

Długoterminowa archiwizacja zasobów cyfrowych w świadomości pracowników polskich bibliotek

ANETA JANUSZKO-SZAKIEL

Krakowska Akademia im. Andrzeja Frycza Modrzewskiego

ajanuszeko-szakiel@afm.edu.pl

Streszczenie

W artykule zdefiniowano podstawowe pojęcia z zakresu długoterminowej archiwizacji publikacji elektronicznych oraz przybliżono podstawowe metody ochrony zasobów stosowane w archiwistyce cyfrowej. Na podstawie prowadzonych badań scharakteryzowano cyfrowe zasoby polskich bibliotek w kontekście zagadnień archiwizacyjnych oraz opisano rozwój świadomości pracowników polskich bibliotek odnośnie do potrzeby długoterminowej ochrony polskich zasobów cyfrowych. W tekście odniesiono się również do zagadnień inicjowania, koordynowania działań archiwizacyjnych w Polsce oraz współpracy na rzecz opracowania strategii długoterminowej archiwizacji polskich zasobów cyfrowych.

Słowa kluczowe: zasoby cyfrowe, długoterminowa archiwizacja zasobów cyfrowych, archiwa cyfrowe, archiwistyka cyfrowa, metody archiwizacji

Wstęp

Nietrwałość zapisów cyfrowych została niejednokrotnie stwierdzona i poparta dobitnymi przykładami. Podstawowe zagrożenia dla materiałów cyfrowych wynikają na przykład z niskiej trwałości, rozpadu, zniszczenia nośnika, zesterzenia się formatu zapisu. Zagrożenie stanowią także 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 aniżeli zmiany technologiczne, jest brak świadomości odnośnie potrzeby ich obserwacji i stosownej reakcji na nie. Dopiero utrata bardzo 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.

W Polsce problematyka długoterminowej archiwizacji publikacji elektronicznych jest jeszcze mało rozpoznana. Z analizy opracowań przedmiotu oraz prowadzonych badań własnych wynika, że polskie instytucje biblioteczne gromadzą publikacje elektroniczne od połowy lat dziewięćdziesiątych ubiegłego stulecia, które jednak dopiero od roku 2001 są rejestrowane w narodowej bibliografii wydawnictw elektronicznych BN. Natomiast zadania ich długoterminowej archiwizacji nie zostały dotychczas uwzględnione w bieżącej działalności bibliotecznej.

W celu rozpoznania, jak dalece powszechny jest w świadomości pracowników polskich instytucji pamięci problem zachowania cyfrowego dziedzictwa nauki i kultury oraz, czy są oni go-

towi podjąć odpowiedzialność w tym zakresie, przeprowadzono badanie metodą ankiety oraz wywiadu. Miało ono również dostarczyć orientacyjnych danych o ilości oraz kondycji cyfrowych zasobów przechowywanych w polskich instytucjach pamięci oraz stosowanych strategiach ich ochrony. Przedstawione poniżej wnioski uzyskano na podstawie badania sondażowego przeprowadzonego w roku 2007 oraz na podstawie badania głównego, które było prowadzone w 2009 roku. Grupę badawczą utworzyły instytucje biblioteczne posiadające status ksiąźnic narodowych, naukowych i akademickich oraz te, których zbiory tworzą narodowy zasób biblioteczny, podlegający szczególnej ochronie.

1. Długoterminowa archiwizacja zasobów cyfrowych – definicja

Termin *długoterminowa archiwizacja publikacji cyfrowych* identyfikowany jest przede wszystkim z ochroną, zabezpieczaniem i przechowywaniem dokumentów elektronicznych, w celu zagwarantowania ich długoterminowej użyteczności. Słowo *długoterminowy* należy przy tym ogólnie rozumieć jako *nieograniczony w czasie* lub *możliwie najbardziej odległy w przyszłości*. Dookreślenia słowa *długoterminowy* dostarczają autorzy wypowiedzi na temat systemów archiwalnych, projektowanych i implementowanych w bibliotekach i archiwach narodowych. Zadaniem takich systemów ma być długoterminowa, tj. trwająca ponad sto lat, archiwizacja zbiorów cyfrowych [5]. W wypowiedziach pojawia się również koncepcja, według której archiwizacja długoterminowa oznacza pełnienie zadań ochrony zbiorów elektronicznych przez określony zespół ludzi do czasu, kiedy zadania te przejmie zespół następczy, zdolny do ich kontynuacji [14].

Do polskojęzycznej terminologii z tego zakresu wprowadzono termin *ochrona informacji cyfrowych* i zdefiniowano go jako „zbiór rozwiązań służących zapewnieniu ciągłości dostępu do materiałów należących do dziedzictwa cyfrowego w okresie, w którym ciągłość taka jest pożądana” [15]. Przy czym dostępność materiału cyfrowego jest utożsamiana z zachowaniem środków, które będą w stanie zapewnić dostęp do autentycznej treści dokumentu cyfrowego i umożliwią jego wykorzystanie zgodnie z pierwotnym celem.

Wracając jednak do głównego wątku definicji, związanej z długoterminowym zapewnieniem użyteczności cyfrowych zasobów, warto przybliżyć znaczenie tego pojęcia. Otóż użyteczność cyfrowych zasobów oznacza przede wszystkim ich dostępność oraz możliwość ich użytkowania, głównie czytania, słuchania, oglądania zapisanych w nich treści przez osoby upoważnione, w ramach posiadanych przez nie praw dostępu [3, 11]. Użyteczność wiąże się z efektywnym korzystaniem z publikacji elektronicznych, a to staje się możliwe wówczas, gdy użytkownik ma pewność, że treści, które czyta, słucha, ogląda, a także, na które powołuje się w swych opracowaniach, są autentyczne i niezafałszowane, tzn., że pochodzą od ich autorów i od dnia opublikowania nie uległy zmianie; przedstawiają dokładnie to, co było zamierzeniem ich twórców [6, 11]. Obok zagwarantowania dostępności publikacji elektronicznych, za podstawowy cel procesu archiwizacji należy uznać zapewnienie ich autentyczności i integralności [1]. Integralność obiektów archiwalnych wiąże się przede wszystkim z ich kompletnością [11]. System archiwalny powinien być odporny na wszelkiego rodzaju nieupoważnione i niewłaściwe modyfikacje na obiektach archiwalnych. Z technicznego punktu widzenia, pojęcie modyfikacji jest opisywane jako zmiana wartości danych, wstawienie lub usunięcie danych [3]. Integralność może zostać naruszona przez celowe działanie nieuprawnionego użytko-

wnika, błędy i zaniedbania uprawnionego użytkownika, wirusy komputerowe lub inne programy szkodliwe, awarie sprzętu, kanałów komunikacyjnych, zasilania, błędy oprogramowania [3]. Z pojęciem integralności związane jest pojęcie niezmienności danych, czyli brak możliwości dokonania jakichkolwiek zmian, zarówno nieautoryzowanych, jak i autoryzowanych [3].

W literaturze przedmiotu, obok ochrony dostępności oraz użyteczności dokumentów cyfrowych, wśród celów, którym długoterminowa archiwizacja ma służyć, wymienia się także poufność. 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 [3].

Warunkiem koniecznym dla wszelkich czynności zapewniających dostępność i użyteczność publikacji elektronicznych jest utrzymanie ich substancji. Pojęcie *substancji publikacji elektronicznej* jest definiowane jako ciąg bitów (kod zerojedynkowy), zapisany na medium elektronicznym. Utrzymanie substancji publikacji elektronicznych jest uzależnione od dwóch zasadniczych czynników, tj. od ograniczonej trwałości nośników zapisu oraz szybkich zmian zachodzących w dziedzinie formatów zapisu danych [12]. Względy bezpieczeństwa dyktują, aby przed upływem pesymistycznie określonej 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. 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 [12].

Samo zabezpieczenie substancji publikacji elektronicznej zapewniłoby użytkownikom jedynie dostęp do kodu zerojedynkowego. Potrzebne są zatem odpowiedni sprzęt i oprogramowanie, które umożliwią odczytanie zakodowanej w postaci zer i jedynek treści publikacji. Z tej racji, że w historii publikowania elektronicznego znajdowały zastosowanie różne platformy programowo-sprzętowe, w bibliotekach zgromadzono publikacje elektroniczne, których odczyt z nośnika i odszyfrowanie za pomocą aktualnego sprzętu i oprogramowania jest obecnie utrudnione, a często nawet niemożliwe. Przykładem mogą być publikacje zapisane na dyskietkach 5,25 cala. Ich odczyt jest możliwy jedynie przy użyciu stacji dysków, które kilkanaście lat temu wyszły z użycia. Zakładając jednak, że uda się zdobyć odpowiednią stację dysków, kolejną, trudną do pokonania barierą, może stać się dostępność platformy programowo-sprzętowej, niezbędnej do zdekodowania treści publikacji. Pomocne w takiej sytuacji okazuje się „zachowane technologii” [7, 14]. Przechowywanie w bibliotekach sprzętu i oprogramowania, które wyszły z powszechnego użycia i wykorzystywanie ich jedynie w celu odczytywania publikacji zapisanych w formatach specyficznych dla tych platform, jest traktowane jako jedna z możliwych metod długoterminowej archiwizacji publikacji elektronicznych, jednak tworzenie tzw. „muzeów komputerowych” nie spotyka się z uznaniem ze strony specjalistów. Wypowiadając się na temat istotnych metod długoterminowej archiwizacji publikacji elektronicznych, wymieniają migrowanie treści dokumentów i emulowanie systemów [7, 11, 12].

Migracja jest definiowana jako konwersja dokumentu z oryginalnego formatu do nowszego, gdy format oryginalny staje się przestarzały i wychodzi z użycia [4, 8]. Migracja bywa też określana jako „wyciągnięcie” danych cyfrowych z oryginalnego, starzejącego się formatu zapisu, w celu ich konwersji do formatu aktualnie wykorzystywanego [8, 9].

Natomiast pojęcie *emulacja* jest używane dla określenia procesu naśladowania, symulowania, także imitowania zachowań określonego sprzętu i oprogramowania [1, 4]. W *Komputerowej Encyklopedii Microsoftu* [13] termin *emulacja* jest definiowany 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 wykorzystywanych [4, 17]. Emulacja jest traktowana jako jedyny sposób uniknięcia przekłamań i odstępstw kolejnych wersji dokumentów od oryginałów. Jedyną formą zapewnienia użytkownika, że dokument jest w pełni zgodny z oryginalną wersją jest udostępnienie go w jego pierwotnej aplikacji, czyli wyemulowanie oryginalnego otoczenia programowego dokumentu [4, 17].

Obecnie jednak najczęściej stosuje się metodę polegającą na odłączeniu treści publikacji elektronicznej od oryginalnego nośnika zapisu i umieszczeniu jej w celu ochrony w archiwalnym repozytorium cyfrowym. Repozytoria cyfrowe oparte są 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, wykorzystują do składowania danych głównie macierze dyskowe oraz biblioteki taśmowe [1, 3, 14]. Zastosowanie takiego rozwiązania pozwala na eliminację zadań związanych z utrzymywaniem urządzeń obsługujących poszczególne rodzaje nośników, monitorowaniem stanu technicznego poszczególnych egzemplarzy nośników różnorodnych typów oraz podejmowaniem koniecznych działań naprawczych. Przechowywanie wszystkich dokumentów w jednym systemie repozytoryjnym powoduje jednak postawienie przed takim systemem całej listy wymagań związanych z bezpieczeństwem 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. W literaturze przedmiotu system taki jest 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 a jak udostępniane użytkownikom i partnerom zewnętrznym [2, 5, 10]. Do realizacji polityki bezpieczeństwa wykorzystuje 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 [2, 5, 10].

Wymagania związane z bezpieczeństwem systemu repozytoryjnego można podzielić na dwa obszary: ochrona danych przed skutkami deterioracji sprzętu i nośników wykorzystywanych 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 również klęsk żywiołowych.

2. Wnioski z badań własnych – wybrane wątki

Rozwój świadomości pracowników polskich bibliotek odnośnie potrzeby długoterminowej archiwizacji zasobów cyfrowych

Osobiste rozmowy w ramach wywiadów oraz analiza odpowiedzi ankietowych dostarczyły dowodów, że w zagadnieniu archiwizacji istnieje terminologiczny nieład. W polskim środowisku bibliotecznym używane jest pojęcie elektronicznej archiwizacji i rozumie się je jako digitalizację, w sensie zabezpieczenia materiałów drukowanych. Prawdopodobnie przyjmuje się więc, 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 długoterminowej archiwizacji, czyli stosownych zabiegów głównie natury technicznej. Rozmówcom należało uświadamić, że długoterminowa archiwizacja nie powinna oznaczać tylko ochrony materiałów zdigitalizowanych, składowanych na serwerach i na bieżąco udostępnianych, jak to ma miejsce w przypadku bibliotek cyfrowych. Co najmniej tak samo pilne jest objęcie procesami archiwizacji oryginalnych publikacji elektronicznych, zapisywanych na fizycznych nośnikach przenośnych, zwłaszcza tych najstarszych. Po takim wyjaśnieniu rozmówcy przyznawali, że temat długoterminowej archiwizacji publikacji elektronicznych podejmowano dotychczas w Polsce bardzo rzadko i że konieczne jest jego propagowanie. Respondenci wskazywali na wątki dotyczące długoterminowej archiwizacji pojawiające się na seminariach organizowanych przez Centrum Promocji Informatyki.

Na podstawie literatury przedmiotu oraz prowadzonych badań, nasuwa się wniosek, że w Polsce poziom świadomości i wiedzy o archiwizacji publikacji elektronicznych wychodzi z fazy pierwszej, w której – jak już wspomniano – pojęcie długoterminowej archiwizacji utożsamiano z procesami digitalizacji. Digitalizacja sama w sobie miała zapewnić długoterminową użyteczność publikacji. Nie uświadamiano sobie wówczas, że samo zgromadzenie materiału w postaci cyfrowej nie jest jego długoterminowym zabezpieczeniem, gdyż materiał cyfrowy potrzebuje być może staranniejszej ochrony, szczególniejszych warunków przechowywania i metod konserwacji, aniżeli dokumenty analogowe. W Polsce fazę pierwszą kształtowania się świadomości prawdopodobnie można datować na lata 1999-2004. Jej początek jest związany z pierwszymi projektami digitalizacji najbardziej zagrożonych zbiorów drukowanych i pierwszymi sieciowymi kolekcjami dokumentów cyfrowych, a koniec to okres pojawienia się w polskiej literaturze przedmiotu wypowiedzi dotyczących długoterminowej archiwizacji publikacji elektronicznych oraz projektów z tego zakresu realizowanych w krajach zachodnich. Krzywdząca byłaby jednak powyższa charakterystyka w odniesieniu do absolutnie całego środowiska związanego zawodowo z instytucjami bibliotecznymi. Z całą pewnością, istniały wówczas w Polsce jednostki, które wykazywały się trafną znajomością tematu i świadomością potrzeby archiwizowania publikacji elektronicznych we właściwym znaczeniu tego terminu. Potwierdzeniem niech będzie wypowiedź zamieszczona w publikacji J. Czermińskiego z 2002 r. pod tytułem *Cyfrowe środowisko współczesnej biblioteki*. Autor przytacza przykłady zagrożeń istnienia dokumentów, podkreśla konieczność archiwizacji i słusznie zwraca uwagę, że „sam proces digitalizacji jest często zaliczany do czynności archiwizacyjnych i wielu osobom trudno jest pojąć, po co archiwizować archiwum” [7]. Należy jednak zauważyć, że stwierdzenie to odzwierciedla sposób myślenia właściwy w czasach powstawania i publikacji cytowanej książki. Obecnie poziom wiedzy

na temat potrzeby ochrony cyfrowych archiwów, zwłaszcza w środowisku bibliotecznym jest znacznie wyższy.

Przypuszczalnie, świadomość potrzeby archiwizacji publikacji elektronicznych w Polsce znajduje się obecnie w fazie drugiej, w której pod pojęciem długoterminowej archiwizacji publikacji elektronicznych, bądź długoterminowej archiwizacji zasobów cyfrowych rozumie się stosowną ochronę i zabezpieczenie na przyszłość głównie zasobów bibliotek cyfrowych, zgromadzonych i przechowywanych na serwerach. Archiwizację utożsamia się z zabiegami sporządzania kopii zapasowych, z okresowym sprawdzaniem sum kontrolnych, z terytorialnym oddaleniem jednej kopii od miejsca macierzystego archiwum, przy czym są to zadania cedowane na informatyków bądź techników bibliecznych. Rola bibliotekarzy w tym procesie kończy się na sporządzeniu metodą digitalizacji cyfrowego dokumentu, który ma być udostępniany użytkownikom. Z jednej z wypowiedzi wynika, że wśród pracowników instytucji kultury charakterystyczny jest sposób myślenia i działania, zgodnie z którym „podstawowym celem działalności digitalizacyjnej jest udostępnienie zasobów cyfrowych. Powoduje to umieszczanie w bibliotekach cyfrowych plików o niskiej rozdzielczości, podczas gdy pliki macierzyste często są archiwizowane na nośnikach nietrwałych (dyski wymienne, płyty DVD)”. W fazie tej sporadycznie występuje świadomość na temat potrzeby objęcia ochroną publikacji zapisanych i przechowywanych na fizycznych nośnikach przenośnych. Bardzo rzadkie są opinie o pilnej potrzebie prób odczytu publikacji najstarszych.

W obecnej fazie rozwoju świadomości raczej nie mówi się o roli twórców, głównie wydawców publikacji elektronicznych w procesach archiwizacji i nie prowadzi się prac w zakresie opracowywania wytycznych dotyczących publikowania elektronicznego. Zwraca się uwagę na wytyczne i standardy, ale w odniesieniu do procesów digitalizacji i tworzenia bibliotek cyfrowych. Przedstawiciele bibliotek przekonują się jednak, że współpraca z twórcami zasobów cyfrowych jest nieunikniona. Na podstawie rozmowy prowadzonej w Bibliotece Narodowej wiadomo, że realizowany jest projekt tworzenia Narodowego Repozytorium Dokumentów Elektronicznych BN oraz gromadzenia polskich witryn internetowych, czyli tzw. harwestowania polskiego Internetu. Doświadczenia instytucji bibliecznych innych krajów pokazują, że tego typu przedsięwzięcia wymagają ścisłej współpracy, w różnych zakresach, twórców zasobów oraz instytucji, które je gromadzą, chronią i organizują do nich dostęp.

Przypuszczalnie, opisana druga faza poziomu świadomości występuje w dość wąskich kręgach ludzi związanych zawodowo z instytucjami bibliotecznymi. Jej upowszechnienie wymaga stosownych działań propagujących.

Ze wszystkich uzyskanych odpowiedzi, zarówno w badaniu sondażowym, jak i głównym jednoznacznie wynika, że w Polsce istnieje potrzeba propagowania zagadnień związanych z długoterminową archiwizacją zasobów cyfrowych. Mało znane są oferty szkoleń, kursów, konferencji, etc. z tego zakresu tematycznego. Świadomość potrzeby archiwizowania publikacji elektronicznych osiągnie w Polsce pełną dojrzałość wówczas, gdy procesami archiwizacji objęte zostaną zarówno publikacje zdigitalizowane, jak i pochodzące z procesów publikowania elektronicznego, zapisywane niegdyś na dyskietkach, obecnie najczęściej na dyskach typu CD oraz DVD. Nastąpi wówczas trzecia faza rozwoju świadomości, charakterystyczna dla wielu krajów i instytucji bibliecznych, w których od lat odczytuje się regularnie treści publikacji elektronicznych i w zależności od przyjętej strategii archiwizacji, przenosi się 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 kopie zapasowe publikacji. Wszelkie operacje i zabiegi konserwatorskie przeprowadzane na całych kolekcjach cyfrowych, bądź poszczególnych jej obiektach, są szczegółowo dokumentowane. Powołuje się centralną instytucję koordynującą, która kieruje działaniami archiwizacyjnymi w skali kraju, organizuje szkolenia, zarządza diagnozę stanu zasobów elektronicznych w instytucjach pamięci, prowadzi badania i prace w zakresie tworzenia narodowej strategii długoterminowej archiwizacji zasobu cyfrowego zgodnie z najnowszymi wytycznymi i standardami oraz uświadamia instytucjom rządowym i pozarządowym potrzebę finansowania działań archiwizacyjnych.

Z badań wynika, że do polskich instytucji bibliotecznych problematykę długoterminowej archiwizacji publikacji elektronicznych wprowadza się z około dziesięcioletnim opóźnieniem w stosunku do krajów zachodnich. Jest to oczywiście fakt niepokojący i prowokujący pytanie, czy polskim bibliotekarzom i archiwistom uda się zapewnić użyteczność najstarszych polskich publikacji elektronicznych, czy polscy użytkownicy bezpowrotnie utracili możliwość ich odczytu i interpretacji? Nie należy wykluczać, że nie ma podstaw do niepokoju; może okazać się bowiem, że treści najstarszych polskich publikacji elektronicznych nie przedstawiają szczególnej wartości bądź mają swoje odpowiedniki w formie drukowanej. Należałoby jednak przystąpić do stosownych prac diagnostycznych. 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 fazy jej rozwoju.

Zasoby cyfrowe polskich bibliotek w kontekście zagadnień archiwizacyjnych

Z prowadzonych badań wynika, że elektroniczne zasoby polskich bibliotek są ukonstytuowane z dwóch grup publikacji.

Grupę pierwszą stanowią zasoby polskich bibliotek cyfrowych. Powstają one głównie w procesach digitalizacyjnych realizowanych w bibliotekach, ale również jako oryginalne publikacje elektroniczne w wyniku procesów publikowania elektronicznego. Są opublikowane w otwartych formatach, opisane metadanymi, opatrzone identyfikatorami trwałymi i przechowywane na serwerach bibliotek. W większości publikacje te są włączone do ujednoliconego systemu bibliotek cyfrowych dLibra i widoczne w Federacji Bibliotek Cyfrowych. Są objęte ochroną, w sensie zapewnienia bezpieczeństwa typowego dla systemów informatycznych. Zarządzaniem tymi zasobami zajmują się informatycy z Poznańskiego Centrum Superkomputerowo-Sieciowego. W systemie dLibra nie ma dotychczas komponentu odpowiedzialnego za długoterminową archiwizację, jednak dostrzega się potrzebę jego utworzenia. Dla celów archiwizacji zasobów cyfrowych Poznańskie Centrum Superkomputerowo-Sieciowe opracowało Krajowy Magazyn Danych (KMD) z Usługą Platon U-4, określaną też jako Usługa Powszechnej Archiwizacji.

Grupa druga natomiast to materiały elektroniczne, opublikowane na przenośnych mediach fizycznych, zgromadzone w bibliotekach na podstawie ustawy o egzemplarzu obowiązkowym oraz zakupów. W grupie tej występuje dodatkowo podział wewnętrzny na dwie podgrupy, tj. na elektroniczne publikacje, którym postanowiono przypisać określenie „nieużyteczne”, bądź „nieczytelne” oraz na publikacje „użyteczne” i „czytelne”.

Do podgrupy publikacji „nieużytecznych” zaliczono pierwsze polskie dokumenty, opublikowane w cyfrowej postaci, w latach 1994-1999, na dyskietkach typu 5,25 oraz 3,5 cala. Są one składowane przede wszystkim w zbiorach bibliotek, które otrzymały je na podstawie przepisów ustawy o obowiązkowych egzemplarzach bibliotecznych. Określenie „nieczytelne” przyznano im dlatego, że biblioteki utraciły możliwości odczytu i prezentacji ich treści. Odczyt treści z dyskietek typu 5,25 cala jest niemożliwy zarówno z powodu braku sprzętu, jak i oprogramowania, natomiast w przypadku dyskietek typu 3,5 cala w większości tylko z powodu braku oprogramowania; w bibliotekach są jeszcze komputery z odpowiednim napędem. Biblioteki mają zatem poważny problem identyfikacji przedmiotu treści tych publikacji, określenia stopnia ich merytorycznej wartości. Większość z tych publikacji nie jest odnotowana w bazie systemów bibliotecznych, ponieważ ukazywały się głównie jako materiały uzupełniające do publikacji drukowanych i nie było wówczas wiedzy oraz wytycznych, jak postępować w ich przypadku. Były odłączane od dokumentów drukowanych, gromadzone w jednym miejscu i przechowywane, niekiedy do chwili obecnej, bez poddawania ich jakimkolwiek zabiegom natury konserwatorskiej. Prawdopodobnie w przypadku wielu z nich, jedyny odczyt ich treści miał miejsce w procesie opracowywania i włączania do zasobów bibliotecznych. Nie udało się ustalić, czy zadbano o warunki przechowywania, wymagane dla tego typu nośników, w sensie dostępu światła, temperatury oraz stopnia wilgotności. Działania archiwizacyjne w Polsce powinny rozpocząć się właśnie od diagnozy tej grupy dokumentów.

Do podgrupy elektronicznych publikacji „użytecznych” natomiast zaliczono dokumenty publikowane od 2000 r. na dyskach typu CD-ROM oraz DVD-ROM. Ich treści mogą być odczytywane i prezentowane przy użyciu otoczenia sprzętowo-programowego, dostępnego w bibliotekach. Również te dokumenty należało poddawać zabiegom odczytu i prezentacji treści, następnie odświeżeniu bądź zmiany generacji nośnika, albo migracji do nowych formatów. Prace te jednak nie były dotychczas wykonywane.

W przypadku tych publikacji, możliwy i dużo łatwiejszy będzie proces oceny ich treści oraz określenia stopnia przydatności dla użytkowników. Są one bowiem odnotowane w katalogach bibliotecznych, opisane hasłami przedmiotowymi. Podczas prac identyfikacyjnych należałoby ustalić, które z nich to publikacje samoistne wydawniczo, bez odpowiedników w wersji drukowanej, a które mają charakter materiału uzupełniającego dokument drukowany i zdecydować o ewentualnym oddzieleniu treści tych publikacji od oryginalnych nośników w celu umieszczenia i ochrony na serwerze archiwalnym.

Preferencje pracowników bibliotek odnośnie modelu narodowego archiwum zasobów cyfrowych oraz koordynacji polskich działań archiwizacyjnych

Respondenci częściej preferują tworzenie lokalnych archiwów polskich zasobów cyfrowych, zorganizowanych i działających na podstawie ujednoczonych zasad. Zdaniem niewielu respondentów, w Polsce należałoby utworzyć jedno, centralne narodowe archiwum, gromadzące kompletny zasób cyfrowy Polski. W jednej z odpowiedzi argumentuje się jednak, że aby zbudować profesjonalny system archiwalny, zgodny ze światowymi standardami potrzebna jest jednolitość procedur; w przypadku archiwów lokalnych, bardzo prawdopodobne są rozbieżności, które utrudnią funkcjonowanie jednorodnego, rozproszonego archiwum.

Z jeszcze innej opinii wynika, że łatwiejsze i sprawniejsze organizacyjnie wydaje się utworzenie jednego centralnego archiwum. Należałoby jednak rozważyć, czy wadą, czy zaletą jest złożenie funduszy i odpowiedzialności „w jedne ręce”. 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.

Z badań wynika, że nie ma zgodności co do modelu archiwum polskiego dziedzictwa cyfrowego. Niewątpliwie, potrzebne jest 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, zgodnie z odpowiedziami większości respondentów, powinna inicjować i koordynować Biblioteka Narodowa. Oprócz Biblioteki Narodowej najczęściej wymieniano również Poznańskie Centrum Superkomputerowo-Sieciowe, Ministerstwo Kultury i Dziedzictwa Narodowego, także Naczelną Dyрекcję Archiwów Państwowych.

Podczas wywiadu próbowano uzyskać opinię respondenta z Biblioteki Narodowej w kwestii ewentualnego podjęcia przez BN roli koordynatora działań archiwizacyjnych w Polsce. Z odpowiedzi jednak wynika, że za bardziej kompetentne przedstawiciel BN uznaje Ministerstwo Kultury i Dziedzictwa Narodowego. „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”.

Nasuwa się zatem wniosek, że Biblioteka Narodowa, z bardzo 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 sugeruje się również w innej odpowiedzi ankietowej, „...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 właściwie organizować prace archiwizacyjne i nimi kierować.

Na podstawie dodatkowych komentarzy do odpowiedzi respondentów można wnioskować, że dobrym partnerem w procesach archiwizacyjnych dla Biblioteki Narodowej byłoby Poznańskie Centrum Superkomputerowo-Sieciowe.

W opinii respondentów pracownicy PCSS od wielu lat są skupieni na pracach związanych z tworzeniem i zarządzaniem polskimi bibliotekami cyfrowymi oraz z rozwojem projektu Federacja Bibliotek Cyfrowych, tym samym znają polskie środowisko biblioteczne oraz jego potrzeby. Są również świadomi potrzeby działań na rzecz długoterminowej archiwizacji zasobów cyfrowych, posiadają wiedzę i doświadczenie, potrzebne w tych procesach. Prawdopodobnie byłiby cennymi doradcami w zakresie technicznych zagadnień długoterminowej archiwizacji i tworzenia systemu depozytowego. Ponadto, na podstawie badania ustalono, że PCSS jest firmą znaną i wiarygodną dla polskiego środowiska bibliotecznego. Bibliotekarze chętnie powierzyliby właśnie PCSS opiekę nad polskim zasobem cyfrowym.

Zasadne wydaje się także zaangażowanie przedstawicieli świata polityki w opracowywanie strategii długoterminowej ochrony polskiego dziedzictwa cyfrowego, przede wszystkim z racji potrzeby zrozumienia przez rządzących wagi problemu, pilnej potrzeby działania, a przede wszystkim finansowania tychże działań. Bez finansowego wsparcia, tego typu przedsięwzięć nie da się realizować i, co ważniejsze, nie chodzi o finansowanie jednorazowe, bądź krótkoterminowe. Długoterminowa, wieczysta archiwizacja zasobów cyfrowych, wymaga nakładów permanentnych. Z badań sondażowych wynika jednoznacznie, że jednym z wyzwań, z jakimi polskie instytucje pamięci muszą uporać się w procesie długoterminowej archiwizacji jest właśnie pozyskanie środków finansowych. Wśród pozostałych wyzwań respondenci jednoznacznie wymieniają brak planu działania i zaniedbania organizacyjne. W opinii respondentów wyzwaniem są też zmiany technologiczne oraz brak specjalistów, tym samym wiedzy, jak reagować na zmiany technologiczne oraz brak profesjonalnego sprzętu i oprogramowania potrzebnych dla celów długoterminowej ochrony zbiorów cyfrowych. W jednej odpowiedzi wskazano na wyzwanie natury prawnej; rzecz dotyczy nowelizacji przepisów dotyczących procesów archiwizacji cyfrowych zasobów nauki i kultury.

Podsumowanie

Ochrona dziedzictwa cyfrowego stała się obowiązującą w świetle zaleceń i międzynarodowych standardów. *Karta ochrony dziedzictwa narodowego* ustalona przez UNESCO nakłada na państwa moralny obowiązek wyeksponowania kulturowych osiągnięć i cyfrowego zabezpieczenia dziedzictwa narodowego. Rozwinięte społeczności świata bardzo poważnie potraktowały te zalecenia i realizują je z ogromnym zaangażowaniem środków i ludzi [15].

W zgromadzonym materiale badawczym pojawiły się opinie, wprawdzie bardzo rzadkie, ale za to jednoznacznie świadczące, że problematyka długoterminowej ochrony zasobów cyfrowych została wprowadzona również do polskich instytucji bibliotecznych. Świadomość odnośnie potrzeby organizacji i prowadzenia właściwie rozumianych działań archiwizacyjnych w środowisku polskich bibliotekarzy i specjalistów informacji występuje, jakkolwiek wciąż w dosyć wąskich kręgach. Jest ona wyraźnie widoczna w środowisku Biblioteki Narodowej. Uzasadnia to przyjęcie przez BN odpowiedzialności za długoterminowe zachowanie polskiego dziedzictwa cyfrowego oraz podjęcie się roli organizatora i koordynatora polskich działań archiwizacyjnych.

Polskie instytucje biblioteczne stoją przed zadaniem ustalenia, czy polski biblioteczny zasób cyfrowy będzie archiwizowany w lokalnych archiwach cyfrowych, czy też w jednym centralnym archiwum. W tej kwestii potrzebne jest podjęcie współpracy i stworzenie powszechnie akceptowanego, narodowego programu działań archiwizacyjnych. Bez względu na to, czy będzie to model archiwum rozproszonego, czy centralnego, potrzebne jest powołanie instytucji, bądź organizacji, która przyjmie rolę inicjatora i koordynatora działań archiwizacyjnych w Polsce. Jest bardzo prawdopodobne, że właściwym koordynatorem działań archiwizacyjnych w Polsce byłaby BN wraz MKiDN oraz firmą PCSS.

Polskie instytucje biblioteczne w działaniach archiwizacyjnych powinny wzorować się na działalności instytucji bibliotecznych innych krajów, które odpowiedzialność taką już przyjęły i dysponują stosowną wiedzą oraz doświadczeniem. Ważne jest to, aby o bezpieczne przechowanie i użyteczność w długim czasie „kruchych” z natury publikacji elektronicznych zatroszczyć się, zanim na szeroką skalę zdomowi się w Polsce zjawisko ich utraty.

Dużą zmianę w działalności polskich instytucji pamięci odnośnie tworzenia i ochrony zasobów cyfrowych zapowiada Ministerstwo Kultury i Dziedzictwa Narodowego w opublikowanym we wrześniu 2009 roku *Programie digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020*. Z treści dokumentu wynika, że najbliższe dziesięciolecie będzie w instytucjach pamięci dedykowane działaniom archiwizacyjnym [16].

Piśmiennictwo

- [1] *Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources. An RLG-OCLC Report. Draft for Public Comment* [online]; RLG. The Research Libraries Group, Mountain View, CA 2001. [Dostęp: 15 lutego 2010]. Dostępny w World Wide Web: < <http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070511/viewer/file2172.pdf> >.
- [2] A. Barczak, T. Sydoruk, *Bezpieczeństwo systemów informatycznych*. Siedlce: Wydawnictwo Akademii Podlaskiej, 2002.
- [3] T. Bilski: *Pamięć. Nośniki i systemy przechowywania danych*. Warszawa: Wyd. Nauk.-Techn., 2008.
- [4] Uwe M. Borghoff, et al., *Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente*. Heidelberg: dpunkt.verlag, 2003.
- [5] Uwe M. Borghoff, Vergleich bestehender Archivierungssysteme [online]; Universität der Bundeswehr München. Fakultät für Informatik. *Nestor – Materialien* Nr 3, 2005. [Dostęp: 21 stycznia 2010]. Dostępny w World Wide Web: < http://files.d-nb.de/nestor/materialien/nestor_mat_03.pdf >.
- [6] W. Coy, Perspektiven der Langzeitarchivierung multimedialer Objekte [online]; Humboldt Universität zu Berlin. Institut für Informatik. *Nestor Materialien* Nr 5, 2006. [Dostęp: 10 lutego 2010]. Dostępny w World Wide Web: < http://files.d-nb.de/nestor/materialien/nestor_mat_05.pdf >.
- [7] Jurand B. Czermiński, *Cyfrowe środowisko współczesnej biblioteki*. Gdańsk, Wyd. Uniw. Gdań., 2002.
- [8] Stefan E. Funk: Migration. [In:] *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 1.5. Dezember 2008, Nestor, Goettingen 2008 [Dostęp: 20 lutego 2010]. Dostępny w World Wide Web: < <http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf> >.
- [9] G. Fülle, T. Ott: Langzeiterhaltung digitaler Publikationen. Archivierung elektronischer Zeitschriften (E-Journals) [online]. Pagina GmbH, Tuebingen. *Nestor – Materialien* Nr 4, 2006. [Dostęp: 20 lutego 2010]. Dostępny w World Wide Web: < http://files.d-nb.de/nestor/materialien/nestor_mat_04.pdf >.
- [10] J. Januszewicz, T. Lewandowski: *Bezpieczeństwo systemów informatycznych*. Wałbrzych: Wałbrzyska Wyższa Szkoła Zarządzania i Przedsiębiorczości, 2009.
- [11] Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 1: Entwurf zur öffentlichen Kommentierung [online]; Frankfurt am Main. *Nestor – Materialien* Nr 8, 2006. [Dostęp: 16 lutego 2010]. Dostępny w World Wide Web: < <http://edoc.hu-berlin.de/series/nestor-materialien/2006-8/PDF/8.pdf> >.
- [12] H. Liegmann: Langzeitverfügbarkeit digitaler Publikationen. [In:] *Bibliotheken - Portale zum Globalen Wissen. 91. Deutscher Bibliothekartag in Bielefeld 2001*. Hrsg. M. Rützel - Banz. Frankfurt am Main: Vittorio Klostermann, 2001.
- [13] Microsoft. Encyklopedia komputerowa. Pod red. J. Woodcock, et al.: Warszawa 2002.
- [14] Nestor Handbuch. *Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Hg.v. H. Neuroth, et al. Version 2.0, Juni 2009 [online]; Boizenburg, Verlag Werner Hülsbusch 2009. [Dostęp: 15 lutego 2010]. Dostępny w World Wide Web: < http://nestor.sub.unigoettingen.de/handbuch/nestor-handbuch_20.pdf >.
- [15] *Ochrona dziedzictwa cyfrowego: zalecenia*. Oprac. National Library of Australia. Warszawa: Naczelna Dyrekcja Archiwów Państwowych, 2003.
- [16] *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009-2020*. Warszawa: Ministerstwo Kultury i Dziedzictwa Narodowego, 2009 [online]. Narodowy Instytut Audiowizualny, Warszawa 2009. [Dostęp: 11.02.2010]. Dostępny w World Wide Web: < http://www.nina.gov.pl/files/images/Program_digitalizacji_2009-2020.pdf >.
- [17] Rothenberg: An Experiment in Using Emulation to Preserve Digital Publications. *Nedlib Report Series* 1. Den Haag: Koninklijke Bibliotheek, 2000.