostęp: 10.07.2017].
Hauffe, Heinz (1998). Langfristige Verfügbarkeit elektronischer Medien. [W:] Böllmann Elisabeth [red.]. *Speicherbibliotheken – Digitale Bibliotheken. Wissen Verteilen und Bewahren.* Frankfurt am Main: Vittorio Klostermann.
Hedstrom, Margaret; Montgomery, Sheon (1998). *Digital Preservation Needs and Requirements in RLG Member Institution* [online]. The Research Libraries Group. Dostępny w WWW: https://pdfs.semanticscholar.org/f87a/e50af20b33493edf37cc0145ac92f3702faf.pdf?_ga=2.50156896.77515404.1499784076-1706289594.1499784076 [Dostęp: 10.07.2017].

Hein, Stefan; Schmitt, Karlheinz; Werb, Virginie (2011). LuKII – LOCKSS und kopal Infrastruktur und Interoperabilität. *Dialog mit Bibliotheken* 23, s. 51-53.

Henze, Volker (1999). Langzeitarchivierung von Disketten. *Dialog mit Bibliotheken*, nr 11(3), s. 15-17.

Hiller, Helmut; Füssel, Stephan (2002). *Wörterbuch des Buches*. Frankfurt am Main: Vittorio Klostermann.

History: the KB (b.d.). *History: the KB and digital preservation* [online]. Koninklijke Bibliotheek. National Library of the Netherlands. Dostępny w WWW: https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/history-the-kb-and-digital-preservation (Dostęp: 10.10.2017).

Huth, Karsten (2009). Computermuseum. [W:] Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0*. [online]. Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_343.pdf [Dostęp: 10.07.2017].

IANA. Internet Assigned Numbers Authority (b.d.) [online]. The Internet Assigned Numbers Authority. Dostępny w WWW: http://www.iana.org [Dostęp: 10.07.2017].

ISBD (1999). *ISBD(ER): Intarnational Standard Bibliographic Description for Electronic Resources* [online]. IFLA, International Federation of Library Associations and Institutions. Dostępny w WWW: https://archive.ifla.org/VII/s13/pubs/isbd.htm [Dostęp: 10.07.2017].

ISBD(ER) (1997). *International Standard Bibliographic Description for Electronic Resources.* München: K. G. Saur.

ISO (2001). *Electronic still-picture imaging – Removable memory – Part 2: TIFF/EP image data format.* ISO 12234-2.

ISO (2004). *Graphic technology – Prepress digital data exchange – Tag image file format for image technology (TIFF/IT).* ISO 12639.

ISO (2005). *Electronic document file format for long-term preservation. Part 1.: Use of PDF 1.4.(PDF/A-1).* ISO/CD 19005-1:2005 [online]. International Organization for Standardization, Geneva. Dostępny w WWW: https://www.iso.org/standard/38920.html [Dostęp: 10.07.2017].

ISO (2006a). *About the MPEG-21 Standard (ISO/IEC 21000)* [online]. International DOI Foundation, Oxford. Dostępny w WWW: http://iso21000-6.net [Dostęp: 10.07.2017].

ISO (2006b). *Space data and information transfer systems – Producer-archive interface – Methodology abstract standard.* 20652:2006.

Jakubiec, Anna; Pazdur, Marzena (2013). Długoterminowa archiwizacja obiektów cyfrowych – międzynarodowe projekty. [W:] Januszko-Szakiel, Aneta [red.] *Wokół zagadnień trwałej ochrony zasobów cyfrowych*. Kraków: Oficyna Wydawnicza AFM, s. 45-58.

Janczewska-Sołomko, Katarzyna (2006). Digitalizacja zbiorów dźwiękowych w Europie Wschodniej i Środkowo-Wschodniej: zarys problematyki [online]. *Biuletyn EBIB*, nr 8(78). Dostępny w WWW: http://www.ebib.pl/2006/78/a.php?janczewska [Dostęp: 10.07.2017].

Janiak, Małgorzata (2012). Biblioteka cyfrowa, biblioteka elektroniczna, biblioteka wirtualna. [W:] Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria [red.]. *Biblioteki cyfrowe*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 15-65.

Januszewicz, Jerzy; Lewandowski, Tadeusz (2009). *Bezpieczeństwo systemów informatycznych*. Wałbrzych: Wałbrzyska Wyższa Szkoła Zarządzania i Przedsiębiorczości.

Januszko-Szakiel, Aneta (2003). Archiwizacja publikacji elektronicznych jako wyzwanie dla bibliotek – zarys problematyki. *Biuletyn Biblioteki Jagiellońskiej,* r. LIII, s. 215-225.

Januszko-Szakiel, Aneta (2005). Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych. *Przegląd Biblioteczny,* nr 3(73), s. 341-358.

Januszko-Szakiel, Aneta (2006). Dysertacje via Internet. Projekt elektronicznej archiwizacji rozpraw naukowych w Niemieckiej Bibliotece Narodowej. *Przegląd Biblioteczny*, nr 2, s. 141-152.

Januszko-Szakiel, Aneta (2008). Rola migracji i emulacji w strategii długoterminowej archiwizacji publikacji elektronicznych. [W:] Chmielowski, Wojciech Z.; Pękala, Maciej [red.]. *Informatyka*. Kraków: Krakowskie Towarzystwo Edukacyjne. Oficyna Wydawnicza AFM, s. 121-130.

Januszko-Szakiel, Aneta (2009a). *Kopal. System długoterminowej archiwizacji cyfrowego dziedzictwa nauki i kultury*. VI warsztaty „Biblioteki cyfrowe" Poznań 2009 [online]. Repozytorium Zespołu Bibliotek Cyfrowych PCSS. Dostępny w WWW: http://lib.psnc.pl/Content/225/03-KOPAL.pdf [Dostęp: 10.07.2017].

Januszko-Szakiel, Aneta (2009b). Wiarygodność archiwów cyfrowych. *Przegląd Biblioteczny*, nr 77(3), s. 325-347.

Januszko-Szakiel, Aneta (2010). Problemy organizacji działań w zakresie długoterminowej archiwizacji zasobów cyfrowych w Polsce. *Przegląd Biblioteczny*, nr 4, s. 405-428.

Januszko-Szakiel, Aneta (2011a). Analiza stanu zbiorów elektronicznych i warunków ich archiwizowania w polskich instytucjach bibliotecznych i wydawniczych. *Przegląd Biblioteczny*, nr 1, s. 21-46.

Januszko-Szakiel, Aneta (2011b). Długoterminowa archiwizacja zasobów cyfrowych. Program dla polskich bibliotek. *Przegląd Biblioteczny*, nr 2, s. 211-230.

Januszko-Szakiel, Aneta (2012). Archiwizacja elektronicznych zasobów bibliotecznych. Przegląd stosowanych metod ochrony. [W:] Januszko-Szakiel, Aneta [red.]. *Tradycja i nowoczesność w bibliotece naukowej XXI wieku*. Kraków: Oficyna Wydawnicza AFM, s. 131-149.

Januszko-Szakiel, Aneta (2013). Narodowy program długotrwałej archiwizacji cyfrowego zasobu nauki i kultury – propozycja dla Polski. [W:] Januszko-Szakiel, Aneta [red.]. *Wokół zagadnień trwałej ochrony zasobów cyfrowych*. Kraków: Oficyna Wydawnicza AFM, s. 173-199.

Januszko-Szakiel, Aneta; Kowalewski, Wojciech; Szafrański, Leszek (2016). Polskie biblioteki cyfrowe w kontekście kryteriów wiarygodności archiwów cyfrowych: próba ewaluacji. [W:] Cisek Sabina [red.]. *Inspiracje i innowacje: zarządzanie informacją w perspektywie bibliologii i informatologii*. Kraków: Biblioteka Jagiellońska, s. 189-224. [online]. Repozytorium Uniwersytetu Jagiellońskiego. Dostępny w WWW: https://ruj.uj.edu.pl/xmlui/handle/item/31976 [Dostęp: 10.10.2017].

Jehn, Mathias (2007). Herausforderung: Digitale Langzeitarchivierung in Europa [online]. *Dialog mit Bibliotheken*, nr 19(1), s. 10-11. Dostępny w WWW: http://d-nb.info/1118653351/34 [Dostęp: 10.07.2017].

Jehn, Mathias; Schrimpf, Sabine (2009). LZA-Aktivitäten in Deutschland aus dem Blickwinkel von nestor. [W:] Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen*

Langzeitarchivierung, Version 2.0. [online]. Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_336.pdf [Dostęp: 10.07.2017].

Kalota, Tomasz; Szala, Marcin (2012). Organizacja i logistyka digitalizacji. [W:] Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria [red.]. *Biblioteki cyfrowe*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 437-446.

Karta (2003). *Karta w sprawie ochrony dziedzictwa cyfrowego, Konferencja Generalna UNESCO, 32 sesja: 2003, rezolucja nr 42 z dnia 15 października 2003*. Naczelna Dyrekcja Archiwów Państwowych [tł. na jęz. pol. z jęz. ang.] [online]. Naczelna Dyrekcja Archiwów Państwowych. Dostępny w WWW: https://www.archiwa.gov.pl/pl/zarzadzanie-dokumentacja/dokument-elektroniczny/karta-unesco [Dostęp: 10.07.2017].

KEEP (2011). *KEEP – Keeping Emulation Environments Portable* [online]. Stanford University. Dostępny w WWW: http://library.stanford.edu/blogs/digital-library-blog/2011/12/keep-keeping-emulation-environments-portable [Dostęp: 10.07.2017].

KMD (b.d.). *KMD. Krajowy Magazyn Danych* [online]. Poznańskie Centrum Superkomputerowo-Sieciowe. Dostępny w WWW: https://hpc.man.poznan.pl/modules/project-section/item.php?itemid=3 [Dostęp: 10.07.2017].

Klein, Andy (2017). *Hard Drive Stats for Q2 2017* [online]. Backblaze. Dostępny w WWW: https://www.backblaze.com/blog/hard-drive-failure-stats-q2-2017/ [Dostęp: 10.10.2017].

Kocójowa, Maria (2004). Elektroniczna edukacja dla archiwistów. [W:] *Archiwistyka: Reforma kształcenia: Materiały Seminarium Rady Archiwalnej*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, s. 67-77.

Kolasa, Marek W. (2012). Formaty dokumentów w bibliotekach cyfrowych. [W:] Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria [red.]. *Biblioteki cyfrowe*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 403-422.

Konferencja (2008). *Konferencja Polskie Biblioteki Cyfrowe. Poznań, 24-25 listopada 2008 roku* [online]. Poznańskie Centrum Superkomputerowo-Sieciowe. Dostępny w WWW: http://www.man.poznan.pl/PBC/2008 [Dostęp: 10.07.2017].

Konopa, Bartłomiej (2017). Witryna internetowa – dokumentacja czy publikacja? [online]. *Biuletyn EBIB*, nr 2 (172). Dostępny w WWW: http://open.ebib.pl/ojs/index.php/ebib/article/view/528/684 [Dostęp: 7.07.2017].

Konstankiewicz, Marek (2005). Wykaz ważniejszych resortowych aktów prawnych regulujących zasady postępowania z dokumentacją. Uzupełnienie i kontynuacja (część XVIII). *Archiwista Polski*, nr 3(39), s. 49-62.

Konstankiewicz, Marek (2006). Wykaz ważniejszych aktów prawnych regulujących zasady postępowania z dokumentacją. Uzupełnienie i kontynuacja (część XVIII). *Archiwista Polski*, nr 2(42), s. 53-60.

Kopal (2007). *Kopal: Ein Service für die Langzeitarchivierung digitaler Informationen* [online]. Kopal Deutsche Nationalbibliothek. Dostępny w WWW: http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf [Dostęp: 10.07.2017].

Kopal (b.d.). *Kopal. Daten für die Zukunft* [online]. Kopal Deutsche Nationalbibliothek. Dostępny w WWW: http://kopal.langzeitarchivierung.de/index_ziel.php.de [Dostęp: 10.07.2017].

Kowalska, Małgorzata (2007). *Dygitalizacja zbiorów bibliotek polskich*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Kriterienkatalog (2006). *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 1: Entwurf zur öffentlichen Kommentierung* [online]. edoc Publication server. Humboldt -Universität zu Berlin. Dostępny w WWW: http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf [Dostęp: 10.07.2017].

Król, Witold (2004). Digitalizacja – problemy terminologiczne. [W:] Krzemińska, Wanda; Nowak, Piotr [red.]. *Cyfryzacja w procesach komunikowania*. Poznań: Sorus Wydawnictwo i Księgarnia, s. 25-33.

KRONIK@ (2017). *Krajowe Repozytorium Obiektów Nauki i Kultury*. [online]. Ministerstwo Cyfryzacji. Dostępny w WWW: https://mc.gov.pl/projekty/3903/opis-projektu [Dostęp: 12.07.2017].

LabKey (b.d.) *Life Science Identifier* [online]. LabKey Support. Dostępny w WWW: https://www.labkey.org/Documentation/wiki-page.view?name=lsidOverview [Dostęp: 18.07.2017].

Lamb, David; Prandoni, Claudio; Ricerche, Consorzio Pisa; Davidson, Joy (2009). *CASPAR* [online]. DCC. Dostępny w WWW: http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/caspar [Dostęp: 10.07.2017].

Lee, Stuart D. (2003). Czy dyskretyzacja jest tego warta? [W:] Rosowska, Ewa [red.]. *Archiwa w postaci cyfrowej. Materiały międzynarodowych warsztatów DELOS CEE.* Warszawa: Naczelna Dyrekcja Archiwów Państwowych, s. 117-121.

Leśniewski, Daniel (2002). *Digitalizacja zasobów bibliotecznych* [online]. Wielkopolska Biblioteka Cyfrowa. Dostępny w WWW: http://www.wbc.poznan.pl/dlibra/docmetadata?id=266&from=pubstats [Dostęp: 10.07.2017].

Lewandowska, Agnieszka; Mazurek, Cezary; Werla, Marcin (2007). Federacja Bibliotek Cyfrowych w sieci PIONIER. Dostęp do otwartych bibliotek cyfrowych i repozytoriów. [W:] IV Ogólnopolska Konferencja EBIB. Internet w bibliotekach Open Access. [online]. *EBIB Materiały konferencyjne* 2007, nr 18. Dostępny w WWW: http://www.ebib.pl/publikacje/matkonf/mat18/lewandowska_mazurek_werla.php [Dostęp: 10.07.2017].

Liegmann, Hans (2001). Langzeitverfügbarkeit digitaler Publikationen. [W:] Rützel-Banz, Margit [red.]. *Bibliotheken – Portale zum Globalen Wissen. 91. Deutscher Bibliothekartag in Bielefeld*. Frankfurt am Main: Vittorio Klostermann, s. 106-109.

LIFE (2010). *LIFE. Life Cycle Information for E-literature* [online]. LIFE, Life Cycle Information for E-literature, University College London, The British Library. Dostępny w WWW: http://www.life.ac.uk/ [Dostęp: 10.07.2017].

LoC (2003). *It's about time: research challenges in digital archiving and long term preservation: final report: Workshop on Research Challenges in Digital Archiving and Long-term Preservation*, April 12-13, 2002. Washington, D. C.: Library of Congress.

LoC (b.d. a). *About American Memory. Mission and History* [online]. Library of Congress. Dostępny w WWW: http://memory.loc.gov/ammem/about/index.html [Dostęp: 10.07.2017].

LoC (b.d. b). *National Digital Information Infrastructure and Preservation Program* [online]. The Library of Congress. Dostępny w WWW: https://www.loc.gov/loc/lcib/0601/ndiipp2.html [Dostęp: 18.07.2017].

LOCKSS Program (b.d.). Lots Of Copies Keep Stuff Safe [online]. Stanford University. Dostępny w WWW: https://www.lockss.org/ [Dostęp: 7.07.2017].

Long-term (b.d.). *Long-term usability of digital resources* [online]. Koninklijke Bibliotheek. National Library of the Netherlands. Nedlib Koninklijke Bibliotheek. Dostępny w WWW: https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources [Dostęp: 11.07.2017].

Lupovici, Catherine; Masanès, Julien (2000). *Metadata for Long Term Preservation* [online]. Nedlib Koninklijke Bibliotheek. Dostępny w WWW: https://www.kb.nl/sites/default/files/docs/preservationmetadata.pdf [Dostęp: 10.07.2017].

MARC Standards (2017) [online]. MARC Standards – Library of Congress Network Development and MARC Standards Office. Dostępny w WWW: http://www.loc.gov/marc [Dostęp: 10.07.2017].

Marshall, Dave (1999). *CCIR Standards for Digital Video* [online]. Cardiff School Of Computer Science. Dostępny w WWW: http://www.cs.cf.ac.uk/Dave/Multimedia/node197.html [Dostęp: 10.07.2017].

Martínez, José M. [red.] (2004). *MPEG-7 Overview* [online]. Moving Picture Experts Group (MPEG). Dostępny w WWW: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm [Dostęp: 10.07.2017].

Matlak, Andrzej (2007). *Charakter prawny regulacji dotyczących zabezpieczeń technicznych utworów*. Warszawa: Wolters Kluwer Polska.

Mazurek, Cezary; Parkoła, Tomasz; Werla, Marcin (2013). dArceo: Usługi długoterminowego przechowywania danych źródłowych. [W:] Januszko-Szakiel, Aneta [red.]. *Wokół zagadnień trwałej ochrony zasobów cyfrowych*. Kraków: Oficyna Wydawnicza AFM, s. 101-111.

Mendruń, Renek (2013). *Egzemplarz obowiązkowy a zasoby elektroniczne – kwiecień 2013* [online]. Polska Izba Książki. Dostępny w WWW: http://www.pik.org.pl/komunikaty/90/egzemplarz-obowiazkowy-a-zasoby-elektroniczne-kwiecien-2013 [Dostęp: 10.07.2017].

Metadata (2010) [online]. The Tech Terms Computer Dictionary. Dostępny w WWW: http://www.techterms.com/definition/metadata [Dostęp: 10.07.2017].

METS (b.d.). *METS. Metadata Encoding & Transmission Standards Schema, & Documentation* [online]. Metadata Encoding and Transmission Standard (METS), Library of Congress. Dostępny w WWW: http://www.loc.gov/standards/mets/mets-schema-docs.html [Dostęp: 10.07.2017].

Moats, Ryan (1997). *URN Syntax. Request for Comments: 2141.* AT&T [online]. The Internet Engineering Task Force. Dostępny w WWW: http://www.ietf.org/rfc/rfc2141.txt [Dostęp: 10.07.2017].

MPEG-21 Digital Item Declaration WD (v 2.0) (2001) [online]. CoverPages, OASIS. Dostępny w WWW: http://xml.coverpages.org/MPEG21-WG-11-N3971-200103.pdf [Dostęp: 10.07.2017].

Muir, Adrienne; Davies, Eric J. (2000). Legal deposit of digital material in the UK: recent developments and the international context. *Alexandria*, nr 12(3), s. 151-165.

Müller, Robert W. (1998). *Elektronisches Publizieren. Auswirkungen auf die Verlagspraxis.* Wisbaden: Harrasowitz Verlag.

Muszyński, Józef (2006). *Krótki żywot wypalanych nośników CD* [online]. Serwis Storage-standard.pl. Dostępny w WWW: http://www.computerworld.pl/news/Krotki-zywot-wypalanych-nosnikow-CD,87440.html [Dostęp: 10.07.2017].

NAC (b.d.). *Archiwum Dokumentów Elektronicznych* [online]. Narodowe Archiwum Cyfrowe. Dostępny w WWW: http://ade.ap.gov.pl/ [Dostęp: 10.07.2017].

Nahotko, Marek (2006). Biblioteki cyfrowych książek w środowisku akademickim i bibliotekarskim. [W:] *Cyfrowy świat bibliotek – problemy techniczne, prawne, wdrożeniowe*. Materiały konferencyjne z IX edycji seminarium z cyklu Archiwizowanie i digitalizacja, 17-18 stycznia 2006 r. Warszawa: Centrum Promocji Informacji, s. 177-204.

NARA (b.d.). *History of the Electronic Records and ERA* [online]. The National Archives and Records Administration. Dostępny w WWW: http://www.archives.gov/era/about/history.html [Dostęp: 10.07.2017].

Narodowe Archiwum Cyfrowe (b.d.) [online]. Narodowe Archiwum Cyfrowe. Dostępny w WWW: http://www.nac.gov.pl [Dostęp: 11.07.2017].

National Library of Australia (oprac.) (2003). *Ochrona dziedzictwa cyfrowego. Zalecenia*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych.

NEDLIB publications (b.d.). *The NEDLIB publications* [online]. Koninklijke Bibliotheek, National Library of the Netherlands. Dostępny w WWW: https://www.kb.nl/en/organisation/research-expertise/research-on-digitisation-and-digital-preservation/publications-on-digital-preservation-and-digitization/the-nedlib-publications [Dostęp: 11.07.2017].

NDAP (2015a). *Archiwum Dokumentów Elektronicznych. Publiczna prezentacja projektu*. [online]. Naczelna Dyrekcja Archiwów Państwowych. Dostępny w WWW: https://www.archiwa.gov.pl/files/prez.pdf [Dostęp: 29.08.2017].

NDAP (2015b). *Archiwum Dokumentów Elektronicznych w Programie Operacyjnym Polska Cyfrowa* [online]. Naczelna Dyrekcja Archiwów Państwowych. Dostępny w WWW: https://archiwa.gov.pl/pl/230-archiwum-dokument%C3%B3w-elektronicznych-w-programie-operacyjnym-polska-cyfrowa# [Dostęp: 29.08.2017].

NEDLIB (b.d.) *NEDLIB. Networked European Deposit Library* [online]. Nedlib Koninklijke Bibliotheek. Dostępny w WWW: http://www.dnb.de/EN/Wir/Projekte/Archiv/nedlib.html [Dostęp: 11.07.2017].

Nestor (2003). *Nestor: Kooperatives Kompetenznetzwerk Langzeitarchivierung*. Frankfurt am Main, Die Deutsche Bibliothek, 12-14 November 2003.

Nestor (2009). *Nestor – Das deutsche Kompetenznetzwerk zur digitalen Langzeitarchivierung* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://www.langzeitarchivierung.de [Dostęp: 11.07.2017].

Neuroth, Heike; Liegmann, Hans; Oßwald, Achim; Scheffel, Regine; Jehn, Mathias [hrsg.] (2008). *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 1.5* [online]. Nestor Göttingen. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf [Dostęp: 10.07.2017].

Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.] (2009). *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0* [online]. Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_20.pdf [Dostęp: 10.07.2017].

NLA (b.d.). *About PADI* [online]. National Library of Australia. Dostępny w WWW: http://www.nla.gov.au/padi/about.html [Dostęp: 10.07.2017].

Nowak, Adam (2007). Nośniki pamięci i ich rola w digitalizacji. [W:] Woźniak-Kasperek Jadwiga; Franke, Jerzy [red.]. *Biblioteki cyfrowe. Projekty, realizacje, technologie*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 186-197.

OAIS (1999). *Reference Model for an Open Archival Information System (OAIS). Draft Recommendation for Space Data System Standards. CCSDS 650.0-R-1. Red Book, z. 1.* Consultative Committee for Space Data System [online]. Nost. Dostępny w WWW: https://nssdc.gsfc.nasa.gov/nost/isoas/ref_model.html [Dostęp: 15.07.2017].

OAIS (2002). *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data Systems. CCSDS 650.0-B-1 Blue Book, z. 1.* Consultative Committee for Space Data System, Washington D. C. [online]. CCSDS.org – Publications. Dostępny w WWW: https://www.kb.nl/sites/default/files/docs/oaisbluebook.pdf [Dostęp: 15.07.2017].

OASIS (b.d.). *OASIS Extensible Resource Identifier (XRI) TC* [online]. OASIS. Dostępny w WWW: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xri [Dostęp: 10.07.2017].

OCLC (2004). *OCLC Digital Archive Preservation Policy and Supporting Documentation* [online]. OCLC Online Computer Library Center, Inc. Dostępny w WWW: http://wiki.lib.sun.ac.za/images/e/e5/Oclc-digital-preservation-policy.pdf [Dostęp: 10.07.2017].

OCLC (2009). *About OCLC* [online]. OCLC Online Computer Library Center Inc. Dostępny w WWW: http://www.oclc.org/about/default.htm [Dostęp: 10.07.2017].

Oświadczenie (2015). *Oświadczenie dotyczące obsługi formatu DjVu przez oprogramowanie dLibra* [online]. Poznańskie Centrum Superkomputerowo-Sieciowe. Dostępny w WWW: http://dingo.psnc.pl/2015/08/19/oswiadczenie-dotyczace-obslugi-formatu-djvu-przez-oprogramowanie-dlibra/ [Dostęp: 30.08.2017].

Oltmans, Erik (2004). Cost Models in Digital Archiving: An Overview of Life Cycle Management at the National Library of the Netherlands. *Liber Quarterly – The Journal of European Research Libraries*, t. 14, nr (3/4), s. 380-392.

Oltmans, Erik; Kol, Nanda (2005). A Comparison between Migration and Emulation in Terms of Costs [online]. *RLG DigiNews*, t. 9, nr 2. Dostępny w WWW: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file959.html [Dostęp: 11.07.2017].

Oltmans, Erik; Lemmen, Adrian (2006). The e-Depot at the National Library of the Netherlands [online]. *Serials* – 19(1). Dostępny w WWW: https://www.kb.nl/sites/default/files/docs/Serialsmarch2006.pdf [Dostęp: 15.07.2017].

ONIX (2014). *ONIX. Online Information eXchange* [online]. MIT Libraries. Dostępny w WWW: https://archive.li/i00Dq [Dostęp: 11.07.2017].

OpenXRI Project (2009) [online]. OpenXRI Foundation. Dostępny w WWW: http://www.openxri.org [Dostęp: 11.07.2017].

Opracowanie w sprawie (2017). *Opracowanie w sprawie stanu i potencjału polskich repozytoriów cyfrowych w instytucjach kultury i nauki.* Kraków: Biuro Informacji i Analiz Naukowych.

PADI (b.d.). *PADI. Preserving Access to Digital Information* [online]. National Library Of Australia. Dostępny w WWW: http://www.nla.gov.au/padi/index.html [Dostęp: 10.07.2017].

PAIMAS (2003). *Producer-Archive Interface: Methodology Abstract Model (PAIMAS). Recommendation for Space Data Systems. CCSDS 651.0-R-1. Red Book.* Consultative Committee for Space Data System, System, St. Hubert, Canada [online]. CCSDS.org – Publications. Dostępny w WWW: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.202.307&rep=rep1&type=pdf [Dostęp: 15.07.2017].

Palm, Jonas (2011). *Cyfrowa czarna dziura Komentarz A. D. 2011* [online]. Narodowy Instytut Audiowizualny. Dostępny w WWW: http://nina.gov.pl/media/43930/jonas-palm--cyfrowa-czarna-dziura.pdf [Dostęp: 30.08.2017].

Paradowski, Dariusz [oprac. i red.] (2010). *Digitalizacja piśmiennictwa* [online]. Biblioteka Narodowa. Dostępny w WWW: http://www.bn.org.pl/download/document/1342175805.pdf [Dostęp: 30.08.2017].

Parastatidis, Savas; Watson, Paul; Webber, Jim (2003). *Grid-Resource Specification* [online]. North East Regional e-Science Centre, School of Computing Science, University of Newcastle. Dostępny w WWW: https://pdfs.semanticscholar.org/9d62/6ffc2db243e1a7389b6249bf95d7fbf857c0.pdf [Dostęp: 11.07.2017].

PARSE.Insight (2011). *About PARSE.Insight: Permanent Access to the Records of science in Europe* [online]. Deutsche Nationalbibliothek. Dostępny w WWW: http://www.dnb.de/EN/Wir/Projekte/Archiv/parseInsight.html [Dostęp: 10.17.2017].

Pawska, Beata; Szymorowska, Teresa E. (2001). Projekt linii technologicznej digitalizacji dokumentów bibliotecznych – projekt menedżerski [online]. *Biuletyn EBIB*, nr 3(21). Dostępny w WWW: http://www.ebib.pl/biuletyn-ebib/21/a.php?pawska_szymorowska [Dostęp: 11.07.2017].

Perdereau, Jacques (2012). *Durability of recordable DVD±R and DVD made of glass (Syylex) at elevated temperature and humidity. Investigation on the preservation of digital archives in the framework of GISDON* [online]. LNE. Dostępny w WWW: https://documents.lne.fr/publications/guides-documents-techniques/syylex-glass-dvd-accelerated-aging-report.pdf [Dostęp: 10.07.2017].

Pericles (b.d.) [online]. Pericles. Dostępny w WWW: http://pericles-project.eu/main [Dostęp: 10.07.2017].

Persistent Identifier (2008). *Persistent Identifier: Eindeutige Bezeichner für digitale Inhalte* [online]. Deutsche Nationalbibliothek. Dostępny w WWW: http://www.persistent--identifier.de/?link=204 [Dostęp: 10.07.2017].

Pest, Czesław (2007). Zastosowanie programu TABULARIUM do kompleksowej obsługi archiwum i biblioteki. *Archiwista Polski*, nr 1(45), s. 15-23.

Phillips, Margaret E. (2005). What should we preserve? The question for heritage libraries in a digital world. *Library trends*, t. 54, nr 1, s. 57-71.

Pindlowa, Wanda (2004). Podstawy Informacji Naukowej w programach kształcenia wyższego dla archiwistów. [W:] *Archiwistyka: Reforma kształcenia. Materiały Seminarium Rady Archiwalnej*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, s. 63-66.

Piotrowicz, Grażyna (2005). Nowy wymiar funkcjonowania bibliotek jako instytucji kultury w społeczeństwie informacyjnym. [W:] Kocójowa, Maria [red.]. *Informacja o obiektach kultury i Internet*. Kraków: IINiB UJ, s. 45-52. [online]. Skryba. Dostępny w WWW: http://skryba.inib.uj.edu.pl/wydawnictwa/e01/piotrowicz.pdf [Dostęp: 11.07.2017].

Pisarek, Walery [red.] (2006). *Słownik terminologii medialnej*. Kraków: Universitas.

PLANETS (2007). *Preservation and long-term access through networked services* [online]. Planets. Dostępny w WWW: http://www.planets-project.eu/about/ [Dostęp: 10.07.2017].

Plezia, Marian [red.] (2007). *Słownik łacińsko-polski*. T. IV. Warszawa: Wydawnictwo Naukowe PWN.

Płaza, Jakub (2016). *Szklany dysk – nowy format zapisu danych* [online]. PC Format. Dostępny w WWW: https://www.pcformat.pl/News-Szklany-dysk--nowy-format-zapisu--danych,n,15183 [Dostęp: 10.07.2017].

Płoszajski, Grzegorz (2009). Standardy w procesie digitalizacji dziedzictwa kulturowego na świecie. Wymagania techniczne, metadane, zalecenia. [W:] *Prawo, standardy, technologia i procedury*. XII seminarium z cyklu Składowanie i archiwizacja. Warszawa: Centrum Promocji Informatyki, 20 maja 2009, s. 83-117.

PN (2000). *Opis bibliograficzny – Dokumenty elektroniczne*. PN-N-01152-13:2000.

Preserving (1996). *Task Force on Archiving of Digital Information; Research Libraries Group; Commission on Preservation and Access. Preserving digital information. Report of the Task Force on Archiving of Digital Information, commissioned by the Commission on Preservation and Access and the Research Libraries Group* [online]. CLIR. Dostępny w WWW: http://www.clir.org/pubs/reports/pub63watersgarrett.pdf [Dostęp: 11.07.2017].

Preservation Policy (2009) [online]. National Library of Australia. Dostępny w WWW: http://www.nla.gov.au/policy-and-planning/preservation-policy [Dostęp: 10.10.2017].

Program digitalizacji (2009). *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020*. Warszawa: Ministerstwo Kultury i Dziedzictwa Narodowego. [online]. Kongres Kultury. Dostępny w WWW: http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf [Dostęp: 18.07.2017].

Prytherch, Ray [oprac.] (1996). *Harrod's Librarians' Glossary. 9000 terms used in information management, library science, publishing, the book trades and archive management*. Aldershot: Gower, Brookfield: Ashgate Pub. Co.

Przyłuska, Jolanta (2008). Repozytorium – magazyn dokumentów czy wirtualna społeczność? [preprint] [online]. E-LiS. Dostępny w WWW: http://eprints.rclis.org/11855/ [Dostęp: 11.07.2017].

Radwański, Aleksander (2005). Normalizacja informatycznych systemów archiwalnych. [W:] *Nowe technologie archiwizacji. Digitalizacja archiwów i bibliotek. VIII seminarium z zakresu składowania i archiwizacji, 19-20 maja 2005 r., Wierzba, Dom Pracy Twórczej PAN*. Warszawa: Centrum Promocji Informatyki, s. 104.

Ras, Marcel (2009). The KB e-Depot: Building and Managing a Safe Place for e-Journals [online]. *LIBER Quarterly*. 19(1), s.44-53. Dostępny w WWW: http://doi.org/10.18352/lq.7951 [Dostęp: 15.07.2017].

Ravenwood, Clare; Muir, Adrienne; Matthews, Graham (2015). Stakeholders in the selection of digital material for preservation: relationships, responsibilities, and influence [online]. *Collection Management*, 40 (2), s. 83-110. Dostępny w WWW: https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/18299/3/Lupin%20Stakeholders%20in%20the%20Selection%20of%20Digital%20Material%20for%20Preservation.pdf [Dostęp: 19.07.2017].

Recommended Standards CCSDS (b.d.) [online]. CCSDS Consultative Committee for Space Data Systems. Dostępny w WWW: http://public.ccsds.org/publications/MOIMS.aspx [Dostęp: 11.07.2017]

Reitz, Joan M. (2004). *Dictionary for Library and Information Science*. Westport, London: Libraries Unlimited.

Research Libraries Group; OCLC (2002). *Trusted Digital Repositories. Attributes and Responsibilities. An RLG-OCLC Report* [online]. The Research Library Group. Dostępny

w WWW: http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf [Dostęp: 13.07.2017].
RLG (2005). *An Audit Checklist for the Certification of Trusted Digital Repositories. Draft for public comment* [online]. Dostępny w WWW: http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2416.pdf [Dostęp: 10.07.2017].
RLG (b.d.). *RLG Partnership Council* [online]. OCLC Online Computer Library Center. Dostępny w WWW: http://www.oclc.org/research/partnership/council.html [Dostęp: 11.07.2017].
Rohde-Enslin, Stefan (2004). *Nicht von Dauer. Kleiner Ratgeber für die Bewahrung digitaler Daten in Museen* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: https://d-nb.info/97435970x/34 [Dostęp: 11.07.2017].
Rosowska, Ewa [red.] (2003). *Archiwa w postaci cyfrowej. Materiały międzynarodowych warsztatów DELOS CEE*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych.
Ross, Seamus (2003). *Przesiadka w Wigan. Dygitalne zabezpieczanie i konserwacja zbiorów a przyszłość nauki*. Toruń: Międzynarodowe Centrum Zarządzania Informacją.
Ross, Seamus (2004). The Role of ERPANET in Supporting Digital Curation and Preservation in Europe [online]. *D-Lib Magazine*, 10(7/8). Dostępny w WWW: http://www.dlib.org/dlib/july04/ross/07 r.ss.html [Dostęp: 19.07.2017].
Rothenberg, Jeff (1995). Ensuring the Longevity of Digital Documents. *Scientific American*, 272(1), s. 42-47.
Rothenberg, Jeff (2000). *An Experiment in Using Emulation to Preserve Digital Publications.* Den Haag: The Koniklijke Bibliotheek.
Sałaciński, Krzysztof [red.] (2001). *Ochrona narodowego zasobu bibliotecznego. Materiały i dokumenty ze szkolenia dyrektorów bibliotek, których zbiory w całości lub w części tworzą Narodowy Zasób Biblioteczny, Kraków, kwiecień 2001 r*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.
Sapa, Remigiusz (2013). Realizacja funkcji repozytoryjnych przez największe przedsięwzięcia zarejestrowane w Federacji Bibliotek Cyfrowych tworzone i współtworzone przez uczelnie. *Przegląd Biblioteczny*, 2(81), s. 117-131.
Sasin, Wiesław (2004). *Przechowywanie i archiwizowanie dokumentacji przedsiębiorstwa według nowych zasad normatywnych. Poradnik dla wszystkich firm z instrukcją wzorcową, ze szczególnym uwzględnieniem dokumentacji: założycielskiej i organizacyjnej, osobowej i płacowej, inwestycyjnej i gospodarczej, księgowej i podatkowej*. Skierniewice: Sigma.
Schanze, Helmut [hrsg.] (2002). *Metzler Lexikon Medientheorie, Medienwissenschaft. Ansätze, Personen, Grundbegriffe*. Stuttgart, Weimar: Verlag J. B. Metzler.
Schneider, Hans J. [hrsg.] (1997). *Lexikon Informatik und Datenverarbeitung*. München, Wien: R. Oldenbourg Verlag.
Schöning-Walter, Christa (2008). Persistent Identifier für Netzpublikationen. *Dialog mit Bibliotheken,* nr 1, s. 32-38.
Schöning-Walter, Christa (2009). Der Uniform Resource Name (URN). [W:] Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0.* [online]. Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_336.pdf [Dostęp: 10.07.2017].
Schröder, Kathrin (2009). Persistent Identifier (PI) – ein Überblick. [W:] Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.]. *Nestor*

Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0. [online]. Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_337.pdf [Dostęp: 10.07.2017].

Shenton, Helen (2003). Life Cycle Collection Management. *Liber Quarterly – The Journal of European Research Libraries*, 13(3/4), s. 254-272.

Siedlarz, Bartłomiej (2012). DjVu [W:]. Janiak, Małgorzata; Krakowska, Monika; Próchnicka, Maria [red.]. *Biblioteki cyfrowe*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 423-436.

Siewicz, Krzysztof [oprac.]; Szczepańska, Barbara [współpr.] (2013). *Stanowisko organizacji bibliotekarskich w sprawie konsultacji społecznych dotyczących dozwolonego użytku publicznego* [online]. Centrum Cyfrowe. Dostępny w WWW: https://centrumcyfrowe.pl/blog/2013/07/02/stanowisko-organizacji-bibliotekarskich-w-sprawie-konsultacji--dotyczacych-dozwolonego-uzytku-publicznego/ [Dostęp: 10.10.2017].

Ślaska, Katarzyna (2009). *Digitalizacja i archiwizacja dokumentów elektronicznych w Bibliotece Narodowej.* [online]. Ośrodek Wydawnictw Naukowych (IChB PAN). Dostępny w WWW: http://lib.psnc.pl/dlibra/doccontent?id=166 [Dostęp: 11.07.2017].

Ślaska, Katarzyna; Wasilewska, Anna (2012). Archiwizacja Internetu – sytuacja w polskim prawie z punktu widzenia bibliotekarzy [online]. *Biuletyn EBIB*, 1(128). Dostępny w WWW: http://www.ebib.pl/images/stories/numery/128/128_slaska.pdf [Dostęp: 10.07.2017].

Śliwińska, Maria [red.] (2003). *Raport z Lund – koordynacja w zakresie dygitalizacji*. Kraków: Wydawnictwo Kontekst.

SMIL (2008). *Synchronized Multimedia Integration Language (SMIL 3.0)* [online]. W3C®. Dostępny w WWW: https://www.w3.org/TR/REC-smil/ [Dostęp: 10.10.2017].

Smith, Abby (2003). Dlaczego przekształcać na postać cyfrową? [W:] Rosowska, Ewa [red.]. *Archiwa w postaci cyfrowej: Materiały międzynarodowych warsztatów DELOS CEE*. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, s. 104-116.

SMPTE (2015). SMPTE. Standards Quarterly Report. March 2015 [online]. *SMPTE Standards Quarterly Report*. The Society of Motion Picture and Television Engineers®. Dostępny w WWW: https://www.smpte.org/sites/default/files/images/Standards%20Quarterly%20Outcome%20Report-March%202015.pdf [Dostęp: 17.07.2017].

Sollins, Karen; Masinter, Larry (1994). *Functional Requirements for Uniform Resource Names* [online]. Internet Engineering Task Force. Dostępny w WWW: http://www.ietf.org/rfc/rfc1737.txt [Dostęp: 10.07.2017].

Sompel van de, Herbert; Hammond, Tony; Neylon, Eamonn; Weibel, Stuart L. (2006). *The* info *URI Scheme for Information Assets with Identifiers in Public* [online]. The Internet Engineering Task Force. Dostępny w WWW: http://www.ietf.org/rfc/rfc4452.txt [Dostęp: 13.07.2017].

Sprawozdanie (2016). *Sprawozdanie Biblioteki Narodowej za rok 2016* [online]. Biblioteka Narodowa, s. 81-82. Dostępny w WWW: http://www.bn.org.pl/download/document/1494330300.pdf [Dostęp: 10.07.2017].

Stachowska-Musiał, Ewa [red.] (2006). *Dziedzictwo kulturowe. Zbiory biblioteczne i nowe technologie ich ochrony.* Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich.

Stanisławska-Kloc, Sybilla (2005). Prawo autorskie a biblioteka cyfrowa – opinia prawna [online]. *Biuletyn EBIB*, nr 9(70). Dostępny w WWW: http://www.ebib.pl/2005/70/stanislawska-kloc.php [Dostęp: 11.07.2017].

Steenbakkers, Johan (1999). Developing the Depository of Netherlands Electronic Publications. *Alexandria,* nr 11(2), s. 93-105.

Steenbakkers, Johan (2000). *Setting up a Deposit System for Electronic Publications. The NEDLIB Guidelines.* The Hague: Koninklijke Bibliotheek.

Steinke, Tobias (2007). *Technik der kopal-Lösung. Deutsche Nationalbibliothek* [online]. Kopal Deutsche Nationalbibliothek. Dostępny w WWW: http://kopal.langzeitarchivierung.de/downloads/kopal-goes-live_Technik_der_kopal_Loesung_Steinke.pdf [Dostęp: 11.07.2017].

Strauch, Dietmar; Rehm, Margarete (2007). *Lexikon Buch, Bibliothek, Neue Medien*. München: K. G. Saur.

Suchodoletz, von Dirk (2008). *Funktionale Langzeitarchivierung digitaler Objekte. Erfolgsbedingungen des Einsatzes von Emulationsstrategien* [online]. Nestor Deutsche Nationalbibliothek. Dostępny w WWW: http://d-nb.info/1047826933/34 [Dostęp: 10.07.2017].

Sun, Sam; Lannom, Larry; Boesch, Brian (2003). *Handle System Overview* [online]. The Internet Engineering Task Force. Dostępny w WWW: http://www.ietf.org/rfc/rfc3650.txt [Dostęp: 11.07.2017].

Supruniuk, Monika (b.d.). *Archiwum cyfrowe – ochrona czy dostęp?* [online]. Narodowy Instytut Audiowizualny. Dostępny w WWW: http://www.nina.gov.pl/baza-wiedzy/archiwum-cyfrowe-ochrona-czy-dost%C4%99p-monika-supruniuk/ [Dostęp: 11.07.2017].

Szaniawski, Jacek (1997). *Duży słownik informatyczny angielsko-polski*. Warszawa: ArsKom.

Szczepańska, Barbara (2007). Prawo autorskie – ochrona dzieł elektronicznych. [W:] Woźniak-Kasperek, Jadwiga; Franke, Jerzy [red.]. *Biblioteki cyfrowe. Projekty, realizacje, technologie. Praca zbiorowa*. Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich, s. 51-66.

Task Force (b.d.). *Task Force on Digital Repository Certification* [online]. The Research Library Group. Dostępny w WWW: http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2588.html [Dostęp: 11.07.2017].

Technology Reports (b.d.). *Technology Reports. Data Sharing, Mediation, and Synchrization* [online]. CoverPages. Dostępny w WWW: http://xml.coverpages.org/dataSharing.html [Dostęp: 11.07.2017].

TEI (b.d.). *TEI: Text Encoding Initiative* (2016) [online]. TEIConsortium. Dostępny w WWW: http://www.tei-c.org/index.xml [Dostęp: 17.07.2017].

CLOCKSS Archive (2017). *The CLOCKSS Archive. A Trusted Community-Governed Archive* [online]. CLOCKSS. Dostępny w WWW: https://clockss.org/clockss/Home [Dostęp: 10.07.2017].

Twigge, Stephen (2003). Ocena dokumentów elektronicznych. [W:] Rosowska Ewa [red.]. *Archiwa w postaci cyfrowej: Materiały międzynarodowych warsztatów DELOS CEE.* Warszawa: Naczelna Dyrekcja Archiwów Państwowych, s. 132-139.

Ullrich, Dagmar (2009). Digitale Speichermedien. [W:] Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Mathias [hrsg.]. *Nestor Handbuch. Eine*

kleine Enzyklopädie der digitalen Langzeitarchivierung, Version 2.0. [online] Boizenburg: Verlag Werner Hülsbusch. Dostępny w WWW: http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_331.pdf [Dostęp: 13.07.2017].

Uniform Resource Names. A Progress Report (1996) [online]. *D-Lib Magazine*, 2(2). Dostępny w WWW: http://www.dlib.org/dlib/february96/02arms.html [Dostęp: 13.07.2017].

Ustawa (1983). Ustawa z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach. *Dziennik Ustaw*, nr 38, poz. 173.

Ustawa (1994). Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych. *Dziennik Ustaw,* nr 24, poz. 83.

Ustawa (1996). Ustawa z dnia 7 listopada 1996 r. o obowiązkowych egzemplarzach bibliotecznych. *Dziennik Ustaw,* nr 152, poz. 722.

Ustawa (1997). Ustawa z dnia 27 czerwca 1997 r. o bibliotekach. *Dziennik Ustaw,* nr 85, poz. 539.

Ustawa (2001). Ustawa z dnia 27 lipca 2001 r. o ochronie baz danych. *Dziennik Ustaw,* nr 128, poz. 1402.

Ustawa (2004). Ustawa z dnia 1 kwietnia 2004 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych. *Dziennik Ustaw*, nr 91, poz. 869.

Ustawa (2005). Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne. *Dziennik Ustaw*, nr 64, poz. 565.

VRML (1995). *VRML Virtual Reality Modeling Language* [online]. W3C. Dostępny w WWW: http://www.w3.org/MarkUp/VRML [Dostęp: 20.10.2017].

Werf, Titia van der (2000). *The Deposit System for Electronic Publication. A Process Model*. The Hague: Koninklijke Bibliotheek.

Werf-Davelaar, Titia van der (1999). Long-term Preservation of Electronic Publications. The NEDLIB Project [online]. *D-Lib Magazine* September, 5(9). Dostępny w WWW: http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html [Dostęp: 13.07.2017].

White Papers (2017) [online]. Ultrium LTO. Dostępny w WWW: https://www.lto.org/resources/white-papers/ [Dostęp: 10.10.2017].

Wojciechowski, Jacek (2006). *Biblioteczna wartość naddana*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.

Woodcock, Joanne [red.] (2002). *Microsoft Encyklopedia Komputerowa*. Warszawa: Mikom.

Workshop „Digital Library Konzepte" (2003). Die Deutsche Bibliothek, 6 Oktober 2003.

X3D (2017). *X3D Resources* [online]. Web3D CONSORTIUM. Dostępny w WWW: http://www.web3d.org/x3d/content/examples/X3dResources.html#Examples [Dostęp: 10.10.2017].

Zalecenie (2011). *Zalecenie Komisji z 27.10.2011 w sprawie digitalizacji i udostępniania w Internecie dorobku kulturowego oraz w sprawie ochrony zasobów cyfrowych* [online]. EUR-Lex. Dostępny w WWW: http://eur-lex.europa.eu/legal-content/PL/TXT/?uri=CELEX%3A32011H0711 [Dostęp: 10.07.2017].

Zhang, Michael (2016). *This Glass Disc Can Store 360 TB of Your Photos for 13.8 Billion Years* [online]. PetaPixel. Dostępny w WWW: https://petapixel.com/2016/02/16/glass-disc-can-store-360-tb-photos-13-8-billion-years/ [Dostęp: 10.07.2017].