

WOKÓŁ
ZAGADNIENÍ
TRWAŁEJ
OCHRONY
ZASOBÓW
CYFROWYCH

WOKÓŁ
ZAGADNIEŃ
TRWAŁEJ
OCHRONY
ZASOBÓW
CYFROWYCH

REDAKCJA NAUKOWA
ANETA JANUSZKO-SZAKIEL

Kraków 2013

Rada Wydawnicza Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego:
Klemens Budzowski, Maria Kapiszewska, Zbigniew Maciąg, Jacek M. Majchrowski

Recenzja: prof. UJ dr hab. Wanda Pindłowa

Projekt okładki: zespół

Adiustacja i korekta: zespół

ISBN 978-83-7571-239-1

Copyright© by Krakowska Akademia im. Andrzeja Frycza Modrzewskiego
Kraków 2013

Wszystkie materiały ilustracyjne zostały przygotowane przez autorów
i są zamieszczone na ich odpowiedzialność

Żadna część tej publikacji nie może być powielana ani magazynowana
w sposób umożliwiający ponowne wykorzystanie,
ani też rozpowszechniana w jakiegokolwiek formie
za pomocą środków elektronicznych, mechanicznych, kopiujących, nagrywających
i innych, bez uprzedniej pisemnej zgody właściciela praw autorskich.

Na zlecenie:



Krakowskiej Akademii
im. Andrzeja Frycza Modrzewskiego
www.ka.edu.pl

Wydawca:

Krakowskie Towarzystwo Edukacyjne sp. z o.o. – Oficyna Wydawnicza AFM
Kraków 2013

Sprzedaż prowadzi:

Księgarnia u Frycza

Kampus Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego

ul. Gustawa Herlinga-Grudzińskiego 1, 30-705 Kraków

tel./faks: (12) 252 45 93

e-mail: ksiegarnia@kte.pl

Skład: Joanna Sroka

Druk i oprawa: Krakowskie Towarzystwo Edukacyjne sp. z o.o.

Spis treści

Aneta Januszko-Szakiel

Słowo wstępne 7

Sabine Schrimpf

Cyfrowe przechowywanie długoterminowe –
najnowsze osiągnięcia i doświadczenia w Niemczech 11

Mary Ren

Przechowywanie zasobów cyfrowych Rosetta Ex Libris 19

Matthias Töwe

Ochrona cyfrowych zasobów w Politechnice Federalnej w Zurychu.
Potrzeby naukowców i wizja Biblioteki Uniwersyteckiej
ETH w Zurychu 27

Anna Jakubiec, Marzena Pazdur

Długoterminowa archiwizacja obiektów cyfrowych –
międzynarodowe projekty 45

Justyna Adamus-Kowalska

Zarządzanie dokumentacją elektroniczną i zabezpieczanie materiałów
archiwalnych dla przyszłych pokoleń według zaleceń specyfikacji
MoReq2 Komisji Europejskiej 59

Dagmara Bubel, Łukasz Kuczyński, Lidia Szczygłowska

PLATON U4 – i wszystko jasne 71

Dorota Witczak, Kamil Panaś, Krzysztof Sobkowiak

W oczekiwaniu na PLATONa. Współpraca Pedagogicznej
Biblioteki Cyfrowej z Akademickim Centrum Komputerowym
Cyfronet AGH w zakresie długoterminowej archiwizacji danych 89

Cezary Mazurek, Tomasz Parkoła, Marcin Werla

dArceo: Usługi długoterminowego przechowywania
danych źródłowych 101

Justyna Maczuga, Jolanta Przytuska

Czy publikacje naukowe powinny mieć Digital Object Identifier? 113

Zdzisław Pietrzyk

Archiwizacja zasobów cyfrowych Jagiellońskiej
Biblioteki Cyfrowej 119

Spis treści

Dorota Olejnik, Elżbieta Gongala, Zofia Kukurowska

Repozytorium Cyfrowe Uniwersytetu Technologiczno-Przyrodniczego
w Bydgoszczy jako przykład opracowania, udostępniania
i archiwizacji dorobku naukowego wykładowców uczelni 133

Marta Stąporek

Organizacja i funkcjonalność wiarygodnego
archiwum cyfrowego uczelni wyższej na przykładzie
Repozytorium Politechniki Krakowskiej 145

Marek Szepski

Biblioteka naukowa w Internecie – od formy do funkcji 161

Aneta Januszko-Szakiel

Narodowy program długoterminowej archiwizacji
cyfrowego zasobu nauki i kultury – propozycja dla Polski..... 173

Słowo wstępne

Cyfrowe kolekcje wartościowych dokumentów powstają w wyniku procesów elektronicznego publikowania oraz digitalizacji materiałów analogowych. Są one świadectwem dorobku polskiej działalności naukowej i kulturowej. Digitalizacja upowszechnia się w instytucjach pamięci jako forma zabezpieczenia zagrożonych zniszczeniem oryginalnych materiałów. Oficyny wydawnicze coraz częściej publikują treści naukowe w cyfrowej postaci, rezygnując z wersji drukowanych. Instytuty naukowe i badawcze oraz uczelnie wyższe tworzą instytucjonalne repozytoria, pozwalające na szybkie i samodzielne udostępnianie publiczności wyników prowadzonych prac badawczych. Powstające w ten sposób cyfrowe zasoby współtworzą światowe dziedzictwo. Podlegają one trwałej ochronie z myślą o potrzebach obecnych i przyszłych użytkowników.

Dokumenty cyfrowe, w odróżnieniu od drukowanych książek i czasopism, odłożone na magazynową półkę, nie przetrwają dłuższego czasu. Nie ulega wątpliwości, że bez poddawania ich okresowym zabiegom archiwizacyjnym, zostanie utracony dostęp do ich treści. Stąd szczególnego znaczenia nabiera zagadnienie bezpiecznego przechowywania dokumentów cyfrowych.

W wielu krajach, także w Polsce, długoterminowa archiwizacja zasobów cyfrowych została wpisana na listę priorytetowych zadań i włączona na stałe w zakres obowiązków instytucji ustawowo powołanych i odpowiedzialnych za zabezpieczenie narodowego dziedzictwa nauki i kultury.

Publikacja ta stanowi zbiór czternastu artykułów skoncentrowanych wokół zagadnień tworzenia, organizowania i trwałej ochrony kolekcji cyfrowych dokumentów. Swoją wiedzę, doświadczeniami i pomysłami dotyczącymi digitalizacji, elektronicznego publikowania oraz bezpiecznego archiwizowania cyfrowych zasobów dzielą się przedstawiciele polskich oraz zagranicznych instytucji nauki i kultury.

Tom otwiera cykl tekstów prezentujących przedsięwzięcia i doświadczenia archiwizacyjne instytucji zagranicznych.

W pierwszym artykule, Sabine Schrimpf, przedstawicielka niemieckiej księżnicy narodowej we Frankfurcie nad Menem, opisuje osiągnięcia oraz perspektywy rozwoju działalności archiwizacyjnej w Niemczech. Szczególną uwagę autorka zwraca na działania grupy roboczej *Nestor*, znanej w świecie między innymi z aktywnego udziału w pracach nad współtworzeniem katalogu kryteriów organizowania i funkcjonowania wiarygodnych oraz certyfikowanych archiwów cyfrowych.

Przedmiotem drugiego tekstu są funkcjonalności systemu *Ex Libris Rosetta*, dedykowanego zadaniom zarządzania i trwałego przechowywania zasobów cyfrowych, opracowanego przez *Ex Libris Group*. Mary Ren, reprezentantka tej firmy, podkreśla istotne cechy systemu, wynikające z uwzględnienia przez twórców zaleceń modelu *OAIS*, uznanego przez Komitet Normalizacyjny *ISO* za standard w zakresie długoterminowej archiwizacji publikacji elektronicznych.

Autor kolejnego artykułu, Matthias Töwe, opisuje koncepcję zarządzania publikacjami elektronicznymi i ich trwałej archiwizacji na przykładzie projektu realizowanego w bibliotece politechniki w Zurychu, której jest pracownikiem. W projekcie zastosowano system trwałej archiwizacji *Rosetta*.

Międzynarodowe projekty i programy długoterminowej archiwizacji stanowią także temat rozważań Anny Jakubiec i Marzeny Pazdur, reprezentujących Bibliotekę Politechniki Krakowskiej. Autorki skupiają uwagę na projektach archiwizacyjnych realizowanych w ramach programów Unii Europejskiej.

W prezentowanym zbiorze znajduje się opracowanie dotyczące problematyki zabezpieczania materiałów archiwalnych, z uwzględnieniem zaleceń Komisji Europejskiej. Justyna Adamus-Kowalska z Uniwersytetu Śląskiego w Katowicach omawia specyfikację *MoReq2* Komisji Europejskiej dotyczącą sprawnego zarządzania i właściwej ochrony archiwaliów.

W kolejnych prezentowanych artykułach autorzy odnoszą się do rozwiązań długookresowej archiwizacji, zaproponowanych przez Poznańskie Centrum Superkomputerowo-Sieciowe (*PCSS*).

Przedstawiciele Politechniki Częstochowskiej (Dagmara Bubel, Łukasz Kuczyński i Lidia Szczygłowska) oraz Biblioteki Głównej Uniwersytetu Pedagogicznego w Krakowie (Kamil Panaś, Krzysztof Sobkowiak i Dorota Witczak) w dwóch artykułach podnoszą kwestię rosnącego zapotrzebowania na usługi bezpiecznego i efektywnego przechowywania danych cyfrowych oraz charakteryzują usługi archiwizacyjne opracowane przez *PCSS* w ramach projektu o nazwie *Krajowy Magazyn Danych*.

Efekty prac badawczo-rozwojowych prowadzonych w strukturach *PCSS*, w swym artykule przybliżają także Cezary Mazurek, Tomasz Parkoła oraz Marcin Werla. Autorzy są pracownikami *PCSS-u*, bezpośrednio zaangażo-

wanymi w rozwój usług archiwizacyjnych. Przedmiotem artykułu, prezentowanego w niniejszej publikacji, jest usługa dArceo opracowana z myślą o długoterminowym przechowywaniu danych źródłowych.

Swoimi spostrzeżeniami dotyczącymi tworzenia bibliotek i repozytoriów cyfrowych dzieli się autorzy pięciu kolejnych tekstów.

Identyfikatory trwale stanowią temat artykułu Justyny Maczugi i Jolanty Przyłuskiej z Instytutu Medycyny Pracy w Łodzi. Autorki analizują proces nadawania numeru DOI publikacjom cyfrowym, składowanym w systemach depozytowych oraz wpływ systemu identyfikacyjnego na zarządzanie zdeponowanymi zasobami.

O zasadach ochrony, konserwacji i zabezpieczania zarówno tradycyjnych, jak i cyfrowych materiałów bibliotecznych traktuje artykuł Zdzisława Pietrzyka. Dyrektor Biblioteki Jagiellońskiej odnosi się do zagadnień bezpiecznego przechowywania zbiorów cyfrowych w kontekście prac nad tworzeniem Jagiellońskiej Biblioteki Cyfrowej.

Opisu organizacji i funkcjonowania instytucjonalnych repozytoriów cyfrowych uczelni wyższych podjęły się Dorota Olejnik, Elżbieta Gongala i Zofia Kukurowska z Biblioteki Głównej Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy oraz Marta Stąporek reprezentująca Bibliotekę Politechniki Krakowskiej. Autorki odnoszą się również do zagadnień trwałej archiwizacji zasobów repozytoryjnych.

Kontrowersyjne spojrzenie na procesy digitalizacji i współczesną bibliotekę prezentuje Marek Szepski z Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego. Autor podejmuje próbę odpowiedzi na pytanie, co tak naprawdę przechowuje współczesna biblioteka. Wymienia również nowe cele, zadania i możliwości funkcjonowania bibliotek.

Ostatni artykuł, mojego autorstwa, to propozycja organizacji działań archiwizacyjnych na rzecz trwałego zachowania cyfrowego zasobu polskiej nauki i kultury.

W zgromadzonych tekstach autorzy odnoszą się do wybranych wątków tworzenia i długoterminowej archiwizacji zasobów cyfrowych. Podejmują kwestie interesujące, trudne, niekiedy rozpatrywane w kategoriach wyzwania. Przekazują Państwu tę publikację z nadzieją, że dostarczy inspiracji do prac teoretycznych i badawczych na rzecz zachowania dla przyszłych pokoleń świadectwa współczesnej naukowej i kulturowej działalności.

Aneta Januszko-Szakiel

Sabine Schrimpf

Deutsche Nationalbibliothek

Cyfrowe przechowywanie długoterminowe – najnowsze osiągnięcia i doświadczenia w Niemczech

Słowa kluczowe: Deutsche Nationalbibliothek, Kopal, Nestor, długoterminowa archiwizacja w Niemczech

Abstrakt: system informacyjny i biblioteczny w Niemczech jest zdecentralizowany. Na tej podstawie przyjęto założenie o tworzeniu zdecentralizowanego modelu infrastruktury archiwizacji długoterminowej. W ramach projektu Kopal Deutsche Nationalbibliothek opracowała koncepcję wspólnie użytkowanego trwałego archiwum. Dalszemu rozwojowi koncepcji dedykowano projekt DP4Lib, w wyniku którego powstał serwis długoterminowego archiwizowania. Idea wspólnie opracowywanej i użytkowanej infrastruktury trwałej archiwizacji zainspirowała grupę kompetencyjną Nestor. Działania grupy Nestor koncentrują się wokół zagadnień łączenia kompetencji i doświadczeń, standaryzacji, kształcenia i dokształcania.

Wstęp

Obecnie niemieckie instytucje dziedzictwa kulturowego odwołują się do dziesięcioletniego doświadczenia w rozwijaniu struktur współpracy na rzecz przechowywania zasobów cyfrowych. W przeciągu tych lat sfinansowano kilka projektów, rozwinięto pilotażowe programy nowych rozwiązań oraz rozważono wiele strategii i zagadnień. Przedłożony artykuł skupia się na rozwoju systemu *Nestor* (*Network of Expertise in long-term Storage of digital Resources*) – sieci eksperckiej w dziedzinie długoterminowego przechowywania zasobów cyfrowych, która powstała w 2004 roku, oraz na jej osiągnięciach. Omówiono również cechy platformy, w której *Nestor* został opracowany, oraz inne związane z nim projekty.

Rozproszony system bibliotek i archiwów

Podobnie do systemu politycznego, system bibliotek i archiwów w Niemczech jest zdecentralizowany i rozproszony; charakteryzuje go dobrze rozwinięty poziom bibliotek i archiwów krajowych oraz regionalnych. System biblioteczny rozdzielił również tematykę posiadanych

zbiorów na pewną grupę instytucji bibliotecznych¹. Wszystkie biblioteki i archiwa zajmują się publikacjami i aktami cyfrowymi: nie tylko je pozyskują, klasyfikują i udostępniają, ale również archiwizują długoterminowo. Digitalizacja doprowadziła do głębokich zmian w rutynie pracy wszystkich tego rodzaju instytucji, co więcej, rzuciła nie lada wyzwanie ich możliwościom przechowywania, które muszą zostać dostosowane do wymagań materiałów cyfrowych. W celu maksymalnego wykorzystania ograniczonych zasobów instytucji, dwie zasady muszą być przestrzegane: podział pracy i współpraca skupiona na osiągnięciu konkretnych celów. Zasady te leżą u podstaw koncepcji projektu *Nestor*.

Nestor – cele i misja

Nestor był pierwotnie finansowany ze środków niemieckiego Federalnego Ministerstwa Edukacji i Badań Naukowych. W pierwszym okresie finansowania (2003–2006) głównym celem było przygotowanie struktur współpracy dla przechowywania cyfrowego w Niemczech, włączając w to koordynację zadań dla poszczególnych bibliotek, archiwów i muzeów, rekomendacje dla narodowej polityki przechowywania długoterminowego, kryteria i rekomendacje dla certyfikacji długoterminowych archiwów cyfrowych oraz system informacyjny oparty na sieci Web. Do założeń drugiego okresu finansowania (2006–2009) należało wypracowanie i koordynacja technicznych standardów przechowywania długoterminowego, rozwój programów treningowych, jak i programu formalnych kwalifikacji oraz przygotowanie trwałych struktur organizacyjnych dla sieci eksperckiej. Od momentu odcięcia dotacji w roku 2009, partnerzy projektu *Nestor* kontynuują pracę w ramach własnych środków. Misja *Nestor* została nieco zmieniona i obejmuje teraz grupowanie dostępnej ekspertyzy o cyfrowym przechowywaniu, łączenie aktywnych instytucji i inicjatyw oraz usprawnianie wymiany informacji. W ten sposób działania projektu *Nestor* nabrały międzysektorowego charakteru. Łączą się tu interesy takich instytucji, jak biblioteki, archiwa, muzea, uczelnie wyższe, instytucje naukowe oraz centra multimedialne. W kwestiach organizacyjnych podstawą działania grupy *Nestor* jest Porozumienie Współpracy zawarte pomiędzy dwunastoma partnerami, na mocy którego, zadania i obowiązki są rozdzielone. Biuro projektu *Nestor* mieści się w Niemieckiej Bibliotece Narodowej i jest finansowane przez tę instytucję.

¹ Deutsche Forschungsgemeinschaft, *Richtlinien zur überregionalen Literaturversorgung der Sondersammelgebiete und Virtuellen Fachbibliotheken* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.dfg.de/download/pdf/foerderung/programme/lis/richtlinien_lit_versorgung_ssg.pdf.

Nestor – działalność

Nestor za trzy podstawowe obszary swojej działalności przyjmuje: nawiązywanie kontaktów, kwalifikację oraz standaryzację. W zakresie tworzenia bazy kontaktów, *Nestor* prowadzi kilkanaście grup roboczych, które organizują różnego rodzaju wydarzenia i warsztaty. Na stronie internetowej *Nestor* można odnaleźć najważniejsze projekty związane z przechowywaniem długoterminowym w krajach niemieckojęzycznych. Grupy robocze *Nestor* (WG – *working group*) zajmują się szerokim spektrum zagadnień i dostarczają podstawowe informacje zainteresowanym instytucjom oraz osobom indywidualnym, chcącym wziąć udział w projekcie. Współuczestnictwo instytucji, które nie są członkami *Nestor*, jest mile widziane. Grupa *WG Cooperation and Networking* stanowi pewnego rodzaju forum dla wszystkich zainteresowanych podjęciem wspólnych wysiłków na rzecz przechowywania cyfrowego. Grupa ta zajmuje się wszelką tematyką w tym zakresie na zasadach współpracy. Ostatnim, pomyślnie zakończonym projektem tej struktury było zdefiniowanie kryteriów dla komponentów i rozwiązań w przechowywaniu długoterminowym. Grupa *WG Costs* zrzesza osoby zainteresowane finansowymi aspektami przechowywania cyfrowego. Uczestnicy tej grupy starają się wspólnie wypracować wytyczne dla określenia kosztów przechowywania długoterminowego. *WG Media* z kolei zajmuje się ustaleniem najlepszego sposobu postępowania w celu zapewnienia długotrwałej dostępności cyfrowych materiałów nietekstowych.

W kwestiach kwalifikacji *Nestor* współpracuje z 12 instytucjami szkolnictwa wyższego (ISW). Podpisały one umowę o porozumieniu w celu wspólnego rozwijania pojęcia treningu, edukacji i instruktażu w dziedzinie przechowywania cyfrowego oraz na rzecz stworzenia wspólnego programu rozwoju. Partnerzy ISW *Nestor* organizują także coroczną „Szkołę *Nestor*” – trzydniowe seminarium dla studentów i profesjonalistów, wydają podręcznik *Nestor* oraz niemieckojęzyczną encyklopedię przechowywania cyfrowego². Wszystkie publikacje *Nestor* są ogólnodostępne na oficjalnej stronie projektu, niektóre z nich są przetłumaczone na język angielski³.

W dziedzinie standaryzacji *Nestor* nawiązał współpracę z Niemieckim Instytutem Standaryzacji (DIN). W okresie kiedy projekt był dofinansowany, wypracowano kilkanaście wytycznych i katalogów kryteriów. W roku 2009 trzy projekty zostały przekazane DIN w celu formalnej standaryzacji: zbiór

² H. Neuroth i in., *Nestor Handbuch, Eine kleine Enzyklopädie der digitalen Langzeitarchivierung* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf.

³ *Nestor* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.langzeitarchivierung.de>.

kryteriów *Nestor* dla wiarygodnych repozytoriów cyfrowych (opublikowany jako DIN 31644), przewodnik dla transferu informacji do repozytoriów cyfrowych (opublikowany jako DIN 31645) oraz zbiór kryteriów do oceny wiarygodności systemów identyfikacji trwałej (opublikowany jako DIN 31646). Komisja odpowiedzialna za standaryzację w tym obszarze w strukturach DIN, Komisja Standardów Informacji i Dokumentacji oraz Podkomisja ds. zarządzania i długoterminowego przechowywania obiektów cyfrowych (NABD 15) jednocześnie reprezentują niemieckie interesy w kwestiach procesów standaryzacji w organizacji ISO.

Certyfikacja

Szczególnie ważnym i długo wyczekiwany krok w budowie infrastruktury był standard DIN 31644, dotyczący wiarygodnych repozytoriów cyfrowych. Był to krok nie tyle ważny ze względu na możliwość standaryzacji kryteriów archiwistyki cyfrowej, ale przede wszystkim dlatego, że będzie on podstawą budowy systemu certyfikacji repozytoriów. W 2010 roku grupa robocza *Nestor WG Certification* opracowała procedurę certyfikacji, opartą na DIN 31644. Etap pierwszy składa się z rozszerzonej autoewaluacji na podstawie DIN, wraz z oceną i nagrodą w postaci „pieczęci *Nestor*”. Czynności te są koordynowane przez inicjatywy międzynarodowe, mianowicie: Pieczęć Zatwierdzenia Danych (*Data Seal of Approval*), dostosowaną do charakteru naukowych danych archiwalnych, Audyt Repozytoriów ISO oraz Roboczą Grupę Certyfikacji, która za podstawę procesu certyfikacji uznaje ISO 16363 „Audyt i certyfikacja wiarygodnych repozytoriów cyfrowych”. Punktem startowym grupy roboczej ISO, w kontekście *Nestor*, jest referencyjny model OAIS. W swojej ostatecznej formie ISO 16363 ściśle odpowiada jednostkom funkcjonalnym OAIS, podczas gdy DIN 31644 bierze pod uwagę narodowe podejście i rezultaty prac, takie jak DINI – Certyfikat dla Dokumentów i Usług Publicznych⁴.

Na skutek sugestii Komisji Europejskiej, trzy inicjatywy dotyczące certyfikacji zgodziły się być częścią umowy o porozumieniu na rzecz stworzenia europejskiej platformy audytu i certyfikacji repozytoriów cyfrowych. Umowa przewiduje zintegrowane podejście do audytu i certyfikacji archiwów cyfrowych, składające się z trzech poziomów. Poziom pierwszy, Certyfikacja Podstawowa, wymaga posiadania Pieczęci Zatwierdzenia Danych. Poziom drugi, Certyfikacja Rozszerzona, wymaga Certyfikacji Podstawowej oraz

⁴ Deutsche Initiative für Netzwerkinformation e.V., *DINI Certificate Document and Publication Services* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://edoc.hu-berlin.de/series/dini-schriften/2010-3-en/PDF/dini-zertifikat-2010-3-en.pdf>.

ustrukturuwanego, zewnętrznego i publicznie dostępnego autoaudytu, opartego na ISO 16363 lub na ekwiwalencie DIN 31644. Poziom trzeci, Certyfikacja Formalna, wymaga Certyfikacji Podstawowej oraz pełnego, zewnętrznego audytu i certyfikacji opartej na ISO 16363 lub DIN 31644.

Wszystkie archiwa cyfrowe, którym przyznano pieczęć *Nestor*, są dostępne w rejestrze na oficjalnej stronie projektu. Partnerzy *Nestor* są przekonani, że taka certyfikacja wprowadza atmosferę zaufania, która jest niezbędna w infrastrukturze przechowywania danych. Przejrzysty system oceny repozytoriów, który za podstawę ma standardowe, wspólnie obrane kryteria, sygnalizuje zdolność danego repozytorium do niezawodnego i długoterminowego przechowywania materiałów. Certyfikacja ma na uwadze zapewnienie użytkownikom i sponsorom poczucia zaufania, promowanie partnerstwa certyfikowanych repozytoriów, ale również uproszczenie współpracy z (certyfikowanymi) dostawcami zewnętrznymi.

Systemy i usługi przechowywania

Od roku 2004 projekt o nazwie *Kopal* (Kooperacyjny Rozwój Długoterminowego Cyfrowego Archiwum Informacji) jest finansowany ze środków niemieckiego Federalnego Ministerstwa Edukacji i Badań Naukowych. Partnerami inicjatywy są Deutsche Nationalbibliothek (DNB), Staats- und Universitätsbibliothek Göttingen (SUB Göttingen), Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) oraz IBM Niemcy. *Kopal* jest uważany za techniczny odpowiednik inicjatywy *Nestor*. Bazujący na DIAS, implementacji modelu OAIS do celów standaryzacji, system *Kopal* został opracowany w celu wspólnego użytkowania i zdalnego dostępu do zasobu archiwalnego. Oprogramowanie otwarte koLibRI (*kopal Library for Retrieval and Ingest*) utworzono, aby umożliwić indywidualnym użytkownikom dostęp do systemu. Pozwalało ono na zmianę konfiguracji w celu zaspokojenia odmiennych potrzeb. Ogólnie rzecz biorąc, to rozwiązanie otwiera użytkownikom drogę do współdzielenia jednego systemu archiwalnego, podczas gdy wciąż zachowują oni swoje własne dane poprzez aplikację koLibRI.

Pomimo faktu, że rozwiązania te wzbudziły powszechne zainteresowanie wśród niemieckich instytucji dziedzictwa kulturowego, złożoność i koszt systemu *Kopal* odstraszyły wielu potencjalnych użytkowników. DNB oraz SUB Göttingen połączyły siły z sześcioma instytucjami partnerskimi w celu opracowania innych rozwiązań, tak by spełniały wymagania partnerów. Ich wspólny wysiłek nazwano *DP4Lib* (*Digital Preservation for Libraries*) i był on finansowany przez Niemiecką Fundację Badawczą (DFG) w latach

2009–2012. Założeniem projektu było ustanowienie gotowych do natychmiastowego zastosowania usług wspierających długoterminowe przechowywanie materiałów cyfrowych. W tym celu partnerzy skompilowali swoje wymagania w katalogu. Rozważano różne modele usług, od ogólnych do modułów zamawianych indywidualnie. Ostatecznie, model ogólny został wdrożony do infrastruktury niemieckiej biblioteki narodowej. Opracowano również kosztorys, by móc ocenić usługi pod kątem liczby i objętości materiałów archiwalnych względem określonych stopni jakościowych.

Równoległe do rozwoju *Kopal/DP4Lib*, kilka bibliotek i archiwów w Niemczech albo wypracowało swoje własne rozwiązania archiwistyczne albo nabyło rozwiązania komercyjne, stąd niejednorodność zaimplementowanych w instytucjach systemów do archiwizowania. W tym kontekście istnienie struktur takich jak *Nestor*, pozwala instytucjom na podzielenie się swoimi doświadczeniami, połączenie sił wobec ogólnych, pokrywających się w wielu obszarach, kwestii (np. standardy, koszty), umożliwia uporanie się z ryzykiem, że zróżnicowanie techniczne doprowadzi do wytworzenia odmiennych rozwiązań.

Podsumowanie

Rozwój infrastruktury w omawianym zakresie będzie uwydatniony na poziomie polityki rozwoju w przeciągu najbliższych lat. Zgodnie z koncepcjami rozwoju infrastruktury informacji naukowej w Niemczech⁵, która została zatwierdzona na konferencji naukowej rządów krajów związkowych i federacji oraz przy rekomendacji Niemieckiej Rady Naukowej dla dalszego rozwoju infrastruktur informacji naukowej dla Niemiec do roku 2020⁶, Niemiecka Fundacja Badawcza ogłosiła apel o zgłaszanie propozycji projektów w tym zakresie. Zgodnie z założeniami programu dedykowanego uaktualnieniu narodowych usług informacyjnych, na które zgłoszono zapotrzebowanie, utworzono międzyregionalny, skalowalny, wieloklientowy system do przechowywania cyfrowego. Uruchomienie tych projektów zaplanowano na połowę 2013 roku.

⁵ Kommission Zukunft der Informationsinfrastruktur, *Gesamtkonzept für die Informationsinfrastruktur in Deutschland* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf.

⁶ Wissenschaftsrat, *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.

Bibliografia

1. Deutsche Forschungsgemeinschaft, *Richtlinien zur überregionalen Literaturversorgung der Sondersammelgebiete und Virtuellen Fachbibliotheken* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.dfg.de/download/pdf/foerderung/programme/lis/richtlinien_lit_versorgung_ssg.pdf.
2. Deutsche Initiative für Netzwerkinformation e.V., *DINI Certificate Document and Publication Services* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://edoc.hu-berlin.de/series/dini-schriften/2010-3-en/PDF/dini-zertifikat-2010-3-en.pdf>.
3. Kommission Zukunft der Informationsinfrastruktur, *Gesamtkonzept für die Informationsinfrastruktur in Deutschland* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf.
4. Nestor [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.langzeitarchivierung.de>.
5. Neuroth H. i in., *Nestor Handbuch, Eine kleine Enzyklopädie der digitalen Langzeitarchivierung* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf.
6. Wissenschaftsrat, *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.

Mary Ren

Ex Libris Group

Przechowywanie zasobów cyfrowych Rosetta Ex Libris

Słowa kluczowe: archiwizacja zasobów cyfrowych, system Rosetta Ex Libris, model referencyjny OAIS

Abstrakt: czy Twoja organizacja ma plany dotyczące zapewnienia w najbliższych latach dostępności wyników prowadzonych badań? Czy Ty masz system zarządzania ryzykiem, który ostrzega przed czasem, że format materiału cyfrowego staje się przestarzały?

Z uwagi na przyrastającą ilość cyfrowych materiałów i krótki cykl ich życia, a także ciągły rozwój formatów zapisu danych cyfrowych, poszukiwanie odpowiedzi na powyższe pytania, i tym samym, temat ochrony zasobów cyfrowych jest w ostatnich latach podejmowany coraz częściej.

Ex Libris, we współpracy z Biblioteką Narodową Nowej Zelandii oraz międzynarodową grupą ekspertów Peer Review Group, opracował rozwiązanie Ex Libris Rosetta, wychodzące naprzeciw wyzwaniu ochrony zasobów cyfrowych.

System Rosetta został opracowany na podstawie modelu Open Archival Information System oraz standardu PREMIS, z uwzględnieniem zaleceń Trusted Digital Repository. Jest to system elastyczny i skalowalny, idealnie nadający się do zarządzania cyfrowymi zbiorami różnych rozmiarów, umożliwiając efektywne przyjmowanie nowego materiału. Dzięki aplikacji Ex Libris Primo, Rosetta zapewnia również wyszukanie oraz udostępnienie wprowadzonego materiału, niezależnie od jego formatu.

W artykule przybliżono, w jaki sposób Rosetta pomaga instytucjom nie tylko zarządzać ich rosnącymi zbiorami materiałów cyfrowych, ale także zapewnić ich dostępność w długim czasie.

Wstęp

Mimo posiadania przez wiele organizacji systemów przechowywania i zarządzania zasobami cyfrowymi, rozwiązania te nie były w zamyśle stworzone w tym celu. Przechowywanie cyfrowe polega na zagwarantowaniu ciągłej użyteczności oraz dostępu do treści cyfrowych jutro, a także w dalekiej przyszłości. Systemy zarządzania zasobami cyfrowymi oraz cyfrowe repozytoria mają na celu ułatwienie codziennego użytkownika zasobów cyfrowych, podczas gdy systemy cyfrowego przechowywania umożliwiają znalezienie, poprzez różne opcje dostępu, pożądaných danych, zapewniają funkcjonalność oraz pozwalają na przepływ wprowadzanych materiałów, bieżących analiz ryzyka

oraz oferują niezakłóconą integralność przechowywanych danych. Mimo iż przechowywanie skupia się na zarządzaniu ryzykiem, byłibyśmy w błędzie, przyrównując przechowywanie do zwykłego tworzenia kopii zapasowych lub krytycznego odzyskiwania danych.

Potrzeba przechowywania cyfrowego

Badania międzynarodowe wykazały, że „świat cyfrowy” rośnie w niewyobrażalnym tempie. Obecne prognozy przewidują, że liczba treści cyfrowych – od nagrań telewizji przemysłowej, poprzez wyniki badań naukowych przedstawione w formie cyfrowej, po liczbę stron internetowych oraz korporacyjnych sprawozdań finansowych – będzie się podwajała w stosunku do ostatniego okresu, co 18 miesięcy¹.

Choć możliwości tworzenia i duplikowania zasobów cyfrowych mnożą się, realność ich przechowywania jest wciąż ograniczona. Dokumenty, obrazy i filmy wyprodukowane w ostatnich dekadach nie będą dostępne dla przyszłych pokoleń z powodu rosnącej fali rozwiązań innowacyjnych, które sprawiają, że starsze systemy przechowywania stają się anachroniczne.

Jak to zostało przedstawione w białej księdze opublikowanej przez organizację Planets (*Preservation and Long-term Access through Networked Services* – Przechowywanie i Długoterminowy Dostęp poprzez Usługi Sieciowe; projekt finansowany ze środków unijnych w celu ewaluacji potrzeb długoterminowego przechowywania cyfrowego): „Im więcej informacji produkujemy oraz przechowujemy w danym miejscu, tym krótszy czas, w którym mamy do nich dostęp².

Biblioteki, muzea oraz archiwa muszą stawić czoła brutalnej rzeczywistości środowiska cyfrowego. Instytucje te powinny nie tylko przechowywać i zapewniać dostęp do tysięcy obiektów i zasobów, którymi dysponują, ale są również odpowiedzialne za to, by te olbrzymie ilości danych cyfrowych, każdego miesiąca pozyskiwane i magazynowane, były dostępne dla przyszłych pokoleń.

Wraz ze wzrostem natywnie cyfrowych treści, w przeciwieństwie do dokumentów digitalizowanych z kopii fizycznych, wyzwania „instytucji pamięci publicznej” (biblioteki, muzea i archiwa) przyjmują wyjątkowy charakter.

¹ Biała księga przygotowana przez IDC, sponsorowana przez EMC Corporation, J. Gantz, D. Reinsel, *As the Economy Contracts, the Digital Universe Expands* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.emc.com/collateral/leadership/digital-universe/2009DU_final.pdf.

² Biała księga organizacji Planets, P. Sinclair, *The Digital Divide. Assessing Organisations' Preparations for Digital Preservation* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.planets-project.eu/docs/reports/planets-market-survey-white-paper.pdf>.

Podczas gdy sprawozdania instytucji rządowych, przedsiębiorstw czy agencji mogą stać się przestarzałe i nieistotne z biegiem czasu, repozytoria wiedzy publicznej powinny zabezpieczyć swoje zbiory, zapewniając ich długotrwałą dostępność.

Badania przeprowadzone przez Planets wykazały, że chociaż w chwili obecnej 80% organizacji przechowuje dokumenty i obrazy, to do roku 2019 ponad 70% organizacji będzie musiało przechowywać także bazy danych, strony internetowe oraz pliki audio i wideo³. Przed bibliotekami stało zatem nowe zadanie, oprócz magazynowania zasobów cyfrowych różnego charakteru, muszą one wyposażyć instytucje w narzędzia umożliwiające skalowanie ich wysiłków w zakresie zabezpieczenia zasobów, by sprostać szybko zwiększającej się ilości danych.

Nawet eksperci nie są w stanie precyzyjnie określić prędkości, z jaką rozwija się świat cyfrowy. Badania przeprowadzone przez International Data Corporation (IDC) w roku 2010 zbyt nisko oszacowały tempo wzrostu w roku poprzedzającym o 3%, podczas gdy we wcześniejszej publikacji z 2008 roku przekłamano globalną liczbę zbiorów cyfrowych o 10%⁴. Dla instytucji, które przewidują przyjęcie milionów obiektów cyfrowych każdego roku, skalowalność jest fundamentalnym wymaganiem stawianym przed architekturą systemu przechowywania cyfrowego.

System przechowywania cyfrowego

Model *Open Archival Information System (Otwarty System Archiwizacji – OAIS)* określa cechy systemu przechowywania cyfrowego. Model ten został zaakceptowany przez organizacje i ekspertów w dziedzinie przechowywania zasobów na całym świecie i jego podstawy są uznawane za wytyczne podczas ewaluacji nowo wdrażanych systemów archiwizowania.

Model OAIS precyzuje sześć funkcji, które muszą być zawarte w systemie przechowywania cyfrowego:

- Wprowadzanie danych
- Przechowywanie
- Zarządzanie danymi
- Administrowanie

³ *Ibid.*

⁴ Biała księga IDC, sponsorowana przez EMC Corporation, J.F. Gantz et al., *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011*, [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.

- Planowanie przechowywania
- Udostępnianie

Wymienione funkcje są integralną częścią systemu dla bibliotek cyfrowych Rosetta Ex Libris, który został wydany w styczniu 2009 roku. Tworzony przy współpracy z Biblioteką Narodową Nowej Zelandii, poddany przeglądowi światowej sławy ekspertów w dziedzinie przechowywania długoterminowego, Rosetta Ex Libris wychodzi naprzeciw potrzebom bibliotek oraz archiwów i wspomaga gromadzenie, zarządzanie i przechowywanie różnego rodzaju obiektów cyfrowych, w różnych formatach i strukturach.

Rosetta Ex Libris umożliwia instytucjom bibliotecznym zarządzanie zbiorami cyfrowymi w pełnym zakresie – od przyjęcia obiektu do jego rozpowszechnienia. Silnik aplikacji o charakterze regułowym (*rule-based*) oraz otwarta architektura systemu ułatwiają instytucjom używanie funkcji systemowych do tworzenia unikalnych wtyczek programowych oraz aplikacji usprawniających wprowadzanie, zarządzanie, przechowywanie i procesy dostarczania informacji w tym systemie.

Architektura wysokiego stopnia

Rosetta Ex Libris to rozwiązanie do przechowywania obiektów cyfrowych, zgodne ze standardem ISO OAIS oraz międzynarodowymi standardami, takimi jak: *Metadata Encoding and Transmission Standard (METS)*, *Preservation Metadata: Implementation Strategies (PREMIS)*, *Dublin Core* oraz *Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*. System został zaprojektowany tak, by wspierać pozyskiwanie, walidację, wprowadzanie danych, przechowywanie, zarządzanie i zabezpieczenie oraz rozprowadzanie różnego typu zasobów cyfrowych, zgodnie z wymogami dotyczącymi deponowania zbiorów w świetle e-prawa.

Rosetta Ex Libris składa się z zestawu oddzielnych, ale interaktywnych modułów, które odpowiadają cyklowi życia obiektu cyfrowego. Są to:

A. Depozyt

Moduł depozytowy umożliwia producentom zewnętrznym (wydawcom), jak i wewnętrznym (personel instytucji) zamieszczanie i zapisywanie wsadowe zbiorów informacji (*Submission Information Package – SIP*), które mają zostać wdrożone do repozytorium. Moduł ten współpracuje z różnymi typami przepływu pracy – automatycznym, półautomatycznym oraz ręcznym, które mogą być przypisane konkretnemu producentowi, w zależności od jego umowy z instytucją. Moduł depozytowy zapewnia producentom interfejs oparty na sieci Web, dzięki któremu mogą oni zamieszczać pliki,

metadane, jak i określać ograniczenia dostępu. Pliki i metadane mogą być również przesyłane na stronę FTP, z której moduł depozytowy będzie mógł je pobrać. Zestaw programisty (*Software Development Kit* – SDK) zapewnia bezproblemową integrację modułu z systemami archiwizacyjnymi i, dzięki temu, usprawnia proces tworzenia dużych depozytów.

B. Obszar Roboczy

Obszarem roboczym jest repozytorium, dzięki któremu można zarządzać wewnętrznymi procesami wdrażania informacji o materiałach wprowadzanych do systemu. Stanowi on miejsce, w którym wsadowe zbiory informacji (SIP) stają się archiwalnymi zbiorami informacji (*Archival Information Package* – AIP) w modelu OAIS.

Obszar roboczy wspiera zautomatyzowane oraz manualne procesy, takie jak:

- skanowanie wirusowe, sumy kontrolne, identyfikacja i walidacja formatu w celu zapewnienia integralności przepływu danych i metadanych SIP;
- procesy kontrolne i katalogujące, które monitorują jakość wprowadzanego materiału w celu weryfikacji kompletności i występowania ewentualnych uszkodzeń;
- usprawnienie procesów konwersji informacji do stabilnego formatu, gdy będzie to konieczne, oraz uzupełnienie obiektów o istotne metadane.

C. Stałe Repozytorium

Moduł stałego repozytorium to miejsce w systemie służące do przechowywania długoterminowego zatwierdzonych i uzupełnionych materiałów. Jego główną funkcją jest długotrwałe przechowywanie informacji. Zasoby tu przechowywane są już zapisane i są plikami tylko do odczytu oraz mają swoje kopie bezpieczeństwa.

D. Repozytorium Operacyjne

Oprócz repozytorium stałego, Rosetta oferuje również repozytorium operacyjne, które daje możliwość przeszukiwania, indeksowania i szybkiego dostępu do zbiorów. Tutaj właśnie odbywają się czynności związane z przechowywaniem długoterminowym.

E. Planowanie Przechowywania Długoterminowego

Według modelu referencyjnego OAIS, system przechowywania powinien zawierać moduły planowania procesu przechowywania. Rosetta posiada moduł planowania, który składa się z biblioteki formatów, narzędzi analizy

ryzyka, ewaluacji i czynności archiwistycznych. Moduł ten umożliwi instytucjom zarządzanie i przeprowadzanie całego procesu zabezpieczenia danych poprzez identyfikację ryzyka, wyboru najkorzystniejszego rozwiązania oraz test/aktywację wybranego procesu archiwizacji.

Wynikiem czynności przygotowawczych do przechowywania jest ponowne wprowadzenie zbiorów do repozytorium stałego i przechowywanie ich tam, jako nowych wersji źródłowych dóbr intelektualnych.

F. Administracja

W celu wsparcia systemu zarządzania moduł administracyjny zawiera aplikacje o następujących funkcjach:

- konfigurator systemu;
- raportowanie – przy użyciu systemu raportowania będącego otwartym oprogramowaniem – BIRT (*Business Intelligence and Reporting Tools*);
- zarządzanie przez użytkownika, w tym personel, patronów oraz producentów;
- monitoring prac konserwacyjnych i innych procesów, umożliwiających działanie systemu.

G. Dostęp

Udostępnione zbiory informacji (*Dissemination Information Package – DiP*) są tworzone w module dostępu. Zbiory DIP są używane do udostępnienia informacji przechowywanych w stałym repozytorium. Moduł ten składa się z dwóch elementów:

- komponent publikacyjny umożliwia systemom zewnętrznym (np. silnikom przeszukiwania oraz narzędziom odszukiwania zasobów, jak Primo® od Ex Libris) wykorzystanie standardowych protokołów komunikacyjnych, takich jak OAI-PMH, *Search/Retrieve via URL* (SRU) oraz *Search/Retrieve Web Service* (SRW) do uzyskania dostępu do zbiorów przechowywanych w stałym repozytorium;
- komponent dostawczy zapewnia dostęp do przechowywanych obiektów intelektualnych poprzez przeglądarki, które są zawarte w systemie (np. serwer do strumieniowania wideo) lub poprzez aplikacje zewnętrzne. System przetwarza prośbę o dostarczenie pliku i sprawdza uprawnienia dostępu do danych przed udostępnieniem ich użytkownikowi.

Bibliografia

1. Biała księga IDC, Gantz J. Reinsel D., *As the Economy Contracts, the Digital Universe Expands* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.emc.com/collateral/leadership/digital-universe/2009DU_final.pdf.
2. Biała księga IDC, Gantz J.F. et al., *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2011* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>.
3. Biała księga Planets, Sinclair P., *The Digital Divide. Assessing Organisations' Preparations for Digital Preservation* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.planets-project.eu/docs/reports/planets-market-survey-white-paper.pdf>.

Matthias Töwe

Politechnika Federalna w Zurychu

Ochrona cyfrowych zasobów w Politechnice Federalnej w Zurychu. Potrzeby naukowców i wizja Biblioteki Uniwersyteckiej ETH w Zurychu

Słowa kluczowe: archiwizacja danych naukowych, system Rosetta, Biblioteka Politechniki w Zurychu, kuratorstwo cyfrowe

Abstrakt: Biblioteka Główna Politechniki Federalnej w Zurychu realizuje projekt cyfrowej archiwizacji danych naukowych, administracyjnych oraz materiałów bibliotecznych. Uczelnia jest największą placówką oświatową w zakresie nauk technicznych i przyrodniczych w Szwajcarii. Celem artykułu jest opisanie sytuacji, która doprowadziła do uruchomienia omawianego projektu, oraz przedstawienie wybranych aspektów trwających nad nim prac. Jedną z najważniejszych opisywanych kwestii są wymogi dotyczące archiwizacji cyfrowej w środowisku akademickim, z akcentem na działania dotyczące danych naukowych. Dla wszystkich rodzajów danych centralnym punktem odniesienia w kontekście długotrwałego przechowywania jest model *OAIS Reference Model*¹, wykorzystywany przez aplikację *Rosetta (Ex Libris)* w połączeniu z lokalnie używanymi narzędziami do administrowania danymi.

Wstęp

Politechnika Federalna w Zurychu (*Eidgenössische Technische Hochschule Zürich*) jest największą uczelnią techniczną w kraju, kształcąca obecnie ponad 17 000 studentów, z których około 3700 odbywa studia doktoranckie. Ponad 450 profesorów pracuje tu w następujących dziedzinach: inżynieria, architektura, matematyka, nauki przyrodnicze, nauki społeczne i tzw. systemowe, marketing i zarządzanie.

Biblioteka ETH (*ETH-Bibliothek*) jest jednostką odpowiedzialną za pozyskiwanie i administrowanie informacjami dla Uczelni. Jako

¹ Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System (OAIS) – Recommended Practice (Magenta Book)* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://public.ccsds.org/publications/archive/650x0m2.pdf>; standard ISO OAIS nr 14721:2012 [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284.

główna biblioteka placówki z silnymi tradycjami naukowymi, Biblioteka ETH musiała stale się rozwijać, by dostosować swe usługi do zmieniających się potrzeb naukowców. Poczynając od wczesnych lat siedemdziesiątych XX wieku, ETH-Bibliothek prowadziła prace nad swoim własnym zintegrowanym systemem bibliotecznym i przed ostatnią dekadą XX wieku umożliwiała badaczom dostęp do zasobów w różnych formach, m.in. CD-ROM. Od połowy lat 90. XX wieku aż do chwili obecnej rozwijany jest dostęp online do pełnotekstowych czasopism elektronicznych, jako standardowej usługi. Do roku 2011 Biblioteka ETH udostępniła blisko 13 900 czasopism elektronicznych. Do nowości należy również udostępnienie studentom i pracownikom Uczelni 86 000 książek elektronicznych.

Doktoranci są zobowiązani do dostarczenia wersji elektronicznej swojej rozprawy doktorskiej, która jest publikowana w uczelnianym repozytorium ETH oraz bazie *ETH e-collection*² wraz z innymi dokumentami utworzonymi przez pracowników Politechniki. Ponadto stale rosnąca część zasobów obrazów i grafik Biblioteki ETH jest dostępna w formie elektronicznej. W 2011 roku liczba ta osiągnęła poziom 169 000 obrazów cyfrowych na platformie *e-pics*³, która udostępnia również inne, oddzielne kolekcje obrazów z różnych instytutów i jednostek organizacyjnych Politechniki. Biblioteka ETH korzysta z dwóch krajowych platform zdigitalizowanych treści: *e-rara.ch*⁴, tj. rzadkie książki wydrukowane w Szwajcarii z kolekcji innych szwajcarskich uniwersytetów, oraz *retro.seals.ch*⁵, która jest wspólną platformą online dla wszystkich szwajcarskich naukowych czasopism zdigitalizowanych. Obie te platformy zawierają łącznie prawie 6 milionów stron. Co więcej, właśnie rozpoczyna się trzeci wspólny projekt (*e-manuscripta*), który ma objąć manuskrypty.

Biblioteka ETH zajmuje się także archiwami uczelnianymi Politechniki. Archiwa o charakterze depozytów hybrydowych zawierają rosnącą liczbę obiektów cyfrowych. Wszystko odbywa się z pewnym opóźnieniem spowodowanym zatrzymaniem dokumentów w jednostkach administracyjnych Uczelni, ale spodziewać się można, że coraz większa liczba materiału cyfrowego trafi do archiwów ETH w celu długoterminowego przechowywania.

Biblioteka ETH dba zatem o naukowe i historyczne dziedzictwo Politechniki. Dlatego też biblioteka od dawna rozważa wszystkie prawdopodobne konsekwencje przechodzenia na platformy cyfrowe oraz całkowitą zmianę charakteru pracy jednostki w przyszłości na wyłącznie cyfrowy.

² E-collection. ETH Institutional Repository [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://e-collection.library.ethz.ch>.

³ E-pics. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <https://www.e-pics.ethz.ch>.

⁴ E-rara. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-rara.ch>.

⁵ Retrodigitised journals. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://retro.seals.ch>.

Co jasne, rosnąca zależność od danych cyfrowych dotyka każdej sfery życia, ale w szczególnym stopniu pracy naukowców, co wynika z procesu badawczego. Chociaż naukowcy z określonych dziedzin od dawna próbowali zarządzać swoimi danymi i osiągnęli w tym zakresie pewne sukcesy, niektórzy dopiero niedawno zaczęli dostrzegać ten problem, gdy zostali skonfrontowani z wymogami instytucji finansujących⁶. Zarówno zachowanie istotnych danych, jak i długotrwałe ich przechowywanie od dawna stanowią podstawy działań naukowych, które mają zapewnić możliwość odtworzenia wyników badań. Z perspektywy instytucjonalnej, przestrzeganie i wdrażanie tych zasad w życie jest gwarantem utrzymania wysokiej reputacji i standardów akademickich. Stąd opracowano wytyczne dobrego postępowania w dziedzinie badań dla Politechniki w Zurychu⁷. Jednak aż do chwili obecnej nie jest dostępne żadne odpowiednie narzędzie.

Wstępne kwestie dotyczące projektu

Wraz z powstaniem pierwszych czasopism elektronicznych narodziło się pytanie dotyczące wiarygodności źródeł. Czy można zagwarantować, że za kilka lat zasoby te będą wciąż dostępne? Czy będzie możliwy nieprzerwany dostęp do danych po upływie okresu licencji? O ile pierwsze pytanie odnosi się do każdego rodzaju danych cyfrowych, o tyle druga kwestia dotyczy w szczególności treści licencjonowanych. Nie można udzielić prostych odpowiedzi, zatem zagadnienia te były ciągle w spektrum zainteresowań Biblioteki ETH, jak i Konsorcjum Szwajcarskich Bibliotek Akademickich⁸. Konsorcjum koordynuje licencjonowanie i pozyskiwanie głównie, ale nie jedynie, materiałów elektronicznych dla bibliotek uniwersyteckich, a jej biuro mieści się w Bibliotece ETH. Jako że wszyscy członkowie Konsorcjum spotkali się z podobnymi wyzwaniami, około 2002 roku rozpoczęto pracę nad projektem pod nazwą *E-Archiwistyka*⁹, mającym na celu przechowywanie długoterminowe. Więk-

⁶ Na przykład informacje zebrane dla Wielkiej Brytanii przez DCC. Digital Curation Centre [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.dcc.ac.uk/resources/policy-and-legal>.

⁷ ETH Zürich Executive Board, *Guidelines for Research Integrity and Good Scientific Practice at the ETH Zurich* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.rechtssammlung.ethz.ch/pdf/414_Integrität_Forschung_engl.pdf; warto zauważyć, że wersja niemiecka była uaktualniona 25 października 2011 [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.rechtssammlung.ethz.ch/pdf/414_Integrität_Forschung.pdf.

⁸ Konsortium der Schweizer Hochschulbibliotheken. Koordinierte elektronische Informationsversorgung für Schweizer Hochschule [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://lib.consortium.ch>.

⁹ Konsortium der Schweizer Hochschulbibliotheken. Projekte [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://lib.consortium.ch/html_wrapper.php?src=earchiving-ea&dir=project&activeElement=2&ea=1%20.

szość rezultatów badań projektu została przedstawiona w postaci osobnej publikacji¹⁰. Projekt doprowadził do stworzenia platformy *retro.seals.ch* dla wcześniej wspomnianych szwajcarskich czasopism naukowych. Ponadto rozpoczęto projekt *E-Skład*¹¹, którego zadaniem jest zbadanie możliwości przechowywania treści licencjonowanych na serwerach lokalnych w celach archiwalnych. Inicjatywy te były kontynuowane w narodowym programie pod nazwą *e-lib.ch* (Szwajcarska Biblioteka Elektroniczna¹²) i wsparte wewnętrznym projektem ETH-Bibliotheken oraz serwisami informatycznymi Politechniki w Zurychu, które miały zawierać testy systemów zarządzania dokumentacją jako możliwych technicznie rozwiązań problemu. Te projekty koncentrowały się na treściach bibliotecznych. W 2008 roku Konsorcjum Szwajcarskich Bibliotek Akademickich, w imieniu Konferencji Szwajcarskich Bibliotek Akademickich, zleciło rozszerzenie badań o naukowe dane pierwotne¹³.

Wspomniane projekty i badania przyczyniły się w znaczącym stopniu do zrozumienia kwestii przechowywania długoterminowego zasobów cyfrowych i pomogły uwrażliwić szwajcarskie środowiska biblioteczne na ten problem. Efekt został pogłębiony przez doświadczenia Szwajcarskiej Biblioteki Narodowej¹⁴ i jej projekt *e-Helvetica*¹⁵, którego celem było znalezienie nowego rozwiązania dla przechowywania długoterminowego, głównie form natywnie cyfrowych, jak strony Web, w celu ochrony szwajcarskiego cyfrowego dziedzictwa kulturowego.

Ogólnym wnioskiem, wyciągniętym z tych inicjatyw, było stwierdzenie, że prędzej czy później biblioteki będą potrzebować rozwiązań spełniających funkcje modelu Open Archival Information System (OAIS). Wówczas nie było żadnych produktów na rynku, które mogłyby sprostać tym wymaganiom, a większość rozwiązań była opracowywana indywidualnie, choć w nie-

¹⁰ M. Töwe, A. Piguet, *Konzeptstudie E-Archiving*, (Version 1.2), Konsortium der Schweizer Hochschulbibliotheken, Zurych 2005, s. 293. Wersja niemiecka [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://dx.doi.org/10.3929/ethz-a-004990905>; wersja francuska (1.2.1) [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://lib.consortium.ch/external_files/Konzeptstudie_F_V1_2_1.pdf.

¹¹ E-Depot [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-lib.ch/en/Offers/E-Depot>.

¹² Swiss electronic library [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-lib.ch/en>.

¹³ P. Keller-Marxer, *Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz*, Ikeep, Berno 2008, s. 96, wersja niemiecka [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://dx.doi.org/10.3929/ethz-a-006070903>; wersja francuska: [online], [dostęp: 24.10.2012], dostępny w World Wide Web <http://dx.doi.org/10.3929/ethz-a-006070907>.

¹⁴ Swiss National Library [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.nb.admin.ch/index.html?lang=en>.

¹⁵ e-Helvetica. Collecting and archiving digital publications [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.nb.admin.ch/nb_professionnel/01693/index.html?lang=en.

licznych przypadkach, przy współpracy z dostawcami usług komercyjnych. ETH-Bibliothek nie miała możliwości rozwinięcia takiego systemu samodzielnie ze względu na niedostatek środków. Zdecydowano zatem, że prace nad bieżącymi projektami wewnętrznymi będą trwały, dopóki jakiś dostawca komercyjny nie zaoferuje takiego produktu. Poglądy te zostały poparte faktem, iż dane w Bibliotece ETH były rozsądnie zarządzane i użytkowane w sposób bieżący, co umożliwiało sprawne wykrycie obszarów mogących stwarzać problemy (obszary ryzyka).

Biblioteka Politechniki nie czekała jedynie na pojawienie się odpowiedniego rozwiązania na rynku. Jako jeden z filarów przyszłego systemu przechowywania długoterminowego, ETH-Bibliothek rozpoczęła projekt *DOI-Desk*¹⁶ stając się członkiem konsorcjum DataCite¹⁷. Projekt *DOI-Desk*, jako inicjatywa w ramach *e-lib.ch*, jest w stanie zaoferować szwajcarskim instytucjom regularną usługę rejestracji identyfikatorów DOI (*Digital Object Identifier* – Identyfikator Obiektów Cyfrowych).

Obecne prace nad kuratorstwem cyfrowym

Po upoważnieniu Biblioteki do objęcia pieczy nad zasobami cyfrowymi Politechniki Federalnej w Zurychu w roku 2010 ETH-Bibliothek stworzyła zespół odpowiedzialny za kuratorstwo cyfrowe. Szczególna uwaga i dbałość powinny zostać zapewnione danym z przeprowadzanych badań. Zasoby biblioteczne oraz akta z archiwów ETH również podlegają ochronie, ale są uznane za mniej zagrożone, przez co nabierają drugorzędного charakteru. tym samym celem projektu jest dostarczenie usług mogących zabezpieczyć oraz umożliwić przechowywanie długoterminowe danych naukowych, administracyjnych oraz bibliotecznych na Politechnice w Zurychu.

Zespół projektowy składa się z trzech etatów, wliczając jeden etat z zespołu serwisowego IT biblioteki. Inni członkowie mają zapewnić wkład naukowy i są wykwalifikowanymi pracownikami bibliotek lub archiwów.

Oczywiste jest, że w kontekście przechowywania, taki projekt może stanowić jedynie punkt wyjścia, a kolejną kwestią będzie posunięcie jego zamysłu naprzód w takim stopniu, by zarząd Politechniki zdecydował się na wspieranie tego typu usług w długim czasie.

Patrząc na tę sprawę z teoretycznego punktu widzenia, można zaobserwować podział obowiązków w tym obszarze pomiędzy ETH-Bibliothek za-

¹⁶ ETH Zürich DOI desk [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.doi.ethz.ch/index_e.html.

¹⁷ DataCite [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.datacite.org>.

rzządzającą usługami i aplikacjami a uniwersyteckim pionem IT zajmującym się obsługą serwerów, magazynowaniem plików i infrastrukturą sieciową. Wspólne zrozumienie wyzwań dotyczących przechowywania długoterminowego zostało zbudowane na podstawie współpracy w ramach wcześniejszego projektu pilotażowego.

Potrzeby naukowców

Biblioteka Uniwersytecka w Zurychu może pochwalić się bogatymi doświadczeniami w dziedzinie zarządzania jej własnymi danymi, lecz sprawą oczywistą jest fakt, że potrzeby naukowców w tym obszarze znacznie się różnią. W pierwszej fazie projektu w 2011 roku wszystkie grupy badawcze zostały poproszone o wypełnienie kwestionariuszy online. Dzięki wyteżonej pracy odpowiedzialnych za to pracowników, grupa respondentów osiągnęła poziom 80% z ówczesnych 450 grup badawczych. Ankieta składała się z dwóch części; pierwsza skupiła się na tworzeniu i użytkowaniu danych badawczych w ogólnym zarysie, natomiast druga dotyczyła szczególnych właściwości tych danych.

Nawet przed samym przeprowadzeniem ankiety nie było wątpliwości, że stworzenie precyzyjnej definicji danych badawczych będzie bardzo trudnym zadaniem. Podobnie wyglądała sprawa wspomnianych już wcześniej *danych pierwotnych* w przeciwieństwie do *danych wtórnych* (publikacji). W ujęciu pragmatycznym *dane badawcze* to wszelkie informacje, które naukowiec uzna za istotne dla jego badań i które będzie chciał zachować.

Warto wspomnieć tu o niektórych wynikach ankiety¹⁸. Jak można się było spodziewać, odpowiedzi z różnych wydziałów odbiegały od siebie w znacznym stopniu, ale odnotowano również różnice wewnątrz instytutów, co było spowodowane specyfiką badanej tematyki oraz stosowanych metod. Ogólnie rzecz ujmując, wyniki nie były sprzeczne z tymi, które pochodziły z podobnych badań¹⁹.

Zważywszy na fakt, że zachowanie danych jest dobrą praktyką naukową, nie powinno dziwić, że badacze często oznajmiali chęć zabezpieczenia danych ze swoich prac na określony przedział czasu (np. 10 lat), a nie „na zawsze”. Może jednak dziwić, że na tę potrzebę wskazało ponad 60% ankietowanych, podczas gdy właściwe przechowywanie długoterminowe zostało wspomniane przez ponad 30% respondentów (można było zaznaczyć wię-

¹⁸ Na podstawie rozmów z dr Susanne Scheid Jakobi, która przeprowadzała badania. Wyniki zostaną opublikowane.

¹⁹ T. Kuipers, J. van der Hoeven, *Insight into digital preservation of research output in Europe* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.

cej niż jedną odpowiedź). Dane te mają bezpośrednie przełożenie na ogólne pojęcie kuratorstwa cyfrowego w kontekście instytucjonalnym.

W zależności od stanu danych podlegających zachowaniu, ponad połowa ankietowanych wyraziła chęć przechowania danych, które były lub będą zamieszczone w recenzowanych publikacjach naukowych (można było zaznaczyć więcej niż jedną odpowiedź). Niespełna połowa respondentów chciałaby zachować nieprzetworzone dane, podczas gdy inne kategorie, takie jak standaryzowane lub ewaluowane statystycznie dane, zostały zaznaczone przez około 25% ankietowanych. Dane zbiorcze osiągnęły poziom nie większy niż 20%, a inne rodzaje danych zaznaczyło blisko 10% respondentów. Ponadto 25% osób nie było zainteresowane przechowywaniem danych z użyciem rozwiązań oferowanych przez bibliotekę.

Na podstawie dalszych dyskusji z naukowcami wyciągnięto wnioski, że chcieliby oni zachować kontrolę nad swoimi danymi. Ta konkluzja stoi w sprzeczności z obecnymi trendami, idącymi w kierunku otwartych danych, które nie są do końca powszechnie uznawane i popierane przez uczonych. Warto wspomnieć, że omawiany projekt Biblioteki Uniwersyteckiej ETH nie zajmuje się tzw. dużymi danymi. Najnowsze raporty międzynarodowe powinny mieć na względzie to rozróżnienie²⁰, jako że nie wszystkie pojęcia stosuje się do każdego typu danych w tym samym stopniu.

W ramach naszego projektu skupiliśmy się na stworzeniu zestawu usług danych dla indywidualnych naukowców. Pokrywa się to z wynikami ankiety, według których większość nagromadzonych danych pochodzi z indywidualnych projektów, głównie rozpraw doktorskich. Jest wiele powodów ostrożności naukowców. Z jednej strony, chcieliby być cytowani, gdy ich prace są przez kogoś wykorzystywane. Z drugiej jednak strony, naukowcy obawiają się, że ktoś posłuży się ich publikacją do osiągnięcia własnego sukcesu i dlatego preferują dwustronne kontakty z potencjalnymi badaczami, bazującymi na ich materiale. Problem ten dotyka również danych badawczych zawierających wszystkie istotne informacje niezbędne do ponownego użycia.

Ogólnie rzecz ujmując, jedynie część naukowców jest świadoma wyzwań, które stawia przed nami zarządzanie i przechowywanie danych. Niektórzy z nich podeszli do tego tematu z zaangażowaniem i osiągnęli już pewne sukcesy. Jednakże zadania tej natury nie powinny zaprzętać głów badaczy i zajęcie się nimi leży w interesie instytucji, które powinny dostarczać takie usługi naukowcom, by ich wysiłek naukowy był wspomagany skuteczną in-

²⁰ J. Wood et al. (High level Expert Group on Scientific Data), *Riding the wave. How Europe can gain from the rising tide of scientific data* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>; S. Reilly et al., *Report on integration of data and publications* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-exesummary_final.pdf.

frastrukturą. Mimo tego, każda dokumentacja musi zawierać informacje o temacie, którego dotyczy, o tym, w jaki sposób została stworzona i jakie mogą być ograniczenia odnośnie do jej wykorzystania. Wszystkie te informacje powinny pochodzić od osoby, która tę dokumentację sporządza.

Różnice pomiędzy typami danych

Długotrwałe przechowywanie zasobów cyfrowych ma za zadanie zabezpieczenie danych i umożliwienie ich ponownego użycia w przyszłości. Dla każdego rodzaju danych mówi się o trzech poziomach wymogów, które muszą być spełnione (rys. 1). Z teoretycznego punktu widzenia, potrzebnym, ale niewystarczającym warunkiem każdego sposobu przechowywania cyfrowego, jest zapewnienie ochrony strumienia bitów. Warunek ten gwarantuje utrzymanie materiałów cyfrowych w takiej formie, w jakiej były utworzone i przechowywane, unikając lub przynajmniej wykrywając rozpad bitów (*bit rot*), czyli nieumyślne uszkodzenie lub celową manipulację oraz wyjście z użycia napędów, mediów itd. Na tym poziomie odbywa się regularne wykrywanie usterek przy użyciu najbardziej zaawansowanych metod.

Co?	Dlaczego?	Kto?
Kuratorstwo danych	Zapewnić intelektualne możliwości ponownego użycia	Producenci danych
Ochrona treści	Zapewnić techniczne możliwości ponownego użycia	Biblioteka ETH
Ochrona strumienia bitów	Zapewnić stabilność techniczną	Pion IT ETH Zurich

Rysunek 1. Trzy poziomy ochrony danych; za: Jens Ludwig, Wissgrid (za: M. Töwe, ETH-Bibliothek)

Źródło: opracowanie własne.

Pomimo że informacje mogą być nienaruszone na poziomie kodu binarnego, nie ma w chwili obecnej żadnej metody na sprawdzenie, czy informacje zostały prawidłowo zakodowane w bitach. Tak może się stać, gdy pewna wersja oprogramowania lub systemu operacyjnego wychodzi z użytku, a dane były utworzone za pomocą właśnie tych starszych wersji. W chwili obecnej najbardziej obiecującym rozwiązaniem, mogącym sprostać temu

problemowi, jest migracja formatów, która jednak może spowodować poważne problemy z odtworzeniem pierwotnego zapisu danych.

Dwa zagadnienia opisane wcześniej odnoszą się do wszystkich rodzajów danych. Gdy obiekt cyfrowy jest stworzony daną metodą, perspektywy typowych obiektów bibliotecznych względem danych badawczych zaczynają nabierać innego charakteru: dokument tekstowy może „mówić za siebie”. Tak długo, jak tekst jest wyświetlony, możemy dojść do tego, o czym on jest, kto go napisał itd., bez dostępu do metadanych lub opisu. W przypadku pliku obrazu pojawiają się dodatkowe pytania: kiedy i gdzie zrobiono fotografię? czy kim jest osoba na niej przedstawiona? Odpowiedzi na tego typu pytania można zazwyczaj znaleźć w metadanych, które są regularnie tworzone w ustrukturyzowany sposób przez biblioteki i podobne instytucje.

W celu zrozumienia danych lub możliwościach przeanalizowania we właściwy sposób, niezbędne stają się złożone informacje, których nie można uzyskać wstecz. Stąd badacze, we własnym zakresie, powinni upewnić się, że do danych dostarczona jest odpowiednia dokumentacja, jeśli dane mają być użyte ponownie. Słowo „odpowiednia”, to w wielu przypadkach taka, która umożliwi samym autorom zrozumienie wyników ich pracy za kilka lat. Jeśli dane mają być użyte przez innych, wymogi są znacznie większe. Idealna sytuacja to taka, w której publikacja dostarcza wszelkich istotnych informacji. Jednakże w praktyce dla wielu dyscyplin i metod badawczych stanowi to duże wyzwanie. Wyzwanie, którym jest ochrona danych cyfrowych, czyli zapewnienie możliwości ponownego użycia oraz interpretacji danych, to powód, dla którego termin „kuratorstwo cyfrowe” jest bardziej odpowiedni przy opisywaniu naszym klientom, czym się zajmujemy, ponieważ do przechowywania danych w formie łatwej w użyciu i interpretacji potrzeba czegoś więcej niż tylko narzędzi technicznych.

Obecny projekt

Wcześniej opisano, jak kształtowało się podejście Biblioteki ETH względem kuratorstwa cyfrowego, uznając za swój cel wspieranie rozwiązań długoterminowego przechowywania dla wszystkich rodzajów danych. W fazie początkowej przedsięwzięcia sporządzono uaktualniony księgozbiór danych w bibliotece, by zidentyfikować dane, przy których należy rozpocząć prace w trybie priorytetowym. Co ciekawe, dane licencjonowane, na których projekt miał się oryginalnie koncentrować, zostały z niego wykluczone. Projekt E-Składu (*E-depot*) prowadzony przez Konsorcjum pokazał, że jest technicznie możliwe lokalne hostowanie danych, ale wykazał też, że wydatki będą tutaj czynnikiem hamującym, jeżeli jakość usług miałaby być porównywalna

z tą od wydawców. Co więcej, dane licencjonowane są użytkowane w wielu instytucjach na całym świecie, więc byłoby rozsądne wzięcie udziału w inicjatywach dzielenia się zasobami, takimi jak *LOCKSS*²¹ czy *Portico*²², które poczyniły znaczny postęp i nabrały trwałego charakteru. Oba te projekty są brane pod uwagę w kwestii ich wdrożenia w bibliotekach Szwajcarii. Dlatego właśnie projekt Biblioteki ETH koncentruje się na treściach tworzonych i używanych przez Politechnikę, stanowiących jej dziedzictwo kulturowe w wymiarze zarówno naukowym, jak i historycznym.

Nie ma znaczącej różnicy w postępowaniu z danymi pochodzącymi z prac badawczych lub innymi typami danych na poziomie ochrony kodu zero-jedynkowego (strumienia bitów) lub zapewnienia technicznych możliwości ponownego użycia danych. Bez wątplenia, dane badawcze występują częściej w mniej znanych formatach, ale podstawowe zasady monitorowania i migracji będą takie same dla wszystkich rodzajów obiektów cyfrowych. To jest właśnie powód, dla którego planujemy przechowywać wszystkie typy danych przy użyciu tej samej aplikacji, zgodnej z OAIS, wykorzystując serwery i infrastrukturę przechowywania danych uniwersyteckiego pionu IT. Jednakże nastąpią znaczące zmiany w sposobie, w jakim obiekty cyfrowe trafią do archiwów cyfrowych (rys. 2).

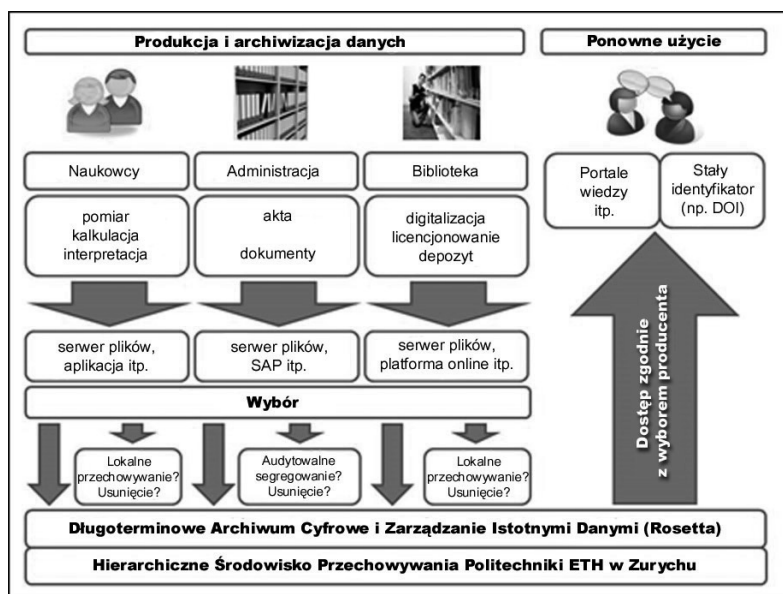
Oprogramowanie Rosetta Ex Libris, zgodne z OAIS, zostało wybrane jako narzędzie do przechowywania wszystkich rodzajów danych. Innymi kryteriami, które zdecydowały o dokonaniu takiego wyboru, były m.in. skalowalność, użycie istniejących narzędzi przechowywania długoterminowego czy możliwość nawiązania współpracy z Ex Libris w rozwoju *Rosetty*, będącej nowatorskim produktem. Jak już zostało wspomniane, wsparcie dostawcy było ważnym kryterium od samego początku. Ponadto system ten został współtworzony z projektem narodowej biblioteki Nowej Zelandii²³ o nazwie Narodowe Archiwum Dziedzictwa Cyfrowego²⁴, dlatego też było kwestią oczywistą, że wiele typowo bibliotecznych rozwiązań doczekało się akceptacji i wdrożenia.

²¹ LOCKSS. Lots of Copies Keep Stuff Safe [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.lockss.org>.

²² Portico [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.portico.org>.

²³ National Library of New Zealand [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.natlib.govt.nz>.

²⁴ About the Library [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.natlib.govt.nz/about-us/current-initiatives/ndha>.



Rysunek 2. Paralelny tryb pracy w cyfrowym rozwiązaniu długoterminowego przechowywania

Źródło: opracowanie własne.

Przechowywanie długoterminowe w zakładanej skali jest możliwe tylko wtedy, gdy większość procesów jest do pewnego stopnia zautomatyzowana. Warunkiem początkowym zautomatyzowanego wdrożenia obiektu do jakiegokolwiek cyfrowego archiwum jest dostęp do ustrukturyzowanych metadanych oraz istnienie źródła dla obiektu cyfrowego. Zazwyczaj wymóg ten jest spełniany przez dane, które już wtedy są pod kontrolą biblioteki. Tak właśnie stało się przy implementacji naszego projektu, kiedy to e-kolekcja Politechniki została wprowadzona do archiwum cyfrowego. Dane tego typu mogą być załadowane do oprogramowania *Rosetta* dzięki narzędziu o nazwie *Submission Application* (aplikacja wsadowa), które tworzy pakiety obiektów cyfrowych i odpowiadających im metadanych, po czym wprowadza je do *Rosetty*.

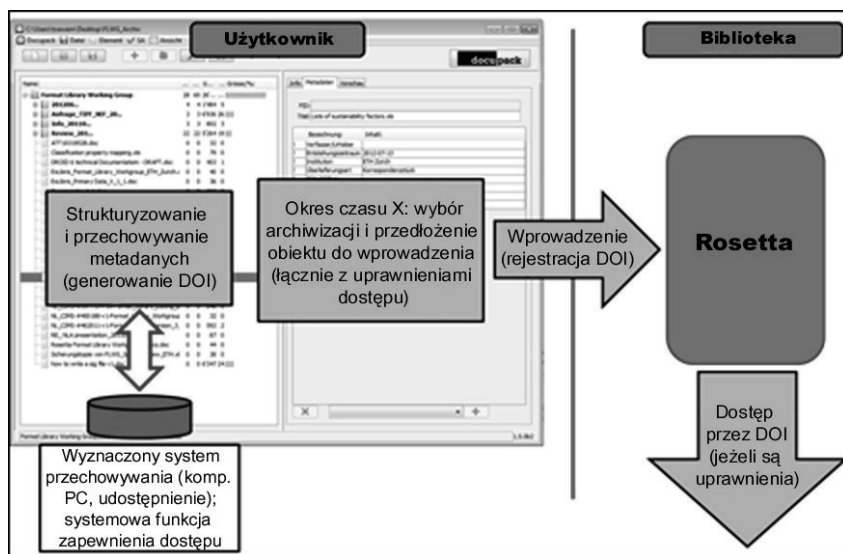
Akta administracyjne przechowywane w archiwach ETH zwykle nie zawierają metadanych od samego początku, lecz dopiero po tym, jak personel archiwum oceni i opíše treści. Z tego wynika, że cały proces od zgłoszenia dokumentu przez jednostki organizacyjne do archiwum aż do końcowego wprowadzenia dokumentu do długoterminowego archiwum cyfrowego podlega obsłudze i nadzorowi przez odpowiednie podmioty.

W przypadku danych badawczych sytuacja jest jeszcze bardziej skomplikowana. Istnieje niewiele ustrukturyzowanych platform lub przejrzystych struktur segregowania danych. Co więcej, nikt nie będzie w stanie dodać konstruktywnego opisu, po tym jak dane zostaną przekazane przez naukowców. Kwestie dotychczas przedstawione wskazały, że nawet rozwiązania zgodne z OAIS nie są w stanie zagwarantować spełnienia wszystkich wymogów w zakresie zarządzania danymi przed podjęciem decyzji o archiwizacji pewnych obiektów. Wymogi względem zarządzania danymi zostały skompilowane na podstawie dyskusji z wybranymi partnerami programu pilotażowego, którzy wyrazili potrzebę nabycia wiedzy i uprawnień do przetwarzania i przenoszenia danych w celu zebrania dokumentacji i metadanych przed stałą archiwizacją. Ma to zastosowanie w przypadku, gdy dane są jedynie materiałem uzupełniającym podczas przygotowań do wprowadzenia manuskryptu do systemu, jako możliwe zastosowanie przez jednego z partnerów programu pilotażowego. Po tym następuje proces *wykrywania luk* przez Ex Libris, by ustalić, czy funkcja ta powinna zostać dodana do oprogramowania *Rosetta* w przyszłości i które działania powinny być wykonywane przed wprowadzeniem do aplikacji. Głównymi wymogami były: przyjazność względem użytkownika oraz potrzeba niezaburzania wiarygodności archiwum długoterminowego przez wymaganą elastyczność w zarządzaniu danymi. Dlatego też kilka ulepszeń funkcjonalnych zostało zleconych twórcom programu *Rosetta*, podczas gdy inne zadania, takie jak prace przed wprowadzeniem do systemu, są wykonywane na prostych narzędziach, używanych lokalnie.

Tok pracy

Niniejszy opis odnosi się do trybu pracy zastosowanego w pierwszej wersji, pod koniec 2012 roku; ukazany został na rysunku 3. Tok pracy wykorzystuje aplikację środowiska Java o nazwie *Docupack*, która zajmuje się obsługą tworzenia zgłoszeniowych zbiorów informacji (*Submission Information Package* – SIP). W terminologii modelu OAIS, SIP-y to samodzielne pakiety obiektów cyfrowych wraz z metadanymi, które mogą być zgłoszone do systemu archiwizacji zgodnego z modelem OAIS w celu dalszego przetworzenia. W naszym zamyśle naukowcy rozpoczynają pracę od *Docupacka* w swoim własnym środowisku, by potem przeorganizować dane w strukturę hierarchiczną, odpowiednią dla ich potrzeb. Co więcej, dodają metadane na każdym poziomie strukturalnym. W celu ułatwienia tego procesu i uniknięcia powtarzających się ciągle czynności, minimalny zestaw metadanych może być skopiowany z wyższych poziomów hierarchii. Jeżeli wszyscy członkowie

zespołu mają wysłać swoje dane w ten sam sposób, warto skonfigurować wzory struktur dla poszczególnych grup badawczych. Dane strukturalne, jak i metadane są zapisywane w pliku METS²⁵, który stanowi medium do przenoszenia tych informacji. Już na tym poziomie można wygenerować pliki DOI jako proste identyfikatory obiektów cyfrowych. Jednakże identyfikatory te nie mogą być zarejestrowane na platformach międzynarodowych, ponieważ dane z narzędzi lokalnych nie mogą stać się publicznie dostępne, co często jest wymogiem przy próbie takiej rejestracji.



Rysunek 3. Przewidywany tok pracy dla danych badawczych przy użyciu *Docupacka* jako administratora pakietów

Źródło: opracowanie własne

Razem z obiektami cyfrowymi plik METS jest przechowywany pod kontrolą badacza, tj. w jego własnym środowisku przechowywania, które jest całkowicie niezależne od archiwum cyfrowego. Z ekonomicznego i instytucjonalnego punktu widzenia, za optymalną uznaje się sytuację, w której takie pakiety są przechowywane w centralnie zarządzanym środowisku magazynowym, lecz do chwili obecnej nie jest jasne, czy takie rozwiązanie byłoby dobrze przyjęte, jako że obecna praktyka w tym zakresie jest niejednorodna.

Tak długo, jak dane będą przechowywane lokalnie, naukowcy mogą je edytować i zamieniać zgodnie z wolą. Gdy zdecydują się wysłać treści do

²⁵ Metadata Encoding and Transmission Standard [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.loc.gov/standards/mets>.

archiwum przechowywania długoterminowego, kopia lokalna musi zostać „zamrożona”, a kopia przeznaczona do archiwizacji powinna być wysłana do narzędzia centralnego (*Documill*), które tworzy wersję ostateczną pakietu SIP i wprowadza ją do oprogramowania *Rosetta*. Z chwilą gdy obiekt zostaje bezpiecznie wdrożony i zarchiwizowany w archiwum długoterminowym, jego identyfikator DOI zostaje zarejestrowany automatycznie na platformie międzynarodowej, pod warunkiem że uprawnienia dostępu zostały odpowiednio skonfigurowane. Jeśli dostęp publiczny nie został przyznany, identyfikator DOI nie może być zarejestrowany. Jeśli DOI jest zarejestrowany, będzie służyć jako identyfikator do cytowania danych i bezpośredniego odwołania się do nich. Funkcja ta spełnia potrzeby wielu badaczy i można ją uznać za znaczącą wartość dodaną.

Podjęcie to pozwala na dużą elastyczność indywidualnych badaczy lub grup badawczych, jako że wiele właściwości lokalnego narzędzia może być skonfigurowanych w celu dopasowania do ich potrzeb, np.: konkretnych dziedzin metadanych, użycia ich jako opcjonalnych lub obowiązkowych, wartości domyślnych, wzorów poszczególnych struktur itd. Z perspektywy archiwum długoterminowego rezultat pracy narzędzia może być zdefiniowany wcześniej tak, aby najlepiej służył w roli dobrze ustrukturuwanego źródła dla procesu wprowadzania danych do baz. Dla obecnie znanych implementacji, dostęp będzie przyznany dzięki komunikacji identyfikatorów DOI. W innych przypadkach można również opublikować metadane w celu umożliwienia szerszego wyszukiwania. Dobrym rozwiązaniem jest publikacja takich metadanych na portalu wiedzy Biblioteki ETH²⁶ (*Primo Ex Libris*).

Porównując wymogi badaczy i archiwów uniwersyteckich ETH, odkryliśmy pewne podobieństwa w zakresie ich potrzeb. Nie powinno zatem być niespodzianką, że narzędzie *Docupack* było oryginalnie rozwijane dla administracji i archiwów. W naszym projekcie będzie używane do wprowadzania akt administracyjnych do naszego systemu, deponenci oddadzą akta personelowi archiwum, który dokona ewaluacji treści w *Docupacku* i wybierze te, które powinny zostać zachowane. Po tych czynnościach do głosu znów dochodzą deponenci, którzy mogą zaakceptować decyzje archiwum o usunięciu pewnych elementów albo odmówić. Następnie personel archiwum kończy opis i wdraża dane do cyfrowego archiwum przechowywania długoterminowego. Jednocześnie dokonywane jest wprowadzenie zasobów do *Rosetty*, a struktura hierarchiczna musi być wyeksportowana do Systemu Informacji Archiwalnej²⁷ (aplikacja: *CMI Star*), który przechowuje metadane

²⁶ ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.library.ethz.ch/en>.

²⁷ Archivdatenbank [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://archivdatenbank-online.ethz.ch>.

zarówno dla akt fizycznych, jak i cyfrowych. Stamtąd metadane, pozbawione szczegółowych informacji strukturalnych, są przekazywane do Portalu Wiedzy i w ten sposób zintegrowane z systemem wyszukiwania zasobów informacji. Dostęp do obiektów cyfrowych w oprogramowaniu *Rosetta* odbywa się poprzez identyfikatory DOI, które wyświetlają się zarówno w narzędziach wyszukiwania, jak i odkrywania.

Podsumowanie

Dotychczas otrzymane informacje zwrotne od naszych partnerów z programu pilotażowego oraz od badaczy są zachęcające. Z pewnością nastąpią pewne poprawki i procesy adaptacyjne, gdy tok pracy zostanie wdrożony w kilku grupach badawczych. Poważnym wyzwaniem dla bibliotek jest zdefiniowanie profilu swoich usług w tym obszarze: musi być to równowaga pomiędzy udostępnieniem narzędzi oprogramowania a ich rolą jako dostawców usług. Badacze nie mogą zostać zaopatrzeni w nowe narzędzia i pozostawieni samym sobie. Z biegiem czasu biblioteka musi przyjąć funkcję centrum kompetencji w zakresie przechowywania cyfrowego w ramach struktur uczelni i być w stanie udzielić porad przed, w trakcie i po tym, gdy dane zostaną utworzone. Informacje zwrotne od badaczy wykazują, że widzą oni bibliotekę w roli wsparcia dla ich wysiłków w zarządzaniu i przechowywaniu danych. Jednakże wielu z nich jest bardzo sceptycznie nastawiona wobec nowej polityki i przepisów (co może spowodować, że staną się oni mniej elastyczni) i potencjalnie przysporzyć prac administracyjnych. W obszarze czynności związanych z przechowywaniem, zgodnie z wcześniejszymi doświadczeniami, kwestie techniczne pozostają tu problemem numer jeden. Wiele uwagi należy poświęcić zagadnieniom dotyczącym wewnętrznego i zewnętrznego przepływu pracy w instytucjach archiwizujących. W najbliższych latach głównym problemem będzie zatrudnienie lub/i przeszkolenie personelu dysponującego pożądanymi umiejętnościami lub przynajmniej posiadającego zdolność poruszania się po wszystkich istotnych obszarach (badania, zarządzanie informacją w ujęciu potrzeb biblioteki i archiwum, długoterminowe przechowywanie cyfrowe itd.).

Pomimo intensywnych prac w ostatnich dwóch latach nad tokiem działań w zakresie późniejszych procesów związanych z przechowywaniem danych (analiza formatu, monitoring, migracje itd.), mamy wrażenie, że to tylko wierzchołek góry lodowej. Pewne fundamentalne zasady przechowywania cyfrowego są znane od jakiegoś czasu. Pierwszym z wymogów jest znajomość własnych danych, by umożliwić wykorzystanie ich w przyszłości. Uważna analiza danych, w celu wdrożenia zautomatyzowanych procesów

i narzędzi, jest w stanie rzucić światło na każdą, najmniejszą nieścisłość w danych i w ten sposób umożliwić poprawę ich jakości. Jednak w kontekście danych badawczych, nie jest już praktycznie możliwe podniesienie standardu danych lub metadanych, po zakończeniu pracy naukowca z danym materiałem. Naukowcy muszą zatem wziąć czynny udział w utrzymaniu jakości dostarczanych przez nich danych, nawet zanim trafią one do archiwum, a biblioteka musi im w tym pomóc.

Bibliografia

1. Archivdatenbank [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://archivdatenbank-online.ethz.ch>.
2. Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System (OAIS) – Recommended Practice (Magenta Book)* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://public.ccsds.org/publications/archive/650x0m2.pdf>; standard ISO OAIS nr 14721:2012 [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284.
3. DataCite [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.datacite.org>.
4. Digital Curation Centre [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.dcc.ac.uk/resources/policy-and-legal>.
5. E-collection. ETH Institutional Repository [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://e-collection.library.ethz.ch>.
6. E-Depot [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-lib.ch/en/Offers/E-Depot>.
7. E-Helvetica. Collecting and archiving digital publications [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.nb.admin.ch/nb_professionnel/01693/index.html?lang=en.
8. E-pics. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <https://www.e-pics.ethz.ch>.
9. E-rara. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-rara.ch>.
10. ETH Zürich DOI desk [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.doi.ethz.ch/index_e.html.
11. ETH Zurich Executive Board, *Guidelines for Research Integrity and Good Scientific Practice at the ETH Zurich* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.rechtssammlung.ethz.ch/pdf/414_Integrität_Forschung_engl.pdf; warto zauważyć, że wersja niemiecka była uaktualniona 25 października 2011 [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.rechtssammlung.ethz.ch/pdf/414_Integrität_Forschung.pdf.
12. ETH-Bibliothek [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.library.ethz.ch/en>.

13. Keller-Marxer P., *Konzeptstudie zur Entwicklung eines Modells für eine zentrale Langzeitarchivierung von digitalen Primär- und Sekundärdaten der Forschung für die Schweiz*, Ikeep, Berno 2008, s. 96; wersja niemiecka [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://dx.doi.org/10.3929/ethz-a-006070903>; wersja francuska: [online], [dostęp: 24.10.2012], dostępny w World Wide Web <http://dx.doi.org/10.3929/ethz-a-006070907>.
14. Konsortium der Schweizer Hochschulbibliotheken. Koordinierte elektronische Informationsversorgung für Schweizer Hochschule [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://lib.consortium.ch>.
15. Konsortium der Schweizer Hochschulbibliotheken. Projekte [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://lib.consortium.ch/html_wrapper.php?src=earchiving-ea&dir=project&activeElement=2&ea=1%20.
16. Kuipers T., van der Hoeven J., *Insight into digital preservation of research output in Europe* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.
17. LOCKSS. Lots of Copies Keep Stuff Safe [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.lockss.org>.
18. Metadata Encoding and Transmission Standard [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.loc.gov/standards/mets>.
19. National Library of New Zealand [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.natlib.govt.nz>.
20. Portico [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.portico.org>.
21. Reilly S. et al., *Report on integration of data and publications* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-exesummary_final.pdf.
22. Retrodigitised journals. ETH-Bibliothek [online], 2011 [dostęp: 24.10.2012], dostępny w World Wide Web: <http://retro.seals.ch>.
23. Swiss electronic library [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.e-lib.ch/en>.
24. Swiss National Library [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.nb.admin.ch/index.html?lang=en>.
25. Töwe M., Piguet A., *Konzeptstudie E-Archiving*, (Version 1.2), Konsortium der Schweizer Hochschulbibliotheken, Zurych 2005, s. 293; wersja niemiecka [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://dx.doi.org/10.3929/ethz-a-004990905>; wersja francuska (1.2.1) [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://lib.consortium.ch/external_files/Konzeptstudie_F_V1_2_1.pdf.
26. Wood J. et al. (High level Expert Group on Scientific Data), *Riding the wave. How Europe can gain from the rising tide of scientific data* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.

Anna Jakubiec, Marzena Pazdur

Biblioteka Politechniki Krakowskiej

Długoterminowa archiwizacja obiektów cyfrowych – międzynarodowe projekty

Słowa kluczowe: archiwizacja, CASPAR, PLANETS, PARSE.Insight, obiekty cyfrowe, międzynarodowe projekty

Abstrakt: rozwój zasobów cyfrowych, różnorodność sposobów dostępu i nośników, a także dynamika zmian w środowisku internetowym sprawiają, iż konieczne jest opracowanie programów, zasad i procedur zapewnienia dostępu do tych zasobów w przyszłości. Na świecie od wielu lat podejmowane są kroki w kierunku opracowania metod ochrony zasobów stanowiących cyfrowe dziedzictwo nauki i kultury. Tworzone są programy długoterminowej archiwizacji oraz realizowane testowe projekty archiwalnych systemów depozytowych. W artykule przedstawiono międzynarodowe projekty długoterminowej ochrony prowadzone w ramach szóstego i siódmego Programu Ramowego Unii Europejskiej. Omówiono główne cele i założenia tych projektów, strategię oraz metody ich realizacji.

Wstęp

Dynamiczny rozwój technologiczny skutkuje ciągłymi zmianami w sposobie zapisu oraz odczytu wszelkiego typu danych. Sytuacja ta niesie ze sobą poważne zagrożenie, polegające na coraz trudniejszym zagwarantowaniu ich użyteczności dla potrzeb obecnych i przyszłych potencjalnych odbiorców. Zapewnienie ich dostępności, autentyczności, spójności oraz poufności, a więc zrealizowanie celów długoterminowej archiwizacji, wymaga podjęcia długofalowej, przemyślanej i konsekwentnie realizowanej strategii. Instytucje nauki i kultury, świadome zagrożeń wynikających z ewentualnej utraty swoich danych przechowywanych w szeroko rozumianej „postaci cyfrowej”, często pod egidą instytucji międzynarodowych podejmują próby tworzenia programów długoterminowej archiwizacji oraz realizują projekty testujące systemy długofalowej ochrony informacji.

Międzynarodowe projekty trwałej archiwizacji zasobów cyfrowych

Parlament Europejski i Komisja Europejska utworzyły Europejską Agendę Cyfrową – plan działań UE na rzecz cyfrowego społeczeństwa i gospodarki – która w ramach tzw. Programów Ramowych wspiera, także finansowo, inicjatywy długoterminowej ochrony danych, jako priorytetowy obszar tematyczny: „Technologie informacyjne i komunikacyjne”. Programy Ramowe – po Wspólnej Polityce Rolnej i Funduszach Strukturalnych – zajmują obecnie trzecią pozycję w budżecie Unii Europejskiej. Szósty Program Ramowy Badań i Rozwoju Technicznego Unii Europejskiej (6PR) zamknął swój budżet kwotą około 20 mld euro. W kwietniu 2005 roku Komisja Europejska opublikowała dokument przedstawiający propozycje struktury, budżetu i zakresu tematycznego siódmego Programu Ramowego (7PR). Zadania w 7PR prowadzone będą przez okres siedmiu lat (2007–2013), a na ich realizację przewidziany jest budżet około 54 mld euro, czyli około trzy razy wyższy od budżetu 6PR. 7PR jest zdecydowanie podporządkowany realizacji Strategii Lizbońskiej oraz zastępującej ją strategii EU 2020 „Europa 2020 – Strategia na rzecz inteligentnego i zrównoważonego rozwoju sprzyjającego włączeniu społecznemu”, czyli budowie gospodarki opartej na wiedzy poprzez maksymalne wykorzystanie innowacyjności opartej na szeroko zakrojonych badaniach naukowych.

W ramach 6PR obejmującego lata 2001–2006 prowadzony był projekt CASPAR: *Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval* (Zabezpieczenie, dostęp i wyszukiwanie wiedzy kulturalnej, artystycznej i naukowej) – międzynarodowy zintegrowany projekt badawczy poświęcony zapisowi, udostępnianiu oraz wyszukiwaniu informacji w sferze nauki, kultury i sztuki. Przedsięwzięcie trwało od 1 kwietnia 2006 roku do 30 września 2009 roku. Ogólny koszt wyniósł ok. 16 mln euro, z czego 8,8 mln euro pokryła Unia Europejska. Prace nad projektem realizowane były przez konsorcjum zrzeszające naukowców, ekspertów kulturalnych i artystycznych, specjalistów w dziedzinie inżynierii wiedzy i ochrony informacji, firmy komercyjne oraz instytucje akademickie z pięciu państw europejskich: Czech, Francji, Grecji, Wielkiej Brytanii i Włoch oraz Izraela. W skład konsorcjum weszły następujące instytucje:

SZKOŁY WYŻSZE:

- University of Glasgow (Humanities Advanced Technology and Information Institute, HATII), Wielka Brytania;
- University of Leeds (Interdisciplinary Centre for Scientific Research in Music, ICSRiM), Wielka Brytania;

- Università degli Studi di Urbino (Istituto di Studi per la Tutela dei Beni Archivistici e Librari), Włochy.

INSTYTUCJE NAUKOWE:

- UNESCO (Organizacja Narodów Zjednoczonych ds. Oświaty, Nauki i Kultury);
- International Centre for Art and New Technologies [CIANT] (Międzynarodowe Centrum Sztuki i Nowych Technologii), Czechy;
- Institut National de l'Audiovisuel [INA] (Narodowy Instytut Audiowizualny), Francja;
- Institut de Recherche et Coordination Acoustique/Musique [IRCAM] (Instytut Koordynacji Badań Akustyczno-Muzycznych), Francja;
- Centre National de la Recherche Scientifique [CNRS] (Krajowe Centrum Badań Naukowych), Francja;
- Foundation for Research and Technology – Hellas [FORTH] (Fundacja na rzecz Badań i Technologii), Grecja;
- Science and Technology Facilities Council [STFC], Wielka Brytania;
- European Space Agency – ESRIN (Europejska Agencja Kosmiczna, Instytut Badawczy ESRIN), Włochy;
- Consiglio Nazionale delle Ricerche – Institute of Information Science and Technologies (Krajowa Rada ds. Badań Naukowych), Włochy.

FIRMY KOMERCYJNE

- Advanced Computer Systems ACS S.p.A., Włochy;
- Asemantics S.r.l, Włochy;
- Metaware S.p.A., Włochy;
- Engineering Ingegneria Informatica S.p.A., Włochy;
- IBM Haifa Research Laboratory, Izrael.

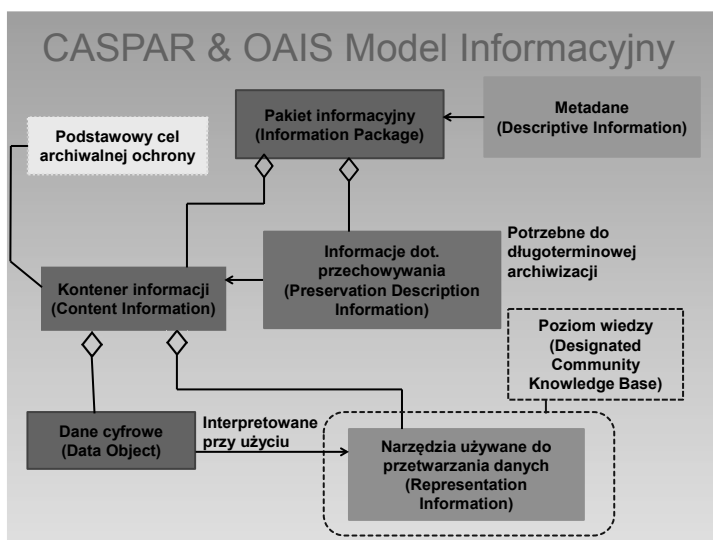
Głównym założeniem projektu było stworzenie, wdrożenie i weryfikacja systemu gwarantującego długoterminową archiwizację cyfrowych zasobów naukowych, kulturalnych i artystycznych. Kluczową kwestię stanowiło przechowywanie informacji i wiedzy. Po opracowaniu teoretycznych podstaw przyjętego rozwiązania zaplanowano przeprowadzenie praktycznych testów przyjętych metod i narzędzi, ze szczególnym uwzględnieniem zmieniającego się otoczenia (w tym sprzętu i oprogramowania), zmieniającej się grupy odbiorców (*Designated Communities*) oraz poziomu jej wiedzy (*Knowledge Base*). Głównym celem zatem było opracowanie systemu archiwizacji, wzbogaconego o interpretację treści (poza ochroną formatów i nośników, CASPAR kładzie nacisk na przechowywanie danych z minimalną utratą informacji, na zachowanie ich oryginalności i autentyczności).

Jako główne cechy charakteryzujące system można wymienić:

- uniwersalność i elastyczność, przydatność dla różnych typów instytucji, dla różnych rodzajów zasobów, współpraca z różnymi systemami oraz skalowalność, czyli możliwość zastosowania w warunkach zmieniającej się liczby użytkowników;
- niezależność – każdy element oprogramowania może działać samodzielnie;
- wymiennność – każdy element może być zastąpiony w ramach interoperacyjności, co ułatwia np. konserwację poszczególnych elementów;
- trwałość – bez względu na zmiany w sprzęcie, oprogramowaniu czy systemach operacyjnych, gwarantuje przydatność poprzez nieustanny rozwój;
- aktualność (przydatność) – poprzez ciągłe aktualizacje także w zakresie wszystkich zmian mających wpływ na interpretację danego obiektu.

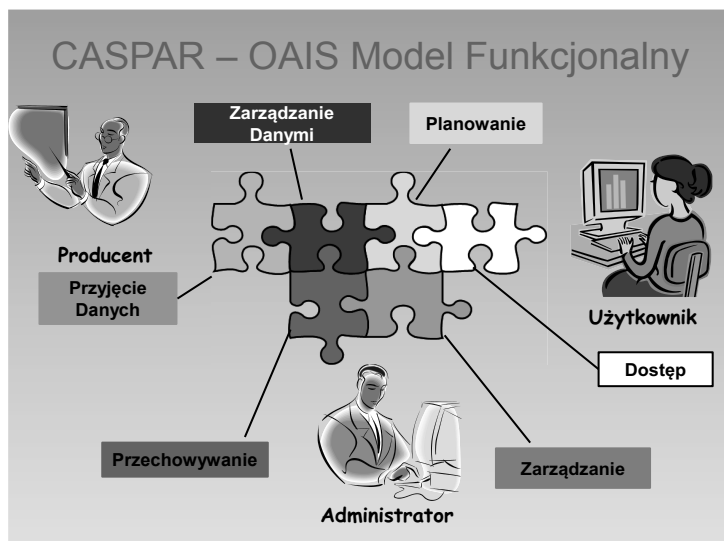
Warte podkreślenia są dwa ważne aspekty archiwizacji zasobów cyfrowych, na które CASPAR zwraca szczególną uwagę. Mianowicie proponuje wzbogacenie treści informacji o kontekst, w celu zapewnienia czytelności i zrozumiałości dla przyszłych pokoleń (określenie wynikających z tego metadanych dotyczących pochodzenia, otoczenia, przechowywania, zmian itp.) oraz proponuje archiwizację nie tylko samych „obiektów”, ale również archiwizację procesu, czyli informacji o tym, w jaki sposób powstał dany obiekt, postrzeganie obiektu jako procesu, a nie jednorazowego wydarzenia. Powinno to zapobiec sytuacji, w której odczyt jest możliwy, jednak interpretacja tego, co odczytano, jest niemożliwa. Może się tak zdarzyć, gdy nośnik przetrwa, ale informacja na nim zapisana jest w języku, który może być niezrozumiały w momencie odczytu. Prawidłowe zrozumienie dokumentu cyfrowego możliwe jest za pośrednictwem oprogramowania, za pomocą którego powstał. Jeśli takiego nie mamy, wówczas dokument pozostanie więźniem własnego kodu. Dlatego tak ważnym aspektem archiwizacji są zagadnienia dotyczące możliwości interpretacji danych.

Za teoretyczną podstawę funkcjonowania systemu przyjęto referencyjny model organizacji i funkcjonowania archiwów elektronicznych OAI (Open Archival Information System); model opisujący otwarty system informacji archiwalnej zarówno na poziomie technicznym, jak i organizacyjnym, definiowany jako archiwum złożone z osób i systemów, których celem jest długoterminowe przechowywanie, ochrona i udostępnianie informacji elektronicznych (rys. 1 i 2).



Rysunek 1. Model informacyjny CASPAR i OAIS

Źródło: L. Briguglio, *CASPAR Preservable Infrastructure Addressing Preservation with an OAIS based Infrastructure*, prezentacja przedstawiona podczas: 3rd Annual WePreserve Conference w Nicei.



Rysunek 2. Model funkcjonalny CASPAR i OAIS

Źródło: L. Briguglio, *op. cit.*

Na podstawie tych modeli zespół CASPAR stworzył infrastrukturę systemu o nazwie CASPAR FOUNDATION, w ramach którego określił 11 kluczowych elementów, podzielonych na 6 bloków, wspierających producentów i użytkowników na poszczególnych etapach archiwizacji. Oprogramowanie funkcjonuje na zasadach typowych dla narzędzi *Open Source*.

Ażeby zweryfikować opracowane rozwiązania, kolejnym etapem projektu CASPAR było powołanie stanowisk testowych. CASPAR-a testowano w zakresie „sztuki komputerowej”, archeologii oraz nauki o Ziemi. Według instytucji biorących udział w projekcie, jego potencjał i proponowane rozwiązania odegrają ważną rolę w przyszłości archiwizacji.

Projekt PLANETS (Zachowanie i długoterminowy dostęp poprzez usługi sieciowe) jest czteroletnim projektem naukowo-badawczym prowadzonym w ramach 6PR, obejmującego lata 2001–2006. Projekt ten prowadzony był w okresie od 1 czerwca 2006 roku do 31 maja 2010 roku. Ogólny koszt wyniósł ok. 14 mln euro, z czego 9 mln euro pokryła Unia Europejska.

Konsorcjum projektu zrzeszało ekspertów z europejskich bibliotek, archiwów, uniwersytetów oraz firm technologicznych. Koordynatorem zadania była British Library, ogółem w zadaniu udział wzięło 16 instytucji:

BIBLIOTEKI

- The British Library (Biblioteka Brytyjska);
- Koninklijke Bibliotheek (Biblioteka Królewska w Holandii);
- Österreichische Nationalbibliothek (Austriacka Biblioteka Narodowa);
- Det Kongelige Bibliotek (Biblioteka Królewska w Danii);
- Statsbiblioteket (Państwowa i Uniwersytecka Biblioteka w Aarhus, Dania).

ARCHIWA

- Nationaal Archief (Archiwum Narodowe Holandii);
- The National Archives of England, Wales and the United Kingdom (Archiwum Narodowe Wielkiej Brytanii);
- Schweizerische Eidgenossenschaft Confédération Suisse (Szwajcarskie Archiwum Federalne).

UNIWERSYTETY

- Universität zu Köln (Uniwersytet Koloński,), Niemcy;
- Albert-Ludwigs-Universität Freiburg (Uniwersytet Alfreda i Ludwika we Fryburgu Bryzgowijskim,), Niemcy;
- University of Glasgow, Humanities Advanced Technology and Information Institute (Instytut Informacji i Zaawansowanych Technologii Społecznych na Uniwersytecie w Glasgow), Wielka Brytania;

- Vienna University of Technology, The Department of Software Technology and Interactive Systems (Wydział Technologii Oprogramowania i Systemów Interaktywnych, Uniwersytet Techniczny w Wiedniu), Austria.

FIRMY TECHNOLOGICZNE

- The Austrian Institute of Technology [AIT] (Austriacki Instytut Technologiczny);
- IBM Netherlands (IBM Holandia);
- Microsoft Research Limited;
- Tessella Plc.

Główne cele projektu to:

- szczególnie nacisk na planowanie długoterminowego przechowywania zasobów;
- zbudowanie praktycznych usług i narzędzi do zapewnienia długoterminowego dostępu do cyfrowych danych kulturowych i naukowych, pozwalających semantycznie wzbogacić zasoby oraz automatycznie identyfikować formaty (migracje);
- opracowanie metodologii, narzędzi i usług potrzebnych do charakteryzacji obiektów cyfrowych;
- opracowanie innowacyjnych rozwiązań dla działań konserwatorskich;
- zapewnienie konsekwentnej i spójnej bazy danych dla obiektywnej oceny różnych protokołów, narzędzi, usług i kompletnych planów konserwacji;
- automatyzacja procesu opracowania i przechowywania w celu maksymalnego obniżenia kosztów;
- opracowanie środowiska testowego;
- stworzenie metod oceny stopnia strat danych cyfrowych.

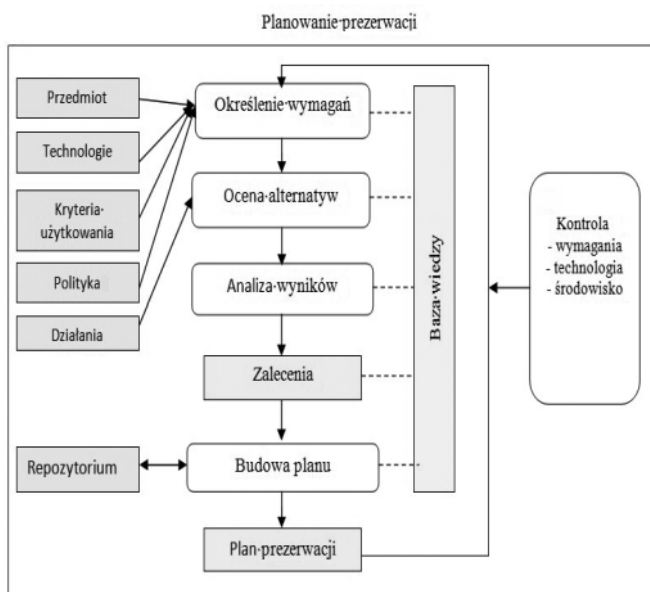
Praca w ramach projektu podzielona została na 6 odrębnych podprojektów: Preservation Planning, Preservation Action, Preservation Characterization, Testbed, Interoperability Framework oraz Dissemination and Training.

Architektura oprogramowania opiera się na zasadzie *click-and-install*, co spowodowane jest wymogiem polityki dużych przedsiębiorstw chcących zabezpieczyć dane na swoich serwerach, i małych przedsiębiorstw, które poszukują łatwo w użyciu programu. Daje on możliwość przetestowania usług i narzędzi przy użyciu rzeczywistych danych, w zabezpieczonym środowisku, tak aby umożliwić podjęcie decyzji w oparciu o naukowe dowody.

Uczestnicy projektu opracowali dwa nowe języki, oparte na języku XML (Extensible Markup Language):

1. XCEL (*eXtensible Characterization Extraction Language*) opisuje, w jaki sposób właściwości obiektu cyfrowego są kodowane i umożliwia ich automatyczne wyodrębnienie;
2. XCDL (*eXtensible Characterisation Description Language*) opisuje właściwości obiektu cyfrowego, takie jak kolor i głębia obrazu.

Ustanowienie polityki archiwizacji (planu konserwatorskiego) jest zasadniczym krokiem w długoterminowym zarządzaniu informacjami cyfrowymi. W dobie kryzysu i niejednokrotnie przy ograniczonym budżecie instytucje muszą dogłębnie zaplanować, jakie dane i dokumenty należy zachować. Plan prezerwacji uwzględnia zasady konserwowania, zobowiązania prawne, organizacyjne i techniczne ograniczenia, potrzeby użytkowników oraz cel przechowywania (rys. 3). Opisuje również kontekst zachowania i ocenia alternatywne strategie konserwatorskie. Planowanie prezerwacji powinno odbywać się w czterech etapach: Określenie wymagań (*Define requirements*), Ocena alternatyw (*Evaluate Alternatives*), Analiza wyników (*Analyze Results*) oraz Budowa Planu Prezerwacji (*Build a Preservation Plan*).



Rysunek 3. Planowanie prezerwacji

Źródło: B. van der Werf, *Plato. Planning digital preservation* [online], [dostęp: 03.09.2012], dostępny w World Wide Web: <http://www.bl.uk/blpac/digievents.html>.

W tym celu uczestnicy projektu PLANETS stworzyli narzędzie o nazwie *PLATO – the Preservation Planning Tool* (rys. 4). Zostało ono opracowane

przez Laboratorium Ochrony Zasobów Cyfrowych na Uniwersytecie Technicznym w Wiedniu i jest ogólnodostępnym narzędziem wspomagania podejmowania decyzji, które realizuje proces planowania archiwizacji, integruje usługi dla scharakteryzowania zawartości, realizuje działania konserwatorskie oraz automatycznego porównania obiektów, tak aby zapewnić maksymalne wsparcie dla planowania przedsięwzięć konserwatorskich. Aby zapewnić dostępność i autentyczność zasobów cyfrowych dla przyszłych pokoleń, plan musi brać pod uwagę ograniczenia techniczne i prawne, takie jak: miejsce przechowywania, infrastrukturę, zasady konserwowania, kwestię praw autorskich, koszty, potrzeby użytkowników oraz właściwości obiektów.

PLANETS do zachowania cyfrowych właściwości obiektów wykorzystuje dwa narzędzia: *JHOVE* i *DROID*. *JHOVE* (*JSTOR/Harvard Object Validation Environment*) służy do sprawdzenia, czy obiekt jest zgodny ze specyfikacją formatu, natomiast *DROID* (*Digital Record Object Identification*) służy do automatycznej identyfikacji typu obiektu.



Rysunek 4. PLATO – the Preservation Planning Tool

Źródło: Institute of Software Technology and Interactive Systems. Plato [online], [dostęp: 03.09.2012], dostępny w World Wide Web: <http://www.ifs.tuwien.ac.at/dp/plato/intro.html>.

18 maja 2010 roku uczestnicy projektu stworzyli i zdeponowali w byłej twierdzy wojskowej w pobliżu szwajcarskiego miasteczka Saanen kapsułę czasu o nazwie „Digital Genome” (Genom Cyfrowy). Plan zakłada przechowanie kapsuły przez 25 lat w schronie wystarczająco silnym, aby przetrwać atak nuklearny. Kapsuła czasu zawiera klucze do formatów danych lub cyfrowy odpowiednik kodu genetycznego różnych formatów danych (JPEG, Java, MOV, HTML oraz PDF) i ma na celu umożliwienie naukowcom w przyszłości dostępu i odczytu danych zapisanych przy użyciu archaicznych systemów pamięci masowej.

PARSE.Insight to dwuletni międzynarodowy projekt współfinansowany przez Unię Europejską w ramach 7PR. Prace nad projektem trwały od marca 2008 roku do czerwca 2010 roku. Ogólny koszt projektu wyniósł ok. 2 mln euro, z czego Unia Europejska pokryła 1 mln 300 tys. euro.

Konsorcjum składało się z pracowników bibliotek, uniwersytetów oraz instytucji naukowo-badawczych z pięciu państw europejskich:

BIBLIOTEKI

- Goettingen State and University Library [UGOE/SUB] (Państwowa i Uniwersytecka Biblioteka w Getyndze), Niemcy;
- National Library of the Netherlands [KB] (Biblioteka Narodowa Holandii);
- German National Library [DNB] (Niemiecka Biblioteka Narodowa).

INSTYTUCJE NAUKOWO-BADAWCZE

- Science and Technology Facilities Council [STFC], Wielka Brytania;
- Max Planck Society [MPG/MPDL] (Towarzystwo Maxa Plancka), Niemcy;
- European Organization for Nuclear Research CERN (Europejska Organizacja Badań Jądrowych CERN), Szwajcaria;
- European Space Agency [ESA/ESRIN] (Europejska Agencja Kosmiczna), Francja;
- International Association of Scientific, Technical & Medical Publishers [STM], Holandia.

UNIWERSYTET

- Fern Universität in Hagen [FUH] (Uniwersytet Eksternistyczny w Hagen), Niemcy.

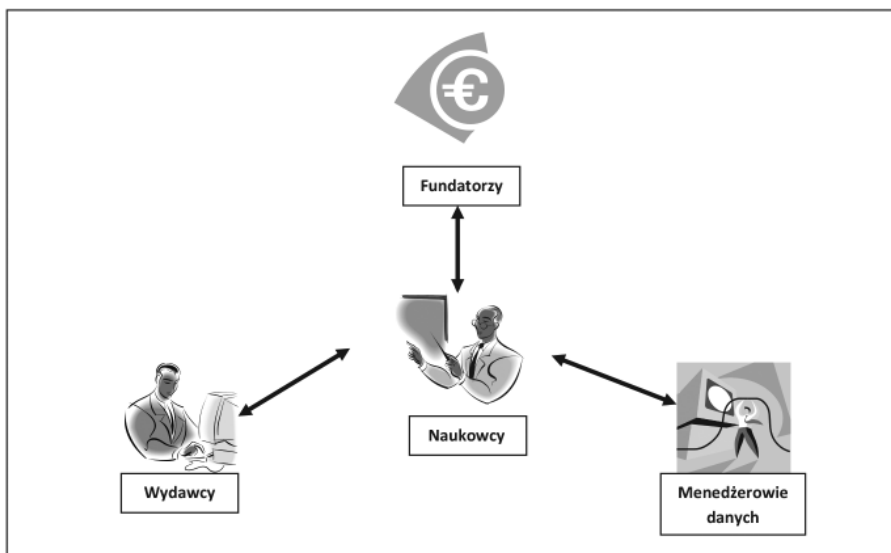
Projekt ten zasadniczo różni się od podobnych tym, że nie skupia się na technicznych rozwiązaniach archiwizacji cyfrowej, lecz na opracowaniu pewnego rodzaju „mapy” (*roadmap*) przeznaczonych dla infrastruktury e-Nauki (*e-Science*).

Podczas badań skupiono się na czterech interesariuszach (*stakeholders*): naukowcach, menedżerach danych, wydawcach oraz fundatorach. W centralnym punkcie znajdują się naukowcy, którzy dostarczają wydawcom i menedżerom danych „obiekty archiwizacji” (dorobek naukowy), finansowane przez fundatorów, a jednocześnie są użytkownikami tychże obiektów (rys. 5).

Wśród tej grupy zostały rozesłane ankiety, na które otrzymano ok. 2000 odpowiedzi, dotyczących praktyki, potrzeb i pomysłów związanych z doku-

mentami cyfrowymi oraz ich archiwizacją. Wszyscy interesariusze byli zgodni, że ochrona dorobku naukowego jest kwestią priorytetową.

Obecny stan archiwizacji dorobku naukowego w Europie jest zróżnicowany i wybiórczy. Istnieje dużo repozytoriów, jednak nie każda dyscyplina naukowa takie posiada, dodatkowo polityka gromadzenia i ochrony często jest opracowana zazwyczaj wewnątrz instytucji lub państwa, podczas gdy wszystkie państwa na świecie borykają się z tym problemem. Większość również skupia się na dostępie do danych niż na ich archiwizacji dla przyszłych pokoleń. Każda z grup zainteresowanych tym tematem zgodnie twierdzi, że infrastruktura ochrony danych naukowych jest jak najbardziej potrzebna, lecz role, jakie powinny one odgrywać w tym procesie, nie są do końca ustalone. Dlatego też pełna koordynacja działań oraz ustalenie, co dokładnie musi zostać zrobione, są konieczne.



Rysunek 5. Interesariusze biorący udział w badaniu

Źródło: T. Kuipers, J. van der Hoeven, *Insight into digital preservation of research output in Europe. Insight Report* [online], [dostęp: 03.09.2012], dostępny w World Wide Web: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.

Dla każdej grupy został zatem określony zakres odpowiedzialności w obsłudze danych naukowych.

Naukowcy, jako twórcy danych, powinni być odpowiedzialni za zarządzanie danymi w trakcie badań, stosowanie przepisów, udostępnianie danych w takiej formie, aby były możliwe do użycia przez innych naukowców. Natomiast jako użytkownicy danych odpowiedzialni są za przestrzeganie licencji

i wymogów użytkowania danych, poprawnego cytowania wykorzystanych prac itp.

Zarządcy danych, jako główni archiwiści, muszą przede wszystkim identyfikować, prowadzić selekcję, przechowywać, udostępniać dane oraz strzec praw autorskich.

Wydawcy są uważani za najważniejszych interesariuszy odpowiedzialnych za upowszechnianie rezultatów badań. Jednak obecnie tylko 12% z wszystkich publikowanych danych jest objęte polityką archiwizacji. Dlatego też w kontekście prezerwacji wydawcy są odpowiedzialni za: upowszechnianie danych, zarządzanie publikacjami w zakresie praw autorskich, wynagradzanie twórców i kuratorów danych, adaptowanie publikacji do cyfrowej archiwizacji.

Zadanie fundatorów natomiast polega m.in. na opracowaniu polityki współpracy z pozostałymi interesariuszami, monitorowaniu i egzekwowaniu polityki archiwizacji.

W wyniku przeprowadzonych ankiet i badań w środowisku bezpośrednio związanym i zainteresowanym ochroną i archiwizacją danych, wyciągnięto kilka wniosków odnoszących się do tworzonej tzw. mapy archiwizacji (*road-map*).

Długoterminowa archiwizacja jest bardzo często mylona z otwartym dostępem, lub nawet z udostępnianiem w ogólnym tego słowa znaczeniu. Dlatego też większy nacisk powinien zostać położony na archiwizację zbiorów niż na ich dostępność. Ważnym zadaniem jest również podniesienie świadomości na temat rozwoju polityki archiwizacji w różnych krajach, gdyż obecnie wysiłki w tym kierunku są fragmentaryczne i prowadzą do różnego rodzaju rozwiązań oraz sposobów realizacji, co, jak i kiedy należy archiwizować. Zatem kolejnym problemem, z którym trzeba się zmierzyć, jest integralność i stworzenie międzynarodowej współpracy związanej z długoterminową archiwizacją danych. Przyczyni się to zapewne do zwiększenia efektywności i szybkości w działaniu na tym polu. Uczestnicy projektu uważają, że organami, które powinny sprawować nad tym pieczę, są odpowiednie instytucje rządowe poszczególnych państw oraz Unia Europejska jako koordynator międzynarodowej współpracy. Prezerwacja zbiorów w poszczególnych państwach przebiega w różnym stopniu i z różnym zaangażowaniem, wiele krajów, zwłaszcza z Europy Wschodniej, nie posiada tak rozwiniętej techniki w tej dziedzinie, jak państwa Europy Zachodniej, dlatego też pożądana byłaby współpraca na tym polu między państwami, a także stworzenie wirtualnej platformy dla naukowców, w której mogliby uzyskać pomoc, w jaki sposób należy udostępniać i archiwizować własne badania naukowe. I tutaj także rolę koordynatorów miałyby spełniać odpowiednie instytucje rządowe poszczególnych państw oraz Unia Europejska. W celu

rozbudowania i wyrównania prac nad archiwizacją powinna również zostać podjęta współpraca z państwami obu Ameryk, Azji i Australii, oraz przydzielony odpowiedni budżet na tego typu działalność.

Zauważalny jest też całkowity brak, lub tylko znikoma liczba repozytoriów czy prac, nad archiwizacją danych w niektórych dziedzinach naukowych. Jednostki rządowe, instytucje naukowe i techniczne powinny wspierać oraz zachęcać do tworzenia repozytoriów i zamieszczania w nich raportów badań. Dobrym przykładem w tym zakresie są Niemcy, gdzie archiwizowanie własnych dokumentów jest bardzo promowane, a najaktywniejsi są nagradzani. Aby skłonić „producentów danych” do zamieszczania swoich prac w repozytoriach i zainteresowania ich długoterminową ochroną, powinny zostać stworzone odpowiednie narzędzia łatwe w użyciu, które zachęcą do takiej działalności, powinny być także prowadzone szkolenia, udzielana pomoc przez odpowiednie jednostki oraz należałoby skupić się na rozbudowie e-learningowych platform z zestawem danych, które w przystępny sposób pokażą, jak archiwizować i udostępniać materiały.

Wśród naukowców istnieje niechęć do deponowania swoich prac w repozytoriach, gdyż obawiają się oni nadużyć swoich badań, dlatego ważne jest ustanowienie ogólnych przepisów dotyczących praw autorskich, stworzenie odpowiednich narzędzi chroniących przed niepożądanym wykorzystaniem czy kopiowaniem.

Producenci pamięci masowych przy współpracy z odpowiednimi instytucjami rządowymi i naukowymi, powinni stworzyć ograniczoną liczbę formatów do przechowywania danych.

Jedną z najważniejszych kwestii jest nieustający przyrost danych, które niekoniecznie muszą być przechowywane, dlatego też uczestnicy projektu uważają za słuszne stworzenie specjalnych paneli z komisją ekspercką z każdej dziedziny, która byłaby organem doradczym w sprawach selekcji zbiorów przeznaczonych do prezerwacji. Panele takie musiałyby oczywiście mieć odpowiednie wsparcie w instytucjach rządowych, wśród menedżerów danych, a także dysponować odpowiednim budżetem na swoją działalność.

Utrata danych zapisanych w formie cyfrowej jest zjawiskiem powszechnym. Obecnie szacuje się, że na całym świecie jest przechowywany ok. 1 trylion płyt CD z ważnymi danymi i wiele z nich nie będzie dostępnych dla przyszłych technologii, tak jak to miało miejsce w przypadku starych 5,25-calowych dyskietek z 1970 roku. Obecnie na jednego mieszkańca ziemi przypada 100 GB danych. Unia Europejska traci informacje cyfrowe o wartości ponad 3 mld euro każdego roku, ze względu na coraz krótszy czas

eksploatacji technologii cyfrowych formatów, których żywotność wynosi od 5 do 7 lat, oraz urządzeń do przechowywania danych, takich jak płyty CD i DVD, których trwałość wynosi ok. 20 lat. Tylko nieliczne rządy i instytucje zdają sobie sprawę z ważności długofalowego zabezpieczania posiadanego dorobku naukowego, kulturowego czy ekonomicznego i próbują zabezpieczać własne dane m.in. poprzez udział w projektach zajmujących się długoterminową archiwizacją. Na świecie prowadzi się wiele takich inicjatyw, bowiem praca nad tym zagadnieniem wymaga nieustannych badań. Niepokoi znikomy udział polskich instytucji w tego typu przedsięwzięciach. Brak odpowiednich procedur, zabezpieczeń i konsekwentnej polityki długoterminowej ochrony danych cyfrowych spowoduje, że w niebezpieczeństwie znajdzie się szereg naukowych baz danych, ważnych dokumentów społecznych, rządowych, wojskowych, w końcu wiele dokumentów firm komercyjnych, których tylko znikomy odsetek posiada strategię nietrwałych przecież archiwów cyfrowych. Taka ignorancja może doprowadzić naszą cywilizację do swoistej choroby Alzheimera, do utraty historycznej, społecznej i kulturalnej pamięci.

Bibliografia

1. CASPAR – Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval [online], [dostęp: 03.09.2012], dostępny w World Wide Web: <http://www.casparpreserves.eu>.
2. Edwards L., *'Digital genome' time capsule stored under the Swiss Alps*, [online], [dostęp: 02.09.2012], dostępny w World Wide Web: <http://phys.org/news193639360.html>.
3. Gałęzowski G., *Trwałość dokumentów na tle nośników tradycyjnych* [online], [dostęp: 02.09.2012], dostępny w World Wide Web: http://www.archiwa.gov.pl/images/stories/Wydawnictwa/AP_Lublin_G_Galezowski.pdf.
4. Januszko-Szakiel A., *Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych*, „Przegląd biblioteczny” 2005, z. 3, s. 341–359.
5. Kuipers T., van der Hoeven J., *Insight into digital preservation of research output in Europe. Insight Report* [online], [dostęp: 02.09.2012], dostępny w World Wide Web: <http://www.parse-insight.eu/publications.php#d3-6>.
6. PARSE.insight Project [online], [dostęp: 02.09.2012], dostępny w World Wide Web: <http://www.parse-insight.eu>.
7. Planets [online], [dostęp: 02.09.2012], dostępny w World Wide Web: <http://www.planets-project.eu>.
8. Prom Ch., *PLATO (Digital Preservation Planning). Software Review* [online], [dostęp: 02.09.2012], dostępny w World Wide Web: <http://e-records.chrisprom.com/plato-digital-preservation-planning-software-review>.

Justyna Adamus-Kowalska

Uniwersytet Śląski w Katowicach

Zarządzanie dokumentacją elektroniczną i zabezpieczanie materiałów archiwalnych dla przyszłych pokoleń według zaleceń specyfikacji MoReq2 Komisji Europejskiej

Słowa kluczowe: przechowywanie dokumentów elektronicznych, specyfikacja MoReq2, działalność administracyjna, zarządzanie dokumentacją

Abstrakt: zasoby cyfrowe polskiej nauki i kultury tworzone są każdego dnia. Działalność naukowa, kulturalna, a także administracyjna, gospodarcza czy społeczna jest dokumentowana i utrwalana na piśmie. Dokumentacja coraz częściej przybiera postać elektroniczną, co świadczy o postępie i nieustannym rozwoju techniki. Wynikiem tak rozumianego rozwoju i postępu jest dokumentowanie wszelkich sfer życia ludzkiego za pomocą urządzeń teleinformatycznych. Przedmiotem artykułu jest problem zabezpieczenia, opracowania i przechowywania oraz udostępniania dokumentów elektronicznych powstających jako efekt pracy administracyjnej na różnych szczeblach zarządzania przy wykorzystaniu międzynarodowej specyfikacji MoReq2. Specyfikacja określa, w jaki sposób można sprawnie zarządzać dokumentacją, która powstaje w organizacjach w sektorze publicznym, jak i w przedsiębiorstwach prywatnych. Sprawne zarządzanie dokumentacją jest także gwarantem właściwej archiwizacji i zachowania dokumentów dla przyszłych pokoleń.

Wstęp

Praca administracyjna polega m.in. na gromadzeniu i przetwarzaniu informacji, co w konsekwencji prowadzi do wydawania odpowiednich decyzji administracyjnych. Integralnym elementem pracy administracyjnej jest także zarządzanie dokumentacją. Wiarygodność każdego postępowania administracyjnego zapewniają dokumenty w danej sprawie gromadzone i przechowywane w sposób odpowiadający przepisom prawa oraz obowiązującym normom. W pracy administracyjnej przyjmowane i stosowane są także odpowiednie praktyki zarządzania dokumentacją, nieuregulowane prawnie, a mające wpływ na jakość pracy administracyjnej.

Obok szeregu postępowania administracyjnych, w których korzysta się z dokumentacji, innym istotnym, choć często niedostatecznie wy-

eksponowanym aspektem pracy administracyjnej, w którym dokumentacja odgrywa znaczącą rolę, jest obrazowanie funkcjonowania państwa, społeczeństwa i gospodarki w danym okresie historycznym, na określonym obszarze i w określonym systemie społeczno-polityczno-gospodarczym. Dzięki zgromadzonej w urzędach i innych podmiotach dokumentacji, składowanej ostatecznie w archiwach, można prowadzić badania historyczne nad funkcjonowaniem państwa i społeczeństwa. Archiwa to niezastąpione skarbnice wiedzy, które sprawują pieczę nad dokumentami po ustaniu ich bieżącej użyteczności. Odpowiedzialność za zabezpieczenie tego dobra kulturalnego, jakim są dokumenty o wartości historycznej, spoczywa w pierwszej kolejności na jego wytwórcy, a następnie na archiwach państwowych, które z mocy prawa są zobligowane do wieczystego przechowywania materiałów archiwalnych¹. W tym miejscu jednak zwrócona będzie szczególna uwaga na początek życia każdego dokumentu, jaki ma miejsce w kancelarii oraz innych komórkach organizacyjnych zajmujących się tworzeniem dokumentów w danej sprawie. Problematyka ta ma niebagatelne znaczenie w kontekście stosowania dokumentów elektronicznych, dla których sam etap tworzenia czy wprowadzania do systemu elektronicznego staje się podstawą stosowania nowych metod i narzędzi, z uwzględnieniem możliwości przetwarzania informacji, jakich nie dawały żadne z dotychczas stosowanych form dokumentów. Ponadto zastosowanie dokumentu elektronicznego przynosi wymierne korzyści finansowe, umożliwia oszczędność czasu i nakładów pracy oraz przyczynia się do budowy społeczeństwa informacyjnego.

Problematyka zabezpieczania materiałów archiwalnych jest szczególnie istotna w kontekście współczesnych form komunikacji i w tym ujęciu często pojawia się potrzeba sięgania do źródeł wskazujących, w jaki sposób sprawnie i skutecznie zarządzać dokumentacją w postaci elektronicznej tak, aby z jednej strony sprostać wyzwaniom społeczeństwa informacyjnego, a z drugiej, spełnić szereg procedur i być w zgodzie z przepisami prawnymi, które określają sposób postępowania z dokumentami. Jednym z najbardziej kompletnych i wieloaspektowych opracowań w zakresie zarządzania dokumentacją elektroniczną jest *Model wymagań dla zarządzania dokumentami elektronicznymi (Model Requirements for the Management of Electronic Records, 2008)*, w skrócie MoReq, opracowany po raz pierwszy w roku 2001 na zlecenie Komisji Europejskiej. Specyfikacja została uzupełniona i opracowana w formie nowej wersji modelu wymagań o nazwie MoReq2². Spe-

¹ Obowiązek ten wynika z zapisów art. 3 ustawy z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach (Dz. U. 1983 Nr 38 poz. 173).

² *Specyfikacja MoReq2. Wzorcowe wymagania dotyczące zarządzania dokumentami elektronicznymi. Aktualizacja i rozszerzenie* [online], [dostęp: 11.09.2012], dostępny w World Wide Web: http://www.archiwa.gov.pl/images/stories/moreq2translation%20iv%20Polish%20versiob%20beta12_10_2011.pdf.

cyfikacja jest bardzo ogólna, nie odnosi się do specjalistycznej platformy sprzętowej ani też nie podaje gotowych przykładów rozwiązania problemów elektronicznego zarządzania dokumentami, jednak jej modelowy charakter sprawia, że jest użyteczna w praktyce. Niewiele jest takich opracowań i zaleceń, które wskazywałyby na sposób postępowania ze współczesną dokumentacją elektroniczną. Ponadto określenie sposobu postępowania z dokumentami, już w momencie ich powstania, jest wsparciem dla kolejnych etapów życia dokumentu, w tym także dla wieczystego przechowywania i udostępniania w archiwach historycznych.

Założenia MoReq2

Specyfikacja MoReq2 zawiera zestaw wymagań niezbędnych dla systemów komputerowych wspierających zarządzanie aktami elektronicznymi, dlatego może okazać się pomocna w pracach nad informatyzacją i wdrażaniem systemów elektronicznego zarządzania dokumentami (EZD). Specyfikacja odnosi się do dokumentów, które nie są ustrukturalizowane, czyli nie mają z góry określonej formy i treści, np. poprzez formularze czy pliki do wypełnienia. Z kolei dokumentacja ustrukturalizowana jest zazwyczaj przedmiotem gromadzenia, przechowywana i przetwarzana w specjalistycznych systemach dziedzinowych, zwanych także systemami merytorycznymi, w których dane są odpowiednio ustrukturalizowane i zarządzanie dokumentacją jest znacznie prostsze. Zestaw wymagań dla systemów służy tworzeniu nowych systemów, dla których dopiero projektowane są zasady funkcjonowania. Zalecenia MoReq2 zapewniają zachowanie dokumentów w długim czasie. Odbiorcami specyfikacji są zarówno podmioty sektora administracji państwowej, jak też sektora przedsiębiorstw prywatnych, gdzie zarządza się dokumentacją w postaci elektronicznej.

Natura i specyfika dokumentu elektronicznego

Dokument rozumiany jest jako informacja wytworzona, otrzymana i uchwycona jako świadectwo przez organizację lub osobę, jako wykonywanie prawnych zobowiązań lub w toku prowadzonych spraw³. Dokument składa się z jednego lub wielu pism, może także zawierać załączniki, może być utrwalony na jakimkolwiek nośniku i w jakimkolwiek formacie. Każdy dokument w systemie zarządzania dokumentacją musi być chroniony przed zmianami. Zasadniczą

³ *Ibid.*, s. 25.

cechą w definiowanym systemie jest grupowanie dokumentów w akta spraw, tj. według wyrozumowanych, opartych na logice podziałach dokumentacji. W specyfikacji mowa jest także o tym, iż każda sprawa, prowadzona w systemie zarządzania dokumentacją elektroniczną, powinna otrzymać określoną nazwę (tytuł sprawy) obejmującą rzeczowe cechy dokumentów.

Natura dokumentu elektronicznego pozwala na wyszczególnienie zbioru elementów, jakie muszą składać się na każdy zapis uznany za dokument:

- zawartość,
- struktura,
- kontekst,
- sposób przedstawienia.

Istotne jest zatem dołączenie do dokumentu metadanych⁴, które opisują kontekst powstania dokumentu, albowiem w zarządzaniu dokumentacją wszystkie cechy związane z powstaniem i obiegiem dokumentów mają wpływ na przebieg załatwiania spraw.

W systemie zarządzania dokumentacją elektroniczną opis kontekstu następuje w procesie klasyfikacji i kwalifikacji dokumentów na podstawie odpowiedniego wykazu akt będącego podstawą zarządzania dokumentami.

Inne wymaganie określone w MoReq2 i powtórzone za normą ISO 15489 wskazuje, że dokument powinien cechować się:

- autentycznością,
- wiarygodnością,
- integralnością,
- użytecznością.

Celem systemu zarządzania dokumentami, zgodnie ze specyfikacją MoReq2, jest zapewnienie, aby dokumentacja obejmowała następujące wartości⁵:

- była wiarygodna i przedstawiała fakty w sposób niezmienny,
- została wytworzona lub wysłana przez osobę podającą się za twórcę lub nadawcę dokumentu,
- została wytworzona lub wysłana w czasie, który podano,
- umożliwiała zweryfikowanie autentyczności,
- umożliwiała zweryfikowanie, przez kogo dokument został wytworzony lub wysłany i czy jest to osoba podająca się za jego twórcę lub nadawcę,
- umożliwiała udowodnienie, że została wytworzona lub wysłana w czasie, który podano,

⁴ Metadane w kontekście zarządzania dokumentacją definiowane są jako dane opisujące kontekst, zawartość i strukturę dokumentów i dzieje zarządzania nimi. Źródło: PN-ISO 15489-1:2006. Informacja i dokumentacja. Zarządzanie dokumentami. Część 1: Zasady ogólne, 28 kwietnia 2006.

⁵ *Specyfikacja...*, *op. cit.*, s. 27–28.

- była wiarygodna i w pełni odwzorowywała postępowania, działania lub fakty, które poświadczą,
- zawierała wszystkie dane i nie podlegała zmianom,
- można tę dokumentację zlokalizować, mieć do niej wgląd i zinterpretować.

Specyfikacja zawiera zagadnienia bezpieczeństwa dokumentacji, które regulowane są poprzez określenie kontroli dostępu do dokumentów. Kontrola ta powinna być przedmiotem polityki dostępu do dokumentów realizowanej w danej organizacji. Każda osoba pracująca w oparciu o dokumenty musi mieć zapewniony dostęp zgodnie z obraną polityką. W systemie zarządzania dokumentami zdefiniowane są określone role, jakie spełniają poszczególne osoby czy grupy osób, i w ten sposób nadawane są określone uprawnienia w systemie. Użytkownicy w systemie są identyfikowani przez odpowiedni login i hasło, stąd dostępność do dokumentu i odpowiednich działań w systemie jest zróżnicowana dla poszczególnych osób mających dostęp do systemu, a także możliwy jest zapis, odczytanie i zapamiętanie wszystkich działań na dokumentach ze wskazaniem osób, które ich dokonywały w trakcie obiegu dokumentu. Administratorzy systemu zarządzania dokumentami w ramach ochrony bezpieczeństwa definiują uprawnienia w dostępie do dokumentów, spraw, jednostek, klas i metadanych dla określonych użytkowników.

Układ dokumentów w systemie EZD

System elektronicznego zarządzania dokumentacją (EZD), budowany zgodnie ze specyfikacją MoReq2, działa w oparciu o określony układ dokumentów. Podobnie jak ma to miejsce w przypadku gromadzenia i zarządzania dokumentacją papierową, dokumenty elektroniczne grupowane są w elektroniczne jednostki i teczki. Grupowanie jest możliwe dzięki dołączaniu do dokumentów odpowiednich metadanych odnoszących się do kontekstu powstania dokumentów, a więc przy uwzględnieniu celów, dla których dokument powstał, czasu powstania dokumentu, autorów i nadawców dokumentu oraz jego treści. W praktyce najczęściej jednostki składające się na cały zasób dokumentacji wydzielane są na podstawie logicznych układów treściowych, które są odzwierciedlone w wykazach akt. Wykazy takie tworzy się w jednostkach wytwarzających dokumentację o wartości historycznej w sposób skoordynowany i nadzorowany przez Archiwa Państwowe, natomiast we wszystkich instytucjach, które nie podlegają nadzorowi archiwalnemu, tworzenie wykazów akt jest dowolne i zazwyczaj przebiega zgodnie z własną wypracowaną metodologią. Niemniej jednak przy wdrożeniu elektronicznego zarządzania dokumentacją warto przyjrzeć się układom do-

kumentów i grupowaniu dokumentacji w akta spraw, łączeniu w jednostki czy tomy, aby przy wyborze odpowiedniej metodologii ułatwić oraz stworzyć bardziej użyteczne i transparentne metody zarządzania aktami. Sama specyfikacja MoReq2 zaleca wprowadzenie wykazu akt jako podstawowego narzędzia dla systemu zarządzania dokumentacją elektroniczną. Wykaz akt pozwala na przechowywanie dokumentów elektronicznych wraz z innymi dokumentami odnoszącymi się do tego samego kontekstu, dzięki czemu dokumenty przyjmują określony układ organizacyjny, grupowane są w jednostki elektroniczne i znane są relacje zachodzące pomiędzy dokumentami⁶.

W instytucjach, które są zobligowane z mocy prawa do wprowadzania jednolitych rzeczowych wykazów akt⁷, grupowanie dokumentów przebiega w obrębie klas z wykazu akt oznaczonych symbolami cyfrowymi i przyjętych dla wszystkich instytucji wymienionych w tych przepisach. Tak ukształtowana polska praktyka kancelaryjna nie odbiega od przyjętych w specyfikacji MoReq2 założeń, jest także podstawą elektronicznego zarządzania dokumentacją w polskich urzędach.

Klasyfikacja dokumentów

Wymagania MoReq2 jasno precyzują potrzebę klasyfikacji dokumentów przy użyciu hierarchicznych wykazów akt oraz wymagania dla prawidłowego konfigurowania wykazu akt⁸. Hierarchiczne wykazy akt służą zapewnieniu efektywnego, stabilnego i czytelnego pogrupowania dokumentów. Wykazy akt mają także szerokie zastosowanie w całej Europie, zgodne są z dotychczas stosowaną i przyjętą praktyką kancelaryjną. Tworzenie wykazów akt nieodłącznie związane jest z zakresem działalności danej organizacji, instytucji czy firmy. Podstawowym założeniem, zgodnie z wymogami specyfikacji, jest uwzględnienie przy tworzeniu i konfiguracji wykazu akt potrzeb biznesowych i zadań realizowanych w organizacji. Klasyfikacja dokumentów jest podstawą dla wyszukiwania dokumentów w systemie, jak również wyznaczania okresów przechowywania dokumentacji i w konsekwencji jest podstawą wskazania czy typowania dokumentacji do zniszczenia, a także wydzielania dokumentacji do wieczystego przechowywania⁹. MoReq2 wska-

⁶ *Ibid.*, s. 35.

⁷ Przykładem wprowadzenia obowiązku stosowania jednolitego rzeczowego wykazu akt jest Rozporządzenie Prezesa Rady Ministrów z dnia 18 stycznia 2011 r. w sprawie instrukcji kancelaryjnej, jednolitych rzeczowych wykazów akt oraz instrukcji w sprawie organizacji i zakresu działania archiwów zakładowych (Dz. U. Nr 14 poz. 67).

⁸ *Specyfikacja...* s. 36–68.

⁹ Wieczyste przechowywanie odnosi się do jednostek będących pod nadzorem archiwów państwowych i wytwarzających materiały archiwalne o wartości historycznej.

zuje na potrzebę określenia grup użytkowników, którzy mają uprawnienia do zarządzania poszczególnymi, pojedynczymi klasami wykazu akt. Wskazuje na potrzebę przyjmowania nieograniczonej ilości poziomów w hierarchii klasyfikacji, choć w praktyce liczba poziomów nie powinna przekraczać dziesięciu. Przykładowy wykaz akt przyjęty dla organów samorządu województwa i urzędów marszałkowskich ma rozwinięcie do czwartego stopnia szczegółowości podziału (najbardziej szczegółowe są symbole czterocyfrowe), co przedstawiono w tabeli. Każda klasa w wykazie akt jest oznaczona poprzez symbol klasyfikacyjny, stanowiący kombinację cyfr według szczegółowości haseł. W podanym przykładzie dla klas pierwszego rzędu są to symbole jednocyfrowe od „0” do „9”, następnie dla klas drugiego rzędu – symbole dwucyfrowe od „00” do „99” i kolejno dla klas trzeciego rzędu – symbole trzycyfrowe od „000” do „999” oraz dla klas czwartego rzędu – symbole czterocyfrowe od „0000” do „9999”. Wszystkie hasła klasyfikacyjne w wykazie akt otrzymują także nazwę, czyli tytuł. Klasy końcowe, które nie są już dalej rozwijane, otrzymują ponadto oznaczenie kategorii archiwalnej, która wskazuje na poszczególne okresy przechowywania dokumentacji.

Tabela 1. Fragment wykazu akt dla organów samorządu województwa i urzędów marszałkowskich

Symbole klasyfikacyjne	Hasło klasyfikacyjne	Oznaczenie kategorii archiwalnej	Uszczegółowienie hasła klasyfikacyjnego
0	ZARZĄDZANIE WOJEWÓDZTWEM I JEGO REPREZENTACJA		
00	Organy kolegialne i jednoosobowe województwa		
000	Sejmik Województwa		
0000	Organizacja pracy Sejmiku Województwa	A	m.in. regulaminy
0001	Planowanie i sprawozdawczość z prac Sejmiku Województwa	A	w tym korespondencja
0002	Sesje Sejmiku Województwa	A	m.in. zawiadomienia, stenogramy, protokoły, materiały pod obrady, uchwały
0003	Wnioski i interpelacje radnych	A	w tym rejestr i odpowiedzi
0004	Przewodniczący i wiceprzewodniczący Sejmiku Województwa	A	m.in. korespondencja kierowana do i od przewodniczącego

0005	Wnioski i postulaty mieszkańców i innych podmiotów do Sejmiku Województwa oraz spotkania z mieszkańcami i innymi podmiotami	A	w tym rejestr i odpowiedzi
0006	Projekty uchwał Sejmiku Województwa	A	proces ich przygotowania i uzgadniania
0007	Uchwały Sejmiku Województwa	A	w tym ich rejestr; przy czym sprawy ich publikowania i badania zgodności z prawem przy klasie 0811
0008	Realizacja uchwał Sejmiku Województwa i jej monitorowanie	A	
001	Komisje, zespoły i inne organy kolegialne Sejmiku Województwa		
0010	Powoływanie i organizacja pracy komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	
0011	Planowanie i sprawozdawczość z prac komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	
0012	Posiedzenia komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	między innymi zawiadomienia, stenogramy, protokoły, materiały pod obrady, uchwały
0013	Projekty uchwał komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	
0014	Uchwały komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	w tym ich rejestr

0015	Realizacja uchwał komisji, zespołów i innych organów kolegialnych Sejmiku Województwa	A	
002	Zarząd Województwa		
0020	Powoływanie i organizacja pracy Zarządu Województwa	A	m.in. regulaminy
0021	Planowanie i sprawozdawczość z prac Zarządu Województwa	A	w tym korespondencja
0022	Posiedzenia Zarządu Województwa	A	m.in. zawiadomienia, stenogramy, protokoły, materiały pod obrady, uchwały
0023	Wnioski i postulaty mieszkańców i innych podmiotów do Zarządu Województwa oraz spotkania z mieszkańcami i innymi podmiotami	A	w tym rejestr i odpowiedzi
0024	Projekty aktów normatywnych Zarządu Województwa	A	proces ich przygotowania i uzgadniania
0025	Akty normatywne Zarządu Województwa	A	w tym ich rejestr; przy czym sprawy ich publikowania i badania zgodności z prawem przy klasie 0811
0026	Realizacja aktów normatywnych Zarządu Województwa i jej monitorowanie	A	
0027	Upoważnianie osób (podmiotów) do działania w imieniu Zarządu Województwa	A	pozostałe upoważnienia przy klasie 087

Źródło: Załącznik nr 4 do Rozporządzenia Prezesa Rady Ministrów z dnia 18 stycznia 2011 r. w sprawie instrukcji kancelaryjnej, jednolitych rzeczowych wykazów akt oraz instrukcji w sprawie organizacji i zakresu działania archiwów zakładowych [Dz. U. Nr 14 poz. 67].

Budowa klasyfikacji jest hierarchiczna i przebiega w oparciu o symbole cyfrowe, zaleca się, aby w systemie zarządzania dokumentacją elektroniczną uwzględnić możliwość nadawania tytułów dokumentom, które są odpo-

wiednikami słownymi haseł klasyfikacyjnych z wykazu akt, jak to podano w przykładzie.

Podsumowanie

Chcąc sprostać szybko rozwijającym się technologiom informacyjnym i komunikacyjnym, współczesne archiwum i archiwista mają do spełnienia szeregu zadań związanych z zarządzaniem dokumentacją. Zadania te znacznie wykraczają poza ramy tradycyjnie pojmowanej archiwistyki. Archiwista w swej pracy administracyjnej musi być doskonale zorientowany w procesach komunikacyjnych, przepisach prawnych, jak również normach i zaleceniach międzynarodowych z zakresu archiwizacji dokumentów elektronicznych. W coraz mniejszym stopniu archiwista zakładowy związany jest z tradycją i praktyką, jaką od lat stosowały archiwa w swej codziennej pracy. Konieczna jest m.in. znajomość modelu zaleceń, jakie wypracowała Komisja Europejska w zakresie elektronicznego zarządzania dokumentacją.

Prawidłowe zarządzanie dokumentacją jest warunkiem niezbędnym dla usprawnienia wszelkich czynności, w których dokument stanowi podstawowe narzędzie i przedmiot działania. Administracja publiczna w Polsce pracuje zgodnie z zaleceniami międzynarodowymi, stosuje szereg reguł, które obowiązują przy sporządzaniu dokumentacji, ale też w procesie obiegu dokumentów i ich przechowywania oraz archiwizowania. Wszelkie prace związane z postępowaniem z dokumentacją w urzędzie, jak też w innych jednostkach sektora publicznego i przedsiębiorstwach sektora prywatnego, można prowadzić przy wykorzystaniu reguł zarządzania dokumentacją. Zagadnienia te są coraz mocniej akcentowane w dobie szybkiego rozwoju technologii informacyjnych i komunikacyjnych, a co za tym idzie, także nowych form dokumentu. Dokument elektroniczny posiada także silne umocowanie prawne w polskim ustawodawstwie¹⁰. Postępowanie się dokumentacją w formie

¹⁰ Ustawa z dnia 18 września 2001 r. o podpisie elektronicznym (Dz. U. 2001 Nr 130 poz. 1450), która zrównała skutki prawne podpisu elektronicznego z podpisem odręcznym; ustawa z dnia 18 lipca 2002 r. o świadczeniu usług drogą elektroniczną (Dz. U. 2002 Nr 144 poz. 1204); ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz. U. 2005 Nr 64 poz. 565); Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie niezbędnych elementów struktury dokumentów elektronicznych (Dz. U. Nr 206 poz. 1517); Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie szczegółowego sposobu postępowania z dokumentami elektronicznymi (Dz. U. Nr 206 poz. 1518); Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 2 listopada 2006 r. w sprawie wymagań technicznych formatów zapisu i informatycznych nośników danych, na których utrwalono materiały archiwalne przekazywane do archiwów państwowych (Dz. U. Nr 206 poz. 1519); Rozporządzenie Prezesa Rady Ministrów z dnia 29 września 2005 r. w sprawie warunków organizacyjno-technicznych doręczania dokumentów elektronicznych

elektronicznej i korzystanie ze współczesnych technologii stanowią podstawę tworzenia takich mechanizmów, które w dużo łatwiejszy sposób umożliwią komunikację i wydawanie decyzji, jak również udzielanie informacji oraz świadczenie usług, stwarzając tym samym podstawy do oszczędności ekonomicznych i organizacyjnych dla państwa i społeczeństwa.

Bibliografia

1. PN-ISO 15489-1:2006. Informacja i dokumentacja. Zarządzanie dokumentami. Część 1: Zasady ogólne, 28 kwietnia 2006.
2. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie niezbędnych elementów struktury dokumentów elektronicznych (Dz. U. Nr 206 poz. 1517).
3. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 30 października 2006 r. w sprawie szczegółowego sposobu postępowania z dokumentami elektronicznymi (Dz. U. Nr 206 poz. 1518).
4. Rozporządzenie Ministra Spraw Wewnętrznych i Administracji z dnia 2 listopada 2006 r. w sprawie wymagań technicznych formatów zapisu i informatycznych nośników danych, na których utrwalono materiały archiwalne przekazywane do archiwów państwowych (Dz. U. Nr 206 poz. 1519).
5. Rozporządzenie Prezesa Rady Ministrów z dnia 11 października 2005 r. w sprawie minimalnych wymagań dla systemów teleinformatycznych (Dz. U. Nr 212 poz. 1766).
6. Rozporządzenie Prezesa Rady Ministrów z dnia 11 października 2005 r. w sprawie minimalnych wymagań dla rejestrów publicznych i wymiany informacji w formie elektronicznej (Dz. U. Nr 214 poz. 1781).
7. Rozporządzenie Prezesa Rady Ministrów z dnia 18 stycznia 2011 r. w sprawie instrukcji kancelaryjnej, jednolitych rzeczowych wykazów akt oraz instrukcji w sprawie organizacji i zakresu działania archiwów zakładowych (Dz. U. Nr 14 poz. 67).
8. Rozporządzenie Prezesa Rady Ministrów z dnia 29 września 2005 r. w sprawie warunków organizacyjno-technicznych doręczania dokumentów elektronicznych podmiotom publicznym (Dz. U. Nr 200 poz. 1651).
9. *Specyfikacja MoReq2. Wzorcowe wymagania dotyczące zarządzania dokumentami elektronicznymi. Aktualizacja i rozszerzenie* [online], [dostęp: 11.09.2012], dostępny w World Wide Web: http://www.archiwa.gov.pl/images/stories/moreq2translation%20iv%20Polish%20versiob%20beta12_10_2011.pdf.
10. Ustawa z dnia 14 lipca 1983 r. o narodowym zasobie archiwalnym i archiwach (Dz. U. 1983 Nr 38 poz. 173).

podmiotom publicznym (Dz. U. Nr 200 poz. 1651); Rozporządzenie Prezesa Rady Ministrów z dnia 11 października 2005 r. w sprawie minimalnych wymagań dla systemów teleinformatycznych (Dz. U. Nr 212 poz. 1766); Rozporządzenie Prezesa Rady Ministrów z dnia 11 października 2005 r. w sprawie minimalnych wymagań dla rejestrów publicznych i wymiany informacji w formie elektronicznej (Dz. U. Nr 214 poz. 1781).

11. Ustawa z dnia 17 lutego 2005 r. o informatyzacji działalności podmiotów realizujących zadania publiczne (Dz. U. 2005 Nr 64 poz. 565).
12. Ustawa z dnia 18 lipca 2002 r. o świadczeniu usług drogą elektroniczną (Dz. U. 2002 Nr 144 poz. 1204).
13. Ustawa z dnia 18 września 2001 r. o podpisie elektronicznym (Dz. U. 2001 Nr 130 poz. 1450).

Dagmara Bubel¹, Łukasz Kuczyński², Lidia Szczygłowska³

¹ Biblioteka Główna Politechniki Częstochowskiej

² Instytut Informatyki Teoretycznej i Stosowanej Politechniki Częstochowskiej

³ Biblioteka Główna Politechniki Częstochowskiej

PLATON U4 – i wszystko jasne

Słowa kluczowe: Usługa Powszechnej Archiwizacji, PLATON U4

Abstrakt: nasza cywilizacja w czasach nazywanych erą informacji generuje olbrzymie ilości danych, przez co ciągle rośnie zapotrzebowanie na ich bezpieczne i efektywne przechowywanie. Na gruncie polskim do największych producentów danych należą biblioteki i archiwa cyfrowe, instytuty, uczelnie i jednostki naukowe oraz akademickie centra komputerowe. Zapotrzebowanie bibliotek cyfrowych na przestrzeń do przechowywania danych jest wprost proporcjonalne do wielkości zbiorów w wersji oryginalnej oraz zdolności do ich digitalizacji. Usługi powszechnej archiwizacji rewolucjonizują podejście do zabezpieczania danych. Ich ścisła integracja, jako funkcjonalność samej sieci, umożliwia zarówno bezpieczny, jak i efektywny dostęp do gromadzonych danych. Usługi te jednocześnie mogą być bazą do powstania laboratorium badawczego dla obszaru tzw. „sieci świadomych treści”. Stwarzają one szansę przechowywania w formie cyfrowej dostępnego w Polsce dziedzictwa kulturowego. Usługa Powszechnej Archiwizacji Platon U4 rozwiązuje większość problemów związanych ze skutecznym i wiarygodnym zabezpieczeniem lub archiwizacją dużych ilości danych, pozwalając na pełne wykorzystanie środowiska cyfrowego.

Wstęp

Paradygmat otwartości zmienia trwale krajobraz świata akademickiego, rewolucjonizując ścieżki komunikacji naukowej. Komunikacja naukowa ulega transformacji wynikającej z przenoszenia jej form do środowiska cyfrowego. Zmiany w komunikacji naukowej powodują m.in. powstawanie jednolitych zasobów informacji, które stanowią podstawę globalnej biblioteki cyfrowej¹. Globalna biblioteka cyfrowa będzie mocniej włączona w procesy badawcze i dydaktyczne. Stanie się częścią wirtualnego laboratorium, a nie dodatkiem czy uzupełnieniem instytucji realizujących procesy badawcze, co zmieni także role uczestników komunikacji naukowej. Te ulegną „wymieszanu”, gdyż

¹ M. Nahotko, *Komunikacja naukowa w środowisku cyfrowym: globalna biblioteka cyfrowa w informatycznej infrastrukturze nauki*, Warszawa 2010, s. 17.

granice między tworzeniem informacji, publikowaniem, gromadzeniem i udostępnianiem zacierają się². Bibliotekarz stanie się członkiem wirtualnego kolaboratorium, zespołu badawczego funkcjonującego w cyberprzestrzeni, jego zadaniem będzie obsługa informacyjna zespołu poprzez dostarczanie informacji niezbędnej do właściwej pracy zespołu, jak i przekazywanie informacji o jego pracach i osiągnięciach³. Zmiany te uwidoczniają się w idei bibliotekarstwa uczestniczącego (*Embedded librarianship*)⁴. Bibliotekarstwo uczestniczące⁵ skupia się na użytkowniku i przenosi bibliotekę oraz bibliotekarzy do użytkowników, gdziekolwiek oni są – do szpitala, sali wykładowej, laboratorium, biura, domu czy nawet na urządzenia przenośne⁶.

Działania, których celem jest dążenie do głębokiej transformacji systemu nauki w Polsce, w czasie globalnej rewolucji cyfrowej, mające wnieść wkład do inicjatywy na rzecz nauki otwartej i transparentnej, polegają na:

- wdrożeniu modelu otwartego dostępu dla całej polskiej nauki i szkolnictwa wyższego;
- wprowadzeniu otwartego mandatu repozytoryjnego dla wszystkich krajowych publikacji naukowych;
- stworzeniu mechanizmów wspierających budowę otwartych repozytoriów naukowych i działaniu na rzecz ich integracji w formie federacyjnej;
- uruchomieniu inicjatywy na rzecz otwartych repozytoriów danych naukowych i oprogramowania naukowego.

Biblioteki naukowe odpowiedzialne za zachowanie i udostępnianie dorobku naukowego znajdują się w centrum tych przemian. Ich działalność może stać się funkcjonalną częścią tej nowo powstającej, złożonej struktury, nazwanej globalną biblioteką cyfrową, stanowiącej zintegrowane środowisko informacji naukowej.

Długoterminowa archiwizacja, obok zapewnienia właściwej liczebności udostępnianych zbiorów, standaryzacji oraz identyfikowalności treści, usług i osób, uznawana jest za najważniejszą cechę właściwie zorganizowanego środowiska informacji cyfrowej⁷.

² *Ibid.*, s. 62.

³ D. Shumaker, *Breaking Out of the Box. The Promise of Embedded Librarianship* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://archive.saoug.org.za/presentations/BreakingOutoftheBox.pdf>.

⁴ J. Siess, *Embedded Librarians. Our Future Or Our Fear?* Annual Conference: Canadian Association of Law Libraries, Halifax, NS, 24–27 maja 2009.

⁵ D. Bubel, L. Szczygłowska, *Bibliotekarz uczestniczący – człowiek orkiestra*, III Ogólnopolska Konferencja Naukowa: Bibliotekarze bez bibliotek, czyli bibliotekarstwo uczestniczące, Białystok 18–20 maja 2011.

⁶ M. Kesselman, S. Watstein, *Creating Opportunities. Embedded Librarians*, „Journal of Library Administration” 2009, Vol. 49, Issue 4, s. 383–400.

⁷ M. Nahotko, *op. cit.*, s. 189.

Skuteczne i wiarygodne zabezpieczenie lub archiwizacja przez biblioteki tak dużych ilości danych jest ogromnym wyzwaniem. Proces ten wymaga bowiem odpowiednich zasobów sprzętowych i programowych, włączając w to urządzenia do przechowywania danych. Nakłady finansowe związane z długoterminowym przechowywaniem danych to nie tylko zakup sprzętu czy oprogramowania, lecz także koszty związane z utrzymaniem systemu w razie konieczności długoterminowego lub wieczystego przechowywania danych.

Podstawową potrzebą producentów danych jest odpowiednio duża przestrzeń do ich przechowywania. Z ekonomicznego punktu widzenia korzystna byłaby konsolidacja pamięci masowych. Oznacza to przejście od przechowywania danych na lokalnych systemach, poprzez użycie specjalistycznych serwerów danych, do wykorzystania centrów przechowywania (ang. *data center*). W ramach centrów danych możliwe jest wprowadzenie bardziej efektywnych, elastycznych i bezpiecznych rozwiązań, takich jak: macierze dyskowe, biblioteki taśmowe, sieci SAN (ang. *Storage Area Network*), a także oprogramowanie do zarządzania danymi, systemy kopii zapasowych i archiwizacji lub systemy hierarchicznego przechowywania danych (ang. *Hierarchical Storage Management – HSM*).

Wiele problemów wiąże się ponadto z szeroko pojętym zabezpieczeniem danych, głównie przed całkowitą ich utratą na skutek klęski żywiołowej, takiej jak powódź czy pożar, lub awarii, aktów wandalizmu czy kradzieży sprzętu. Odpowiednio wyposażone centra komputerowe oraz technologie redundancji przechowywania danych zwiększają bezpieczeństwo informacji, jednak jego wysoki poziom może dać jedynie kopiowanie danych do kilku geograficznie oddległych lokalizacji. To z kolei wymaga zastosowania zaawansowanego oprogramowania zarządzającego danymi i systemem przechowywania tak, aby proces replikacji odbywał się automatycznie, w sposób wydajny i zapewniający równowagę obciążenia infrastruktury, a przede wszystkim integralność danych. Co więcej, informacje trzeba też chronić przed przestępstwami informatycznymi, takimi jak złośliwe usunięcie danych, ich modyfikacja, kradzież czy blokowanie dostępu. Taką ochronę może zapewnić zastosowanie odpowiednich protokołów przesyłania danych, mechanizmów szyfrowania informacji i autoryzacji użytkowników oraz zabezpieczeń programowych i sprzętowych, skonfigurowanych przez specjalistów zajmujących się bezpieczeństwem⁸.

Nakłady finansowe związane z długoterminowym przechowywaniem danych nie kończą się wraz z zakupem sprzętu czy oprogramowania, konieczne jest ponoszenie kosztów związanych z utrzymaniem systemu, włączając

⁸ M. Brzeźniak, *PLATON. Usługi powszechnej archiwizacji* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1374/PLATON_Uslugi_powszechej_archiwizacji.html.

w to koszty osobowe. Wymagana jest specjalistyczna wiedza i doświadczenie z dziedziny zabezpieczenia i przechowywania danych oraz zarządzania systemami informatycznymi. Dodatkowo kosztuje utrzymanie sprzętu do przechowywania danych, należy także przewidzieć nakłady związane z serwisowaniem elementów systemu informatycznego czy zapewnieniem tym systemom właściwych warunków pracy przez odpowiednie zasilanie, klimatyzację itp. Sporym problemem jest także przeciwdziałanie skutkom starzenia się nośników danych. Niezależnie od rodzaju nośnika, jego żywotność jest ograniczona. W razie konieczności długoterminowego przechowywania danych, jak to jest w przypadku bibliotek cyfrowych, konieczna jest migracja danych na nowe nośniki, po upływie określonego czasu. Poza starzeniem się fizycznym, nośniki przechowujące dane starzeją się technologicznie. Zapewnienie długookresowego przechowywania danych wymaga więc migracji między technologiami przechowywania, w praktyce bowiem po upływie 5–10 lat od zapisania danych na nośniku określonego typu dostęp do czytnika mediów danej technologii może okazać się trudny, niemożliwy lub bardzo kosztowny⁹.

Opisanych wyżej wymagań nie jest w stanie spełnić większość instytucji naukowych i akademickich, ze względu na brak kadry specjalizującej się w problematyce zabezpieczania danych i zaawansowanych systemów informatycznych. Proces skutecznego i bezpiecznego zabezpieczenia danych może przekraczać możliwości bibliotek. Rozwiązaniem dla nich może być więc przeniesienie problemu na zewnętrzną instytucję (ang. *outsourcing*).

Usługa Powszechnej Archiwizacji

Dla bibliotekarzy interesująca może być Usługa Powszechnej Archiwizacji, która wychodzi naprzeciw potrzebom zabezpieczania danych w instytucjach naukowych, również w bibliotekach cyfrowych. Realizowana jest przez Poznańskie Centrum Superkomputerowo-Sieciowe w ramach projektu PLATON Platformy Obsługi Nauki, który zakłada stworzenie i uruchomienie pięciu usług działających w oparciu o ogólnopolską akademicką sieć naukową PIONIER. Usługa oparta jest na oprogramowaniu Krajowego Magazynu Danych (KMD), które zostało wdrożone w redundantnej, wysoko wydajnej i skalowalnej infrastrukturze serwerów i systemów przechowywania danych.

Odpowiednimi zasobami do realizacji tych wymagań dysponuje dziesięć ośrodków biorących udział w projekcie PLATON. Usługi powszechnej archiwizacji oparte zostały o sieć PIONIER, zakupiony w ramach projektu

⁹ M. Brzeźniak, *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych* [online], [dostęp: 25.04. 2011], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?brzezniaak>.

PLATON sprzęt i oprogramowanie do przechowywania danych, a także wykwalifikowaną i doświadczoną kadrę. Potrzebne oprogramowanie zostało też stworzone w ramach projektu badawczego Krajowy Magazyn Danych.

Usługi powszechnej archiwizacji obejmują: archiwizację danych, tworzenie i odtwarzanie kopii zapasowych (ang. *backup*) oraz składowanie danych w tzw. wirtualnym systemie plików. Usługi dostępne są online przez całą dobę, 7 dni w tygodniu.

W 10 polskich miastach (Białystok, Częstochowa, Gdańsk, Kielce, Kraków, Lublin, Łódź, Poznań, Warszawa, Wrocław) rozmieszczone są węzły przechowywania danych oraz węzły usługowe, połączone z optycznymi sieciami miejskimi, zapewniając odpowiednią przepustowość tzw. ostatniej mili, i siecią krajową, zapewniając odpowiedniej jakości połączenia szkieletowe. W każdym z tych ośrodków bezpośrednie wsparcie dla klienta świadczą specjaliści Konsorcjum PIONIER. Jest to pierwsze rozwiązanie o tak dużym zasięgu w skali kraju¹⁰.

Jedną z podstawowych technik zapewnienia wiarygodności przechowywania danych i ich trwałości oraz niezawodności usługi przechowywania jest replikacja danych w rozproszonym środowisku przechowywania oraz wykorzystanie redundantnych komponentów infrastruktury. Usługa dostarcza użytkownikom rozmaite interfejsy dostępu do danych składowanych w wirtualnym systemie plików oraz możliwość automatyzacji i optymalizacji procesu wykonywania kopii zapasowych i archiwalnych poprzez wykorzystanie aplikacji klienta opracowanej w projekcie KMD.

Usługa kopii zapasowych i archiwizacji dostarczana jest w postaci specjalnej aplikacji działającej w systemie komputerowym użytkownika, która pozwala na optymalizację czasu wykonania kopii oraz jej objętości poprzez wsparcie dla kopii przyrostowych i wielowątkowego, równoległego zapisu danych do wielu serwerów. Równoległy *backup* pozwala na pełne wykorzystanie dostępnego pasma sieciowego, a zastosowanie wielu serwerów zapewnia kopiowanie informacji równocześnie do wielu centrów przechowywania danych. Dzięki temu, że aplikacja działa na komputerze użytkownika, składowane przez nią dane mogą być szyfrowane jeszcze przed opuszczeniem systemu komputerowego użytkownika przy wykorzystaniu klucza kryptograficznego znanego tylko klientowi usługi.

Już na etapie projektowania systemu wysoki priorytet nadano bezpieczeństwu i poufności danych. W przypadku wirtualnego systemu plików ewentualne szyfrowanie pozostaje w gestii użytkownika, jednakże komunikacja z systemem odbywa się przy zastosowaniu wspomnianych bezpiecznych i standardowych protokołów, co ułatwia integrację usług systemu z narzędziami realizującymi dodatkowe techniki kryptograficzne. Wewnątrz systemu poufność komunikacji

¹⁰ M. Brzeźniak, *PLATON...*, *op. cit.*

zapewniają bezpieczne połączenia w ramach sieci PIONIER. Wirtualny system plików oferuje ponadto oddzielne przestrzenie nazw i oddzielne bazy metadanych dla różnych instytucji-klientów. Takie rozwiązanie, nowatorskie względem dostępnych na rynku produktów, powoduje wysoki poziom izolacji danych i metadanych (zwiększona poufność) i daje potencjał do rozbudowy systemu bez utraty wydajności. Dodatkowe zabezpieczenia to przechowywanie i weryfikacja sum kontrolnych (na poziomie wirtualnego systemu plików), sprzętowe szyfrowanie i deszyfrowanie plików (na poziomie napędów taśmowych i napędów w macierzach dyskowych), a także audyty bezpieczeństwa oprogramowania i konfiguracji systemu prowadzone cyklicznie przez konsorcjum w infrastrukturze PLATONA.

Wykorzystanie usług systemu zwalnia użytkowników z konieczności przeznaczania znacznych środków na potrzeby zabezpieczenia i archiwizacji danych. Interfejsy i funkcjonalność systemu pozwalają na efektywne wykorzystanie usługi na potrzeby zabezpieczenia i archiwizacji danych bibliotek naukowych.

Funkcjonalność Krajowego Magazynu Danych

Krajowy Magazyn Danych (KMD) oferuje dwie usługi wydajnego, wiarygodnego i bezpiecznego przechowywania danych. Dane użytkowników przechowywane są w postaci wielu fizycznych replik danych rozmieszczonych w rozproszonych geograficznie lokalizacjach.

Pierwsza usługa to tzw. wirtualny system plików dostępny dla użytkowników zdalnie za pośrednictwem protokołów SCP/SFTP, HTTP i WebDAV (oraz GridFTP). Dane użytkowników zorganizowane są w wirtualnym systemie plików w plikach i katalogach, analogicznie do przechowywania ich na dysku lokalnym lub sieciowym. Dane te mogą być umieszczane w wirtualnym systemie plików ręcznie, np. poprzez użycie klientów usługi SCP/SFTP (np. WinSCP), przeglądarki internetowej z odpowiednimi modułami – WebDAV lub z wykorzystaniem zewnętrznych narzędzi, takich jak Bacula (<http://www.bacula.org/en>)¹¹ lub Amanda (<http://www.amanda.org>).

W ramach drugiej usługi opracowano aplikację klienta usługi kopii zapasowych i archiwizacji, która pozwala na automatyzację procesu wykonywania kopii zapasowych, np. poprzez cykliczne wykonywanie zdefiniowanego harmonogramu kopii danych, a także optymalizację tego procesu, m.in. przez:

¹¹ M. Haba, *Podstawowa instalacja i konfiguracja oprogramowania Bacula* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://www.bacula.pl/artukul/39/podstawowa-instalacja-i-konfiguracja-oprogramowania-bacula>.

- wsparcie dla tzw. kopii przyrostowych;
- możliwość równoległego przesyłania danych przez sieć wieloma strumieniami od klienta do serwera usługi;
- możliwość składowania wielu replik danych na wielu serwerach jednocześnie.

Dodatkowo, aplikacja klienta backupu i archiwizacji może realizować szyfrowanie i kryptograficzną kontrolę integralności danych składowanych w systemie. Co istotne, klucze kryptograficzne pozostają znane wyłącznie użytkownikowi końcowemu. Szyfrowanie i deszyfrowanie danych użytkownika odbywa się całkowicie po stronie klienta, dzięki czemu żadna porcja danych nie opuszcza systemu komputerowego użytkownika bez uprzedniego zaszyfrowania.

Architektura i mechanizmy systemu KMD zapewniają automatyczną replikację danych użytkownika. Gwarantuje to trwałość zapisanych w systemie danych nawet w przypadku zniszczenia jednego z centrów przechowywania. Replikacja zapewnia także wysoką dostępność danych – mogą być one odczytywane nawet w przypadku awarii jednego z węzłów systemu – a także pozwala optymalizować wydajność operacji na danych. Do ich odczytu z systemu wykorzystywana jest najszybsza, np. najbliższa geograficznie replika¹².

Krajowy Magazyn Danych – ogólny zarys projektu

Głównymi komponentami Krajowego Magazynu Danych (KMD) są Węzły Dostępowe (ang. *Access Node*) oraz Węzły Składowania (ang. *Storage Node*). Użytkownik kontaktuje się z Węzłem Dostępowym, który oferuje kilka Metod Dostępu do Danych (ang. *Data Access Methods*), takich jak: SSH, HTTP czy GridFTP. Operacje wejścia/wyjścia przechwycone od użytkownika przez Metody Dostępu do Danych są tłumaczone na wywołania Wirtualnego Systemu Plików (ang. *Virtual File System*) i obsługiwane przez Demona Danych (ang. *Data Daemon*)¹³. Demon Danych węzła dostępowego implementuje podstawowe operacje wejścia/wyjścia, takie jak: *create*, *open*, *flush*, *release*, *read*, *write*, *getattr*, *setattr* itd. Wykonuje on operacje na replikach, takie jak utworzenie, usunięcie, modyfikacja, odczytanie lub też zapis fizycznego pliku przechowywanego na Węźle Składowania.

Mapowanie pomiędzy logicznymi plikami a ich fizycznymi replikami, podobnie jak logiczna struktura systemu plików oraz atrybuty pliku, położe-

¹² M. Brzeźniak, *Usługa Powszechnej Archiwizacji...*, *op. cit.*

¹³ M. Brzeźniak i in., *Usługa powszechnej archiwizacji w sieci PIONIER i jej zastosowanie do składowania danych sieciowych*, I Konferencja i3. Internet – Infrastruktury – Innowacje. Nauka w obliczu społeczeństwa cyfrowego, Poznań 4–6 listopada 2009.

nie i parametry replik są obsługiwane i utrzymywane przez MetaKatalog KMD (ang. *NDS MetaCatalog*). MetaKatalog jest replikowany w semi-synchroniczny sposób pomiędzy wieloma geograficznie odległymi Węzłami Dostępowymi, co podwyższa niezawodność systemu.

Przetwarzanie danych przez użytkownika końcowego jest wykonywane zgodnie z profilem użytkownika przetrzymywanym w Bazie Danych Użytkowników (ang. *Users DataBase*). Profil opisuje m.in. takie parametry usługi przechowywania danych, jak tryb replikacji (synchroniczny/asynchroniczny), liczba wymaganych replik, dopuszczalne lokalizacje replik, preferowane rodzaje zasobów do przechowywania danych.

Uwierzytelnianie użytkowników względem metod dostępowych jest realizowane z wykorzystaniem warstwy wirtualnego systemu plików: Demon Autentykacji (ang. *Authentication Daemon*) prezentuje dane autentykacyjne (przechowywane w Bazie LDAP) Metodami Dostępowymi w zrozumiałym dla nich formacie, wykorzystując przy tym wirtualny system plików – KeyFS.

Węzły Składowania przechowują fizyczne repliki danych użytkowników. Dane mogą być przechowywane na różnych typach mediów, zarówno na dyskach, taśmach, jak i systemie HSM (ang. *Hierarchical Data Management*).

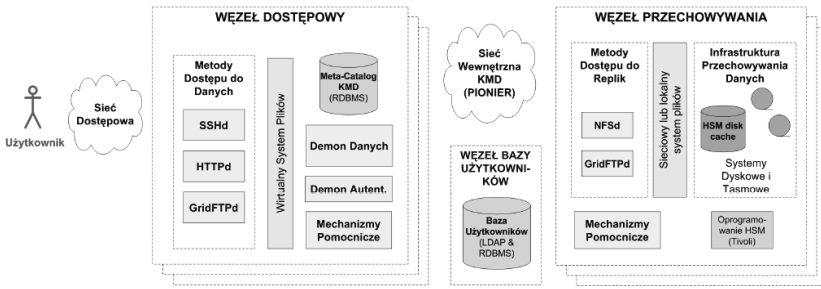
Dostęp do danych przechowywanych na Węzłach Składowania z Węzłów Dostępowych odbywa się z wykorzystaniem wewnętrznej sieci KMD (skonfigurowanej w sieci PIONIER), za pomocą Metod Dostępu do Replik (ang. *Replica Access Methods*), włączając NFS oraz GridFTP. Infrastruktura Zasobów Przechowywania, która przechowuje dane (system HSM, biblioteki taśmowe, macierze dyskowe), prezentowana jest Metodą Dostępu do Replik jako sieciowy lub też lokalny system plików.

Replikacja Metadanych jest jednym z unikalnych mechanizmów KMD. Metadane są automatycznie replikowane w semi-synchroniczny sposób. Oznacza to, że:

- system zarządzania bazą metadanych implementuje asynchroniczny mechanizm replikacji oparty na wyzwalaczach, który kopiuje dane z głównej instancji bazy metadanych do jej podrzędnych instancji;
- każda operacja na danych oraz metadanych jest logowana i przechowywana na Węzłach Składowania razem z replikami plików. Operacja zapisywania logu jest wykonywana synchronicznie z właściwą operacją wykonywaną na fizycznej replice.

Mechanizmy te umożliwiają bezstratne odzyskanie metadanych nawet w przypadku uszkodzenia głównej instancji bazy (rysunek 1)¹⁴.

¹⁴ PCSS. KMD. Krajowy Magazyn Danych [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://kmd.pcss.pl/ogolne.html>.

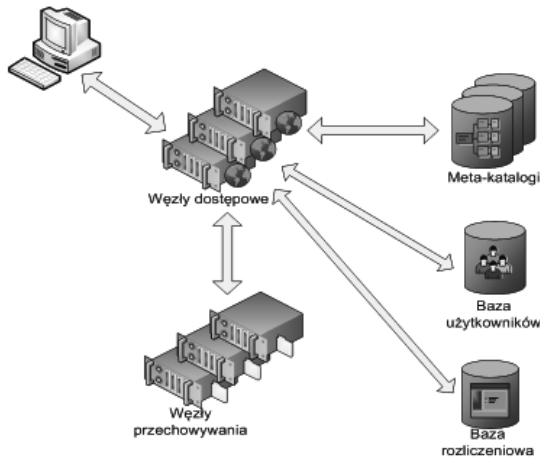


Rysunek 1. Ogólna architektura KMD

Źródło: M. Brzeźniak i in., *Ustuga...*, op. cit.

Charakterystyka architektury:

- decentralizacja danych oraz usług;
- replikacja metadanych;
- liczne punkty dostępne (Węzły Dostępowe);
- liczne lokalizacje przechowywania replik (Węzły Składowania);
- centralna baza danych użytkowników (pojedyncze logowanie, ang. *single sign-on*);
- standardowe interfejsy pomiędzy warstwami: wirtualny system plików oraz standardowe metody dostępu do danych.



Rysunek 2. Zarys architektury

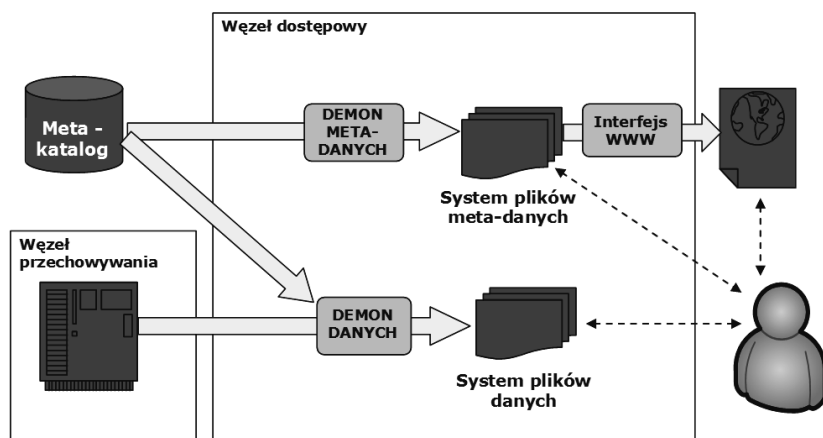
Źródło: M. Jankowski, M. Brzeźniak, *Architektura i mechanizmy systemu* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/online/pl/artykuly/1091/Materialy_archiwalne_z_warsztatow_Usluga_powszechniej_archiwizacji_juz_dostepne.html.

Demon dla danych:

- emuluje logiczny system plików użytkownika na węźle dostępowym;
- realizuje polityki bezpieczeństwa, replikacji, limitów, optymalizacji;
- wytwarza dane accountingowe;
- dostęp dla użytkownika standardowymi metodami oraz przez portal.

Demon dla metadanych:

- emuluje system plików zawierający metadane na węźle dostępowym;
- pliki z metadanymi umieszczone w katalogach odpowiadających logicznym plikom i katalogom;
- dostęp dla użytkownika standardowymi metodami oraz przez portal.



Rysunek 3. Przepływ danych i metadanych

Źródło: M. Jankowski, M. Brzeźniak, *Architektura i mechanizmy systemu*, op cit.

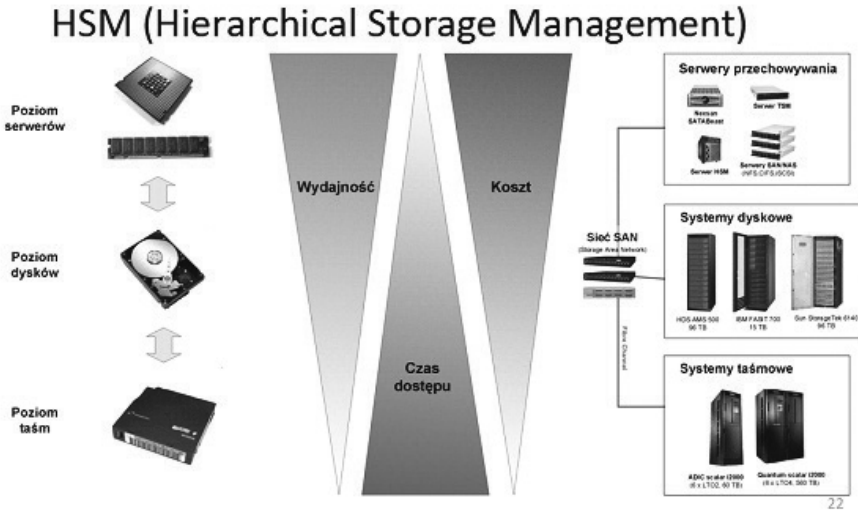
Dostęp do danych

Standardowe protokoły dostępu do danych i metadanych.

Po stronie użytkownika – typowe oprogramowanie klienckie:

- SSH/SCP/SFTP (WinSCP, SSHFS);
- HTTP/WebDAV (przeglądarka internetowa, klient WebDAV, mapowanie dysków w Windows);
- GridFTP.

Po stronie systemu – emulowane systemy plików z danymi i metadanymi.



Rysunek 4. Przechowywanie danych

Źródło: M. Jankowski, M. Brzeźniak, *Architektura i mechanizmy systemu*, op. cit.

Poufność i bezpieczeństwo danych:

- szyfrowanie połączeń klient–system i wewnątrz systemu (X.509);
- oddzielne przestrzenie nazw dla instytucji;
- audyty bezpieczeństwa systemu i oprogramowania;
- przechowywanie odpowiedniej liczby replik;
- wsparcie dla szyfrowania sprzętowego;
- komunikacja przez VPN.

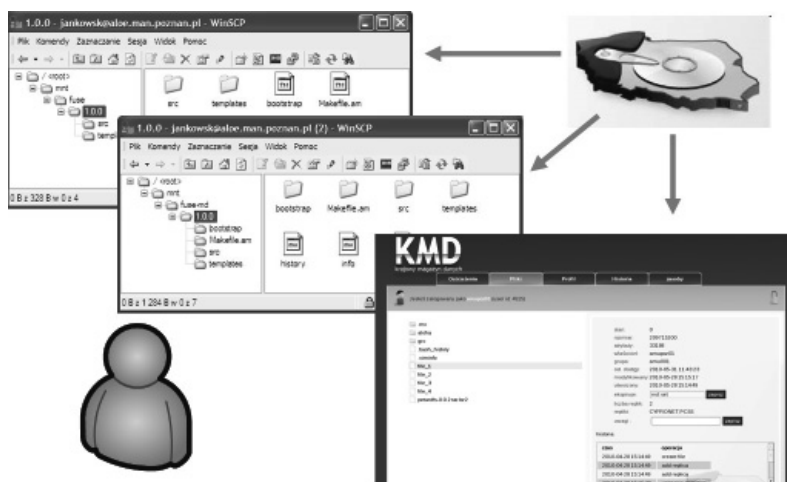
Zarządzanie cyklem życia informacji:

- zarządzanie liczbą replik;
- synchroniczna i asynchroniczna replikacja;
- asynchroniczne usuwanie danych;
- retencja danych.

Jak użytkownik widzi system:

- dostęp przez stronę www:
 - interfejs www użytkownika,
 - certyfikaty użytkownika i połączenie szyfrowane;
- dostęp przez dysk sieciowy WebDAV:
 - bezpośredni dostęp do danych,
 - możliwość podłączenia jako dysk;

- dostęp przez protokół SSH (SCP i SFTP):
 - wykorzystanie klienta SCP (WinSCP),
 - wykorzystanie narzędzia SSHFS¹⁵.



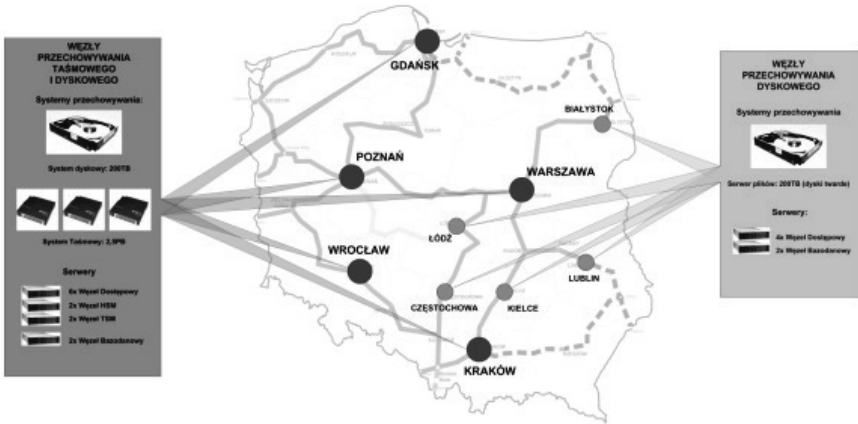
Rysunek 5. Jak użytkownik widzi system

Źródło: D. Bubel, L. Szczygłowska, Ł. Kuczyński, *Biblioteki cyfrowe – bibliotekami uczestniczącymi*, Konferencja Polskie Biblioteki Cyfrowe 2011. VIII Warsztaty „Biblioteki cyfrowe”, Poznań 10–13 października 2011.

Infrastruktura Usługi Powszechnej Archiwizacji

Infrastruktura Usługi Powszechnej Archiwizacji, w której została wdrożona architektura Krajowego Magazynu Danych, składa się z redundantnych, rozproszonych geograficznie elementów. Podstawowym składnikiem tej infrastruktury są systemy przechowywania, w tym macierze dyskowe i serwery plików o łącznej pojemności około 2 petabajtów oraz systemów przechowywania taśmowego o pojemności 12,5 petabajta. Poza systemami przechowywania danych, na infrastrukturę składają się serwery dostępowe, bazodanowe oraz serwery dla oprogramowania HSM – łącznie ponad 70 maszyn. Elementy instalacji rozlokowane są w dziesięciu miastach Polski i połączone za pomocą wydajnych łącz sieciowych w PIONIER.

¹⁵ M. Białokórski, B. Balcerek, *Narzędzia klienta usługi archiwizacji* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/files/platon-u4_narzedzia_klienta_v20100608.pdf.

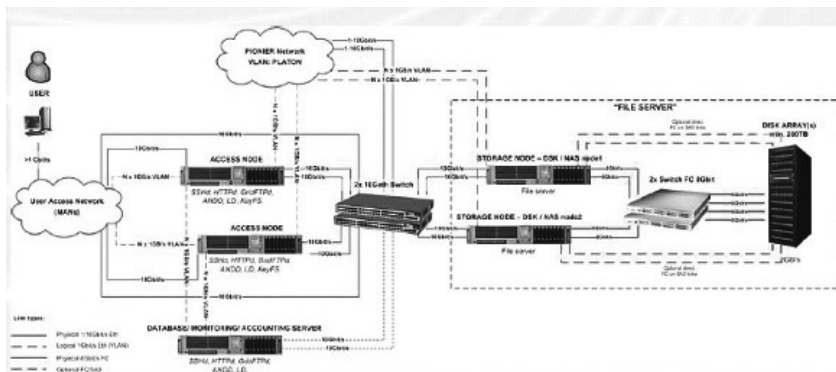


Rysunek 6. Schemat infrastruktury Usługi Powszechnej Archiwizacji

Źródło: M. Brzeźniak, *Ustuga...*, op. cit.

Infrastruktura Usługi Powszechnej Archiwizacji, w której wdrożona jest skalowalna architektura Krajowego Magazynu Danych pozwala na oferowanie usług przechowywania danych, które odpowiadają potrzebom użytkowników w zakresie pojemności systemu, wydajności składowania, trwałości danych w systemie, wiarygodności usługi, a także bezpieczeństwa i poufności danych.

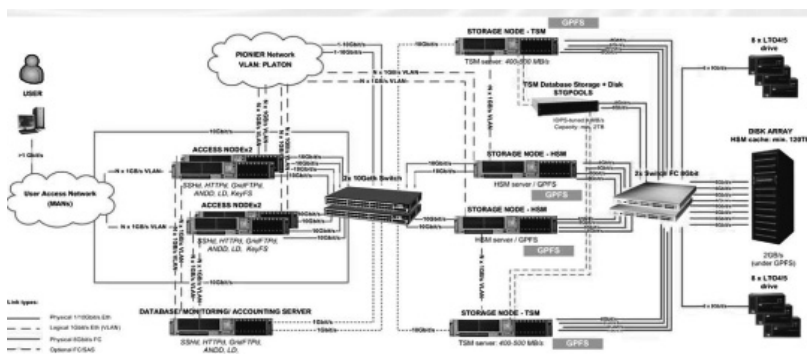
Dzięki zastosowaniu dyskowych i taśmowych technologii przechowywania, użytkownicy usługi mają do wyboru różne klasy usługi przechowywania związane z typem medium przechowującego dane, np. dyski i taśmy. Ponadto możliwe jest wykorzystanie przez użytkowników różnego poziomu (minimalna liczba i odległość geograficzna replik) oraz schematów replikacji. W architekturze KMD/PLATON-U4 obsługiwane są obydwie tryby replikacji: synchroniczna i asynchroniczna. Replikacja synchroniczna oferuje pewność dotyczącą stanu/spójności wszystkich replik przy dłuższym czasie odpowiedzi systemu, natomiast replikacja asynchroniczna gwarantuje pewność dotyczącą stanu/spójności tylko dla jednej repliki przy krótszym czasie odpowiedzi systemu.



Rysunek 7. Dyskowe węzły składowania

Źródło: R. Wyrzykowski, R. Mikołajczak, M. Zawadzki, *Niezawodne usługi outsourcingowe na przykładzie usług kampusowych i Krajowego Magazynu Danych w sieci PIONIER*, I Konferencja i3. Internet – Infrastruktury – Innowacje. Nauka w obliczu społeczeństwa cyfrowego, Poznań 4–6 listopada 2009.

Użytkownicy mogą skorzystać z funkcjonalności wirtualnego systemu plików lub aplikacji klienta kopii zapasowych i archiwizacji, co daje możliwość dostosowania usługi do wymagań funkcjonalnych środowiska i aplikacji użytkownika. Z kolei rozproszenie geograficzne systemu oraz jego dostępność poprzez sieć PIONIER i połączone z nią sieci miejskie (tzw. sieci MAN) pozwala na efektywne składowanie danych w systemie i uzyskiwanie do nich dostępu. Wielu docelowych użytkowników usługi podłączonych jest do wspomnianych sieci lub dysponuje łączami do sieci Internet umożliwiającymi wydajny dostęp do infrastruktury usługi przechowywania danych. Dotyczy to m.in. bibliotek cyfrowych.



Rysunek 8. Taśmowe węzły składowania

Źródło: R. Wyrzykowski, R. Mikołajczak, M. Zawadzki, *op. cit.*

Usługi powszechnej archiwizacji w ramach projektu PLATON są realizowane poprzez konsorcjum ośrodków superkomputerowych i sieciowych doświadczonych w dziedzinie przechowywania danych oraz świadczeniu zaawansowanych usług informatycznych. Posiadają odpowiednie zasoby, począwszy od fizycznych nośników danych, poprzez geograficznie rozproszoną infrastrukturę (wiele centrów danych wyposażonych w nowoczesny sprzęt do przechowywania danych) i szybką sieć PIONIER, łączącą elementy tej infrastruktury, po wysoce wykwalifikowaną kadrę. Usługi powszechnej archiwizacji pozwalają przenieść zadanie długoterminowego przechowywania i zabezpieczenia masowych ilości danych z instytucji naukowych czy akademickich na posiadające odpowiednią infrastrukturę konsorcjum PLATON.

Wyróżniające cechy systemu:

- replikacja synchroniczna i asynchroniczna
 - semi-synchroniczna replikacja metadanych;
- oddzielne (logicznie i fizycznie) przestrzenie dla użytkowników;
- wysoki stopień decentralizacji
 - niezależne metakatalogi.

Brak potrzeby specjalizowanego oprogramowania klienckiego.

Gdzie zamówić usługę?

W jednym z ośrodków biorących udział w projekcie:

- Białystok – Politechnika Białostocka, Centrum Komputerowych Sieci Rozległych, platon-u4-req@biaman.pl;
- Częstochowa – Politechnika Częstochowska, Instytut Informatyki Teoretycznej i Stosowanej, PCz, platon-u4-req@icis.pcz.pl;
- Gdańsk – Politechnika Gdańska, Centrum Informatyczne Trójmiejskiej Akademickiej Sieci Komputerowej TASK, platon-u4-req@task.gda.pl;
- Kielce – Politechnika Świętokrzyska, platon-u4-req@tu.kielce.pl;
- Kraków – Akademia Górniczo-Hutnicza, Akademickie Centrum Komputerowe CYFRONET AGH, platon-u4-req@cyfronet.pl;
- Lublin – Uniwersytet Marii Curie-Skłodowskiej w Lublinie lub MAN UMCS, platon-u4-req@umcs.lublin.pl;
- Łódź – Politechnika Łódzka, Centrum Komputerowe PŁ, Miejska Sieć Komputerowa LOD, platon-u4-req@man.lodz.pl;
- Poznań – Instytut Chemii Bioorganicznej PAN, Poznańskie Centrum Superkomputerowo-Sieciowe PCSS, platon-u4-req@man.poznan.pl;

- Warszawa – Uniwersytet Warszawski, Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, ICM, platon-u4-req@net.icm.edu.pl;
- Wrocław – Politechnika Wrocławska, Wrocławskie Centrum Sieciowo-Superkomputerowe, WCSS, platon-u4-req@kdm.wcss.wroc.pl.

Kroki rejestracyjne:

- wysłanie zgłoszenia na adres jednego z ośrodków biorących udział w projekcie;
- wypełnienie otrzymanego w odpowiedzi formularza rejestracji usługi i odesłanie go na adres kontaktowy;
- po poprawnym zweryfikowaniu formularza otrzymujemy informację z prośbą o dostarczenie swojego certyfikatu X509 lub odebranie certyfikatu wygenerowanego przez CA usługi archiwizacji;
- po zakończeniu wszystkich powyższych kroków otrzymujemy informację potwierdzającą proces zakończenia rejestracji i aktywację usługi.

Dane rejestracyjne obejmują:

- dane identyfikujące instytucję, z jakiej pochodzi użytkownik:
 - dane teleadresowe;
- dane identyfikujące użytkownika końcowego:
 - dane teleadresowe,
 - certyfikat użytkownika, certyfikat CA użytkownika;
- dane określające profil zamawianej usługi:
 - wielkość zasobów,
 - liczba replik,
 - dostępność danych itp.

Certyfikaty użytkowników:

- Użytkownicy korzystający z usługi muszą posiadać certyfikat X509.
- Certyfikat użytkownika może zostać wygenerowany i podpisany przez CA usługi archiwizacji w trakcie procedury rejestracyjnej.
- Użytkownik może dostarczyć swój certyfikat podpisany przez inne CA w trakcie procedury rejestracyjnej¹⁶ (ale wtedy konieczne jest dostarczenie bądź wskazanie miejsca, skąd można pobrać certyfikat CA podpisującego certyfikat użytkownika).

¹⁶ R. Mikołajczak, *Usługa Powszechnej Archiwizacji (zabezpieczanie i archiwizacja danych)*, XVI edycja seminarium z cyklu Archiwizacja Dokumentów Elektronicznych. Archiwizacja danych cyfrowych. Technologie, dobre praktyki, uwarunkowania prawne, Warszawa 19 kwietnia 2011.

Podsumowanie

Oprogramowanie KMD i Usługa Powszechnej Archiwizacji dają użytkownikom dostęp do skalowalnej i rozproszonej infrastruktury opartej na nowoczesnej technologii – nieosiągalnej dla większości instytucji. System został zaprojektowany tak, aby spełnić wymagania użytkowników w zakresie: bezpieczeństwa danych, wysokiej trwałości danych, niezawodności, prostoty użytkowania. System jest innowacyjny względem istniejących rozwiązań¹⁷.

Bibliografia

1. Białoskórski M., Balcerek B., *Narzędzia klienta usługi archiwizacji* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/online/pl/artykuly/1091/Materialy_archiwalne_z_warsztatow_Usluga_powszechnej_archiwizacji_juz_dostepne.html.
2. Brzeźniak M. i in., *Usługa powszechnej archiwizacji w sieci PIONIER i jej zastosowanie do składowania danych sieciowych*, I Konferencja i3. Internet – Infrastruktury – Innowacje. Nauka w obliczu społeczeństwa cyfrowego, Poznań 4–6 listopada 2009.
3. Brzeźniak M., *PLATON. Usługi powszechnej archiwizacji* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1374/PLATON_Uslugi_powszechej_archiwizacji.html.
4. Brzeźniak M., *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?brzezniaak>.
5. Bubel D., Szczygłowska L., Kuczyński Ł., *Biblioteki cyfrowe – bibliotekami uczestniczącymi*, Konferencja Polskie Biblioteki Cyfrowe 2011. VIII Warsztaty „Biblioteki cyfrowe”, Poznań 10–13 października 2011.
6. Bubel D., Szczygłowska L., *Bibliotekarz uczestniczący – człowiek orkiestra*, III Ogólnopolska Konferencja Naukowa: Bibliotekarze bez bibliotek, czyli bibliotekarstwo uczestniczące, Białystok 18–20 maja 2011.
7. Haba M., *Podstawowa instalacja i konfiguracja oprogramowania Bacula* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://www.bacula.pl/artkul/39/podstawowa-instalacja-i-konfiguracja-oprogramowania-bacula>.
8. Jankowski M., Brzeźniak M., *Architektura i mechanizmy systemu* [online], [dostęp: 25.04.2011], dostępny w World Wide Web: http://www.pionier.net.pl/online/pl/artykuly/1091/Materialy_archiwalne_z_warsztatow_Usluga_powszechnej_archiwizacji_juz_dostepne.html.
9. Kesselman M., Watstein S., *Creating Opportunities. Embedded Librarians*, „Journal of Library Administration” 2009, Vol. 49, Issue 4, s. 383–400.

¹⁷ M. Jankowski, M. Brzeźniak, *Architektura...*, *op. cit.*

10. Mikołajczak R., *Usługa Powszechnej Archiwizacji (zabezpieczenie i archiwizacja danych)*, XVI edycja seminarium z cyklu Archiwizacja Dokumentów Elektronicznych. Archiwizacja danych cyfrowych. Technologie, dobre praktyki, uwarunkowania prawne, Warszawa 19 kwietnia 2011.
11. Nahotko M., *Komunikacja naukowa w środowisku cyfrowym: globalna biblioteka cyfrowa w informatycznej infrastrukturze nauki*, Wydaw. SBP, Warszawa 2010.
12. PCSS. KMD. Krajowy Magazyn Danych [online], [dostęp: 25.04.2011], dostępny w World Wide Web: <http://kmd.pcoss.pl/ogolne.html>.
13. Shumaker D., *Breaking Out of the Box. The Promise of Embedded Librarianship* [online], [dostęp 25:04.2011], dostępny w World Wide Web: <http://archive.saoug.org.za/presentations/BreakingOutoftheBox.pdf>.
14. Siess J., *Embedded Librarians. Our Future Or Our Fear?* Annual Conference: Canadian Association of Law Libraries, Halifax, NS, 24-27 maja 2009.
15. Wyrzykowski R., Mikołajczak R., Zawadzki M., *Niezawodne usługi outsourcingowe na przykładzie usług kampusowych i Krajowego Magazynu Danych w sieci PIONIER*, I Konferencja i3. Internet – Infrastruktury – Innowacje. Nauka w obliczu społeczeństwa cyfrowego, Poznań 4-6 listopada 2009.

Dorota Witczak, Kamil Panaś,
Krzysztof Sobkowiak

Biblioteka Główna Uniwersytetu Pedagogicznego w Krakowie

W oczekiwaniu na PLATONa. Współpraca Pedagogicznej Biblioteki Cyfrowej z Akademickim Centrum Komputerowym Cyfronet AGH w zakresie długoterminowej archiwizacji danych

Słowa kluczowe: Pedagogiczna Biblioteka Cyfrowa, PLATON, długoterminowa archiwizacja obiektów cyfrowych, ochrona zasobów bibliotecznych

Abstrakt: Pedagogiczna Biblioteka Cyfrowa (PBC) stanęła przed problemem długoterminowej archiwizacji plików master obiektów cyfrowych. W artykule przedstawiono ewolucję sposobów archiwizacji zasobów PBC, począwszy od wykorzystania płyt CD i DVD, przez wprowadzenie technologii *Network Attached Storage* (NAS), po przechowywanie drugiego poziomu (*Secondary Storage*). Zaprezentowano działania podjęte wspólnie z Akademickim Centrum Komputerowym Cyfronet AGH, poprzedzające udostępnienie środowisku naukowemu usługi powszechnej archiwizacji w ramach projektu PLATON. Współpraca ta przyczynia się do wypracowania procedur i reguł bezpiecznego, uporządkowanego i długoterminowego przechowywania danych. Docelowo, w ramach projektu, dane PBC zgromadzone na serwerze Cyfronetu zostaną przemieszczone na urządzenia docelowe oraz zreplikowane w centrach superkomputerowo-sieciowych. Przedstawione rozwiązanie jest niskonakładową alternatywą opracowaną dla bibliotek cyfrowych borykających się z ograniczeniami infrastruktury i niewielkimi budżetami.

Wstęp

Problemem, z którym wcześniej czy później musi zmierzyć się każda instytucja posiadająca bibliotekę cyfrową jest uporządkowane, bezpieczne i długoterminowe przechowywanie zasobów cyfrowych. Podobnie jak w bibliotekach tradycyjnych, gdzie dbałość o księgozbiór wymaga nakładów finansowych na stworzenie optymalnych warunków długoterminowego przechowywania i dostępu do różnego typu fizycznych obiektów, jak np. książki, rękopisy, mapy, fotografie itp., tak w przypadku obiektów cyfrowych również należy stworzyć odpowied-

nie środowisko informatyczne i wypracować procedury ich właściwego przechowywania i dostępu do nich.

Warunki przechowywania zbiorów tradycyjnych i cyfrowych wymieniono w tabeli 1.

Tabela 1. Przechowywanie zbiorów tradycyjnych i zbiorów cyfrowych. Uwarunkowania.

Zbiory tradycyjne	Zbiory cyfrowe
Elastyczna powierzchnia magazynowa składowania zbiorów	Skalowalny system składowania danych (możliwość zwiększania pojemności)
Odpowiednia temperatura, wilgotność powietrza, oświetlenie	Klimatyzacja, zasilanie awaryjne, specjalistyczne systemy przeciwpożarowe
Półki, segregatory, kontenery	Opakowanie informacji (formaty cyfrowe)
Fizyczny monitoring	Rejestrowanie dostępu do danych (logi)
Ograniczenie do osób upoważnionych	Polityka bezpieczeństwa dostępu do danych
Drzwi, zamki, kraty	Hasła dostępu, tokeny, karty inteligentne, metody biometryczne, szyfrowanie
Kontrola księgozbioru	Kontrola możliwości odczytu i integralności danych
Egzemplarz archiwalny	Replikacja i delokalizacja danych
Konserwacja zbiorów	Migracja danych do nowych technologii, odtwarzanie danych z replik

Źródło: opracowanie własne.

W świadomości pracowników bibliotek utrwaliło się już wystarczająco mocno przekonanie o konieczności zapewnienia tradycyjnym zbiorom bibliotecznym właściwych warunków do przechowywania, tj. zabezpieczenie odpo-

wiednio pojemnej powierzchni magazynowej np. przez składowanie zwarte czy magazyny wysokiego składowania, optymalnych warunków klimatycznych (temperatura, wilgotność powietrza, natężenie oświetlenia), ochrony przeciwpożarowej oraz ograniczenia dostępu do osób upoważnionych.

W przypadku zbiorów cyfrowych nie ukształtowała się jeszcze w bibliotekach powszechna świadomość znaczenia tego problemu. Stąd brak jest wciąż ogólnie przyjętych procedur, które pozwalałyby na opracowanie i wdrożenie strategii długoterminowej archiwizacji danych (*Secondary Storage*) i zapewniłyby finansową i organizacyjną ciągłość dla systemów archiwizacji danych. Niektóre elementy organizacji i przechowywania zbiorów w magazynach tradycyjnych i cyfrowych będą się różnić, a niektóre pozostaną podobne (rysunek 1).



Rysunek 1. Magazyn zbiorów bibliotecznych i serwerownia

Źródło: fot. z lewej: Krzysztof Sobkowiak; fot. z prawej: Torkild Retvedt. Server room. Racks within a server room at CERN [online], [dostęp: 15.09.2012], dostępny w World Wide Web: <http://www.flickr.com/photos/torkildr/3462607995>.

Trwała archiwizacja danych cyfrowych w Pedagogicznej Bibliotece Cyfrowej

Zabezpieczanie danych w instytucjach najczęściej sprowadza się do wykonywania kopii zapasowych. Pojęcia tworzenie kopii zapasowych (*backup*) i archiwizacja danych (*archiving*), dotyczące zbiorów danych w określonym momencie czasu, są nierzadko traktowane jako synonimy. Często obie te funkcje są łączone ze sobą, zatem wydaje się, że mamy do czynienia z tym samym procesem. Różnica między kopią zapasową a archiwum danych

zaciera się jeszcze bardziej, gdy kopię zapasową przechowujemy w długim okresie. Taka sytuacja jest częściowo spowodowana niespójną terminologią używaną w heterogenicznym środowisku informatycznym, w którym w obrębie jednej organizacji wykorzystywane są zarówno zamknięte systemy Windows, jak i otwarte platformy Linux.

Kopie zapasowe są definiowane jako zrzuty zbioru danych wykonanych w określonym momencie czasu, zapisane w powszechnie dostępnym formacie i monitorowane przez cały czas wykorzystywania w systemie tworzącym kopie zapasowe¹. Natomiast archiwum danych jest oryginalnym zbiorem stałych danych przeniesionych, według określonych kryteriów, z jednej lokalizacji do innej, którego rolą jest długotrwałe przechowywanie danych na nośnikach odpornych na degradację w długim okresie czasu². Podobnie jak kopie zapasowe, tak i zarchiwizowane dane są rejestrowane i zarządzane w czasie, ale z uwagi na ich niezmiennosc, w dowolnym momencie czasu istnieje tylko jedna ich kopia. W bibliotekach cyfrowych danymi, które powinny być archiwizowane w pierwszej kolejności, są pliki *master* zdigitalizowanych obiektów. Umożliwi to w przyszłości nie tylko odtworzenie utraconych wersji prezentacyjnych obiektów cyfrowych, ale również utworzenie nowych, doskonalszych wersji w przypadku pojawienia się bardziej zaawansowanych technologii lub pozyskania pracowników o odpowiednio wysokich kwalifikacjach.

Pedagogiczna Biblioteka Cyfrowa, jak i podobne jej instytucje, stanęła przed problemem długoterminowej archiwizacji plików *master* obiektów cyfrowych. Rozwiązaniem okazało się nawiązanie współpracy administratorów PBC z Akademickim Centrum Komputerowym Cyfronet AGH podczas warsztatów, jakie przeprowadzono dla instytucji zainteresowanych wdrożeniem jednej z usług projektu Platformy Obsługi Nauki PLATON: Usługi Powszechnej Archiwizacji Danych (U-4), w ramach Programu Operacyjnego Innowacyjna Gospodarka, lata 2007–2013. Projekt PLATON jest kontynuacją wcześniej realizowanego projektu PIONIER – Polski Internet Optyczny – Zaawansowane Aplikacje, Usługi i Technologie dla Społeczeństwa Informatycznego. Projekt PIONIER został zrealizowany głównie w zakresie nowoczesnej sieci światłowodowej. Natomiast z braku środków nie wdrożono usług i aplikacji. Dopiero PLATON finalizuje wcześniejsze zamierzenia dzięki wdrażanym w ramach pierwszego etapu, pod nazwą „Kontener Usług Wspólnych”, pięciu usług³. Nazwa projektu – PLATON – jest nieprzypadkowa, ma się kojarzyć z greckim filozofem i ukazać inne spojrzenie na sieci komputerowe, jako na medium niosące nowe możliwości i usługi.

¹ S. Nelson, *Profesjonalne tworzenie kopii zapasowych i odzyskiwanie danych*, Gliwice 2012, s. 14–15.

² *Ibid.*

³ Projekt PLATON zakłada stworzenie i uruchomienie następujących usług: wideokonferencji, eduroam, kampusowych, archiwizacji, naukowej interaktywnej telewizji HD.

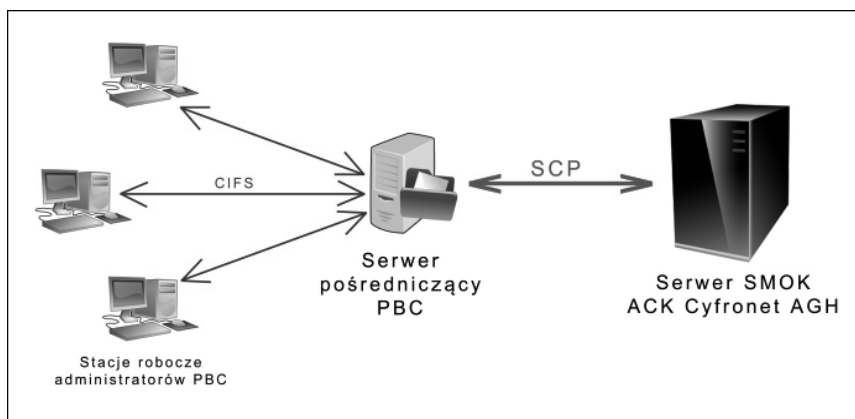
Usługą Powszechnej Archiwizacji Danych (U-4) realizowaną w ramach projektu PLATON zajmie się Konsorcjum⁴ specjalistycznych instytucji, które zapewni odpowiednią infrastrukturę sprzętową, oprogramowanie oraz wykwalifikowany personel. W dziesięciu miastach⁵ rozmieszczone zostaną węzły przechowywania danych oraz węzły usługowe, rozproszone geograficznie, połączone z optycznymi sieciami miejskimi oraz siecią krajową. Dane będą archiwizowane na żądanie w lokalizacji oddalonej geograficznie, z nieograniczonym w czasie dostępem online do archiwum, co w przypadku utraty cyfrowych zbiorów zapewni natychmiastowe ich odzyskanie. Usługa ta, będąca największym tego typu przedsięwzięciem ogólnopolskim, w założeniu ma zagwarantować bezpieczeństwo funkcjonowania jednostek akademickich w całym kraju. Skorzystają z niej nie tylko biblioteki cyfrowe, ale również całe środowisko akademickie, bezpiecznie przechowując np. wyniki eksperymentów naukowych, badań, materiałów edukacyjnych. W Pedagogicznej Bibliotece Cyfrowej posiadamy unikalne materiały edukacyjne, np. filmy z otwartych wykładów, multimedialne materiały dydaktyczne, dokumentacje fotograficzne, np. Kresów Wschodnich wykonane przez studentów Wydziału Sztuki Uniwersytetu Pedagogicznego podczas plenerów malarskich. Zgromadzone fotografie pochodzą z kilku ostatnich lat, a niektóre są już niestety jedynym świadectwem istnienia obiektów, które uległy trwałemu zniszczeniu.

Udostępnienie cyfrowego repozytorium przez Akademickie Centrum Komputerowe Cyfronet AGH spowodowało inne spojrzenie na stosowany model archiwizacji zbiorów Pedagogicznej Biblioteki Cyfrowej. Bezpośrednio po digitalizacji i obróbce materiału, cyfrowy dokument był archiwizowany na nośnikach CD bądź DVD i usuwany z dysku lokalnego komputera. Dotychczas stosowane rozwiązanie było wymuszone brakiem środków finansowych na zakup profesjonalnych urządzeń do archiwizacji oraz brakiem ogólnopolskiej polityki archiwizacji skierowanej przede wszystkim do środowiska akademickiego. Archiwizacja na nośnikach optycznych, oprócz bezdyskusyjnych problemów z trwałością zapisu i późniejszego odczytu, jest czasochłonna i pracochłonna, a stosowana w długim okresie także kosztowna (rosnąca powierzchnia magazynowa, odświeżanie nośników co kilka lat itp.). Nośniki optyczne traktowano więc w PBC jako rozwiązanie doraźne i poszukiwano alternatywnej, taniej i bezpiecznej metody archiwizacji materiałów cyfrowych.

⁴ Konsorcjum PIONIER powstało 25 października 2003 r. w Kazimierzu Dolnym nad Wisłą. Jest porozumieniem zawartym pomiędzy 22 jednostkami MAN (*Metropolitan Area Network*) i KDM (centra Komputerów Dużej Mocy).

⁵ Białystok, Częstochowa, Gdańsk, Kielce, Kraków, Lublin, Łódź, Poznań, Warszawa, Wrocław.

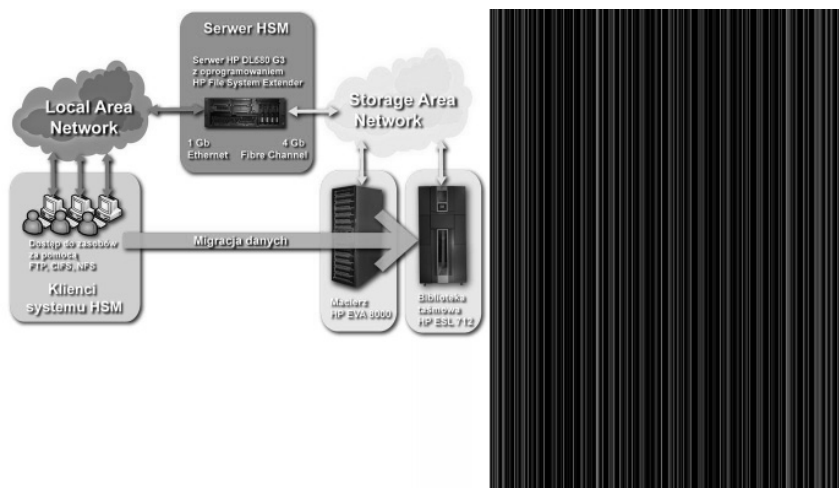
Restrykcyjne zasady bezpieczeństwa ACK Cyfronet AGH umożliwiają dostęp do repozytorium jedynie autoryzowanym użytkownikom posiadającym stosowne certyfikaty. Kontrolę nad przesyłaniem danych archiwizowanych na serwerze *smok* w Cyfronecie sprawują więc administratorzy PBC po uzyskaniu wymaganych uprawnień. Początkowo dane przesyłano ręcznie, aktualnie proces transferu plików przebiega automatycznie. Dla usprawnienia działań zespołu PBC oraz tymczasowego składowania efektów jego pracy, wynikających z pełnionych obowiązków, wprowadzono serwer pośredniczący. Wartością dodaną takiego rozwiązania jest odseparowanie dwóch warstw replikacji: lokalnej oraz docelowej. Implementacja usługi pośredniczącej umożliwia dodatkowo modyfikację archiwizowanych zasobów przez lokalnych użytkowników systemu przed ich ostatecznym zdeponowaniem w repozytorium docelowym przez osoby posiadające stosowne certyfikacje. Niewątpliwą zaletą przyjętego rozwiązania jest możliwość rozbudowy systemu o nowe stanowiska komputerowe bez konieczności występowania o dodatkowe certyfikaty bezpieczeństwa w ACK Cyfronet AGH.



Rysunek 2. Przepływ danych do archiwizacji w PBC

Źródło: opracowanie własne

Rolę zasobu pośredniczącego w przechowywaniu i udostępnianiu roboczych zbiorów cyfrowych pełni serwer pracujący w oparciu o technologię *Network Attached Storage* (NAS).



Rysunek 3. Panel administracyjny – FreeNAS

Źródło: zrzut ekranowy systemu FreeNAS

Przyjęte rozwiązanie umożliwia kontrolowany dostęp do danych znajdujących się w jednym miejscu z różnych punktów sieci. Ponadto implementacja kilkunastu protokołów sieciowych daje dostęp do zasobów bez względu na rodzaj zainstalowanego systemu operacyjnego klienta. Polityka bezpieczeństwa zastosowana w środowisku NAS precyzuje nadawanie ściśle określonych uprawnień do poszczególnych zbiorów archiwizowanych na platformie. W celu zwiększenia bezpieczeństwa przechowywanych danych omawiane środowisko oferuje archiwizację na nośnikach pamięci masowej w oparciu o technologię RAID⁶, która umożliwia łączenie wielu dysków w macierz systemowe tworzące jeden dysk logiczny. Macierz dyskowa na serwerze pośredniczącym jest skonfigurowana w oparciu o system RAID 1 (zwany także lustrzanym, ang. *mirroring*) – dane są równolegle zapisywane na dwóch dyskach – awaria jednego z dysków nie powoduje utraty archiwizowanych danych.

Wdrożenie prezentowanej usługi przebiegało w kilku etapach. Z uwagi na skromny budżet PBC, zastosowano darmowe oprogramowanie oraz wykorzystano komputer klasy Pentium 4. W pierwszej kolejności zmodernizowano przyszły serwer, montując w nim kontroler RAID wraz z dwoma dyskami o pojemności 1 TB każdy. Następnie zainstalowano na maszynie

⁶ RAID (ang. *Redundant Array of Independent Disks* – nadmiarowa macierz niezależnych dysków) to układ współpracujących dwóch lub więcej dysków twardych w celu zwiększenia wydajności lub zabezpieczenia zapisanych danych.

system FreeNAS⁷ dystrybuowany na licencji FreeBSD⁸ oraz przeprowadzono podstawową konfigurację, włączając m.in. protokół CIFS (ang. *Common Internet File System*)⁹ – typu *klient-serwer* – udostępniający zasoby serwera lokalnym stacjom roboczym, pracującym pod kontrolą systemów Windows. W kolejnym etapie utworzono na serwerze konta i grupy użytkowników oraz stworzono system kartotek. Przyjęta struktura drzewa umożliwia szybki i łatwy dostęp do poszczególnych zasobów określonym grupom oraz implikuje określone czynności użytkowników systemu (*workflow*).

↑ Nazwa	Roz.	Wielkość
[1_DIGITALIZACJA]		<DIR>
[2_ARCHIWIZACJA]		<DIR>
[3_PBC]		<DIR>
[4_REDAKTORZY]		<DIR>
[COMMON]		<DIR>

Rysunek 4. Katalogi główne stosowane w workflow w PBC

Źródło: fragment zrzutu ekranowego protokołu CIFS.

W każdym katalogu znajdują się imienne podkatalogi użytkowników odpowiedzialnych za etapy procesu digitalizacji, archiwizacji oraz opracowania zbiorów PBC. System uprawnień przypisany poszczególnym użytkownikom umożliwia im dostęp tylko do określonych katalogów. W końcowym etapie zaimplementowano skrypty pozwalające na zautomatyzowany transfer plików *master* z macierzy NAS na serwer ACK Cyfronet AGH.

Obecnie dane przeznaczone do archiwizacji w ACK Cyfronet AGH przesyłane są poprzez sieć MAN z użyciem bezpiecznego protokołu SCP¹⁰ i trafiają na serwer dostępowy *smok*¹¹. Przestrzeń do składowania danych udostępniona jest za pośrednictwem wysokowydajnej sieci SAN¹², złożonej z macierzy dyskowych HP EVA 8000 w konfiguracji: 240 dysków FATA 500GB 7 200 rpm, 8GB pamięci cache, 8 portów Fibre Channel. Macierze wyposażone są w wewnętrzne mechanizmy kontroli integralności danych wraz z ich zabezpieczeniem na wypadek awarii komponentów sprzętowych.

⁷ FreeNAS jest darmowym serwerem NAS (ang. Network Attached Storage).

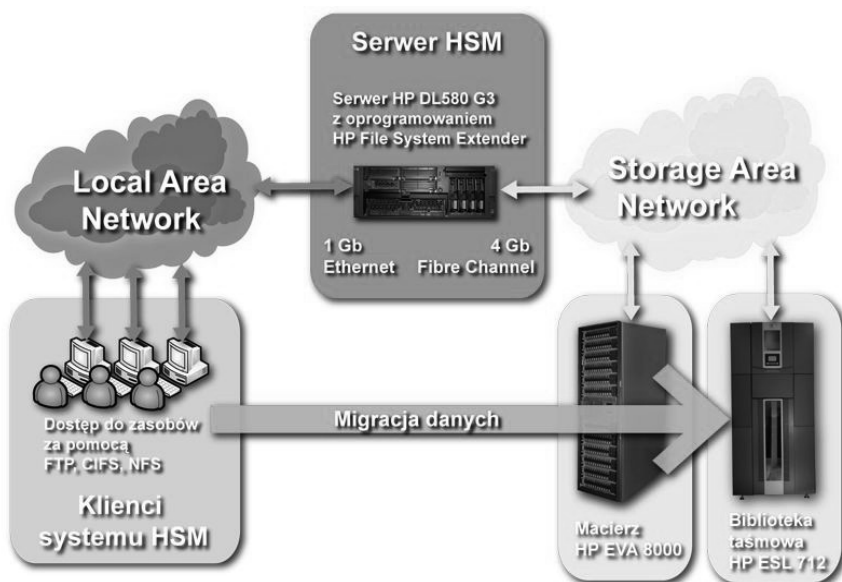
⁸ FreeBSD – system operacyjny z rodziny UNIX.

⁹ CIFS – protokół udostępniania zasobów komputerowych.

¹⁰ SCP (ang. *Secure CoPy*) – protokół bezpiecznego kopiowania plików.

¹¹ *Smok* jest maszyną firmy Hewlett-Packard ProLiant DL580 G3, wyposażoną m.in. w 2 procesory Intel Xeon 3,3 GHz, 4GB pamięci RAM, 2 karty Fibre Channel oraz dysk twardy 72GB o prędkości 10 000 rpm, pracującą pod kontrolą systemu operacyjnego Linux RedHat Enterprise z oprogramowaniem HSM HP File System Extender.

¹² SAN (ang. *Storage Area Network*) – sieć pamięci masowej.



Rysunek 5. Schemat przepływu danych z PBC wewnątrz infrastruktury ACK Cyfronet AGH

Źródło: http://cyfronet.pl/uslugi_obliczeniowe/?a=unitree/dabd_storage [dostęp: 15.09.2012].

Zgodnie z przyjętymi wspólnie z ACK Cyfronet AGH ustaleniami, wielkość wolumenu przeznaczanego na potrzeby Pedagogicznej Biblioteki Cyfrowej początkowo wynosiła 1 TB (*soft quota*), z możliwością rozszerzania w miarę potrzeb, a dane surowe były i są chronione technologią RAID 5. Operacje zapisu i odczytu na wolumenie dedykowanym PBC możliwe są poprzez zasób przypisany do naszego konta użytkownika, zamontowany na serwerze HSM *smok*.

Oprócz zabezpieczeń macierzowych, dane chronione są poprzez cykliczne, przyrostowe kopie zapasowe. Przebiegiem procesów wykonywania i kontroli kopii zapasowych zarządza oprogramowanie HP Data Protector. Codzienne kopie bezpieczeństwa wykonywane są w czasie dogodnych okien backupowych, a ponadto za pośrednictwem sieci SAN kierowane są do biblioteki taśmowej¹³ i zapisywane na taśmy magnetyczne w technologii LTO-4¹⁴. Przyjęta w Cyfronecie polityka kopii bezpieczeństwa przewiduje

¹³ Biblioteka taśmowa jest automatem składającym się m.in. z jednego lub więcej napędów do zapisu/odczytu taśm, magazynków na kasety z taśmami, serwo mechanizmu realizującego zaprogramowany transport kaset z magazynku do napędu i z powrotem oraz systemu identyfikacji kaset.

¹⁴ LTO (ang. *Linear Tape Open*) – standard zapisu danych na taśmach magnetycznych. LTO-4 jest wersją tego standardu z roku 2007.

przechowywanie jednej wersji aktywnej oraz wersji nieaktywnej przez 16 tygodni. Taka konfiguracja w wystarczającym stopniu spełnia większość wymagań skutecznej i bezpiecznej archiwizacji danych.

Usługa Powszechnej Archiwizacji Danych (U-4) dobiega końca fazy testowej, więc niebawem rozpocznie się migracja danych PBC pomiędzy systemami. Zbiory Pedagogicznej Biblioteki Cyfrowej przechowywane tymczasowo w Cyfronecie zostaną przemiegrowane przez pracowników Działu Archiwizacji i Bezpieczeństwa Danych na docelowe urządzenia oraz zreplikowane w innych centrach superkomputerowo-sieciowych w ramach projektu PLATON.

Podsumowanie

Przedstawione rozwiązanie jest niskonakładową alternatywą opracowaną dla bibliotek cyfrowych borykających się z ograniczeniami infrastruktury i skromnymi budżetami. Zaprezentowany model byłby optymalnym rozwiązaniem, gdyby serwer pośredniczący zastąpiono profesjonalną macierzą dyskową dającą możliwość skalowania w miarę potrzeb.

Bibliografia

1. Brzeźniak M., *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych* [online], [dostęp: 15.09.2012], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?brzeznia>.
2. Januszko-Szakiel A., *Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek*, „Przegląd Biblioteczny” 2011, R. 79, z. 2, s. 21–46.
3. Januszko-Szakiel A., *Nowy wymiar zabezpieczania elektronicznych zasobów* [online], [dostęp: 15.09.2012], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?januszko>.
4. Januszko-Szakiel A., *Wiarygodność archiwów cyfrowych*, „Przegląd Biblioteczny” 2009, R. 79, z. 3, s. 325–347.
5. Januszko-Szakiel A., *Długoterminowa archiwizacja zasobów cyfrowych w świadomości pracowników polskich bibliotek* [online], „Polskie Biblioteki Cyfrowe” 2009, s. 91–101 [dostęp: 15.09.2012], dostępny w World Wide Web: <http://lib.psnc.pl/Content/370/11-Januszko-ER.pdf>.
6. Lorie R., *Long-term archiving of digital information* [online], [dostęp: 15.09.2012], dostępny w World Wide Web: [http://domino.watson.ibm.com/library/CyberDig.nsf/papers/BE2A2B188544DF2C8525690D00517082/\\$File/RJ10185.pdf](http://domino.watson.ibm.com/library/CyberDig.nsf/papers/BE2A2B188544DF2C8525690D00517082/$File/RJ10185.pdf).

7. Nelson S., *Profesjonalne tworzenie kopii zapasowych i odzyskiwanie danych*, Wydawnictwo Helion, Gliwice 2012.
8. „PIONIER Magazine” 2011, Nr 1(06) [online] [dostęp: 15.09.2012], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1365/PLATON_specjalnie.html.
9. Platforma Obsługi Nauki PLATON [online], Polski Internet Optyczny PIONIER [dostęp: 15.09.2012], dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.

Cezary Mazurek, Tomasz Parkoła, Marcin Werla

Poznańskie Centrum Superkomputerowo-Sieciowe

dArceo: Usługi długoterminowego przechowywania danych źródłowych

Słowa kluczowe: długoterminowa archiwizacja, długoterminowe przechowywanie danych, PLATON, model OAIS, dArceo

Abstrakt: celem niniejszego artykułu jest przybliżenie efektów prac badawczo-rozwojowych, które realizowane są przez Poznańskie Centrum Superkomputerowo-Sieciowe w ramach projektu SYNAT, a dotyczą opracowania dArceo – systemu usług sieciowych dedykowanego dla długoterminowego przechowywania danych źródłowych. Usługi te oferują szerokie spektrum funkcji, od bezpiecznego składowania i zaawansowanego dostarczania wysokiej jakości danych źródłowych na potrzeby uprawnionych użytkowników końcowych, poprzez konwersję danych w celu budowania bibliotek i repozytoriów cyfrowych, aż po migrację danych na potrzeby długoterminowego przechowywania, w zgodzie z modelem OAIS. Usługi te opracowano z uwzględnieniem uznanych i powszechnie wykorzystywanych formatów, standardów i narzędzi, w tym: METS, PREMIS, PRONOM/UDFR, DROID, FITS oraz FFMpeg. W zakresie składowania danych uwzględniona została możliwość zapewnienia wsparcia dla różnych systemów i technologii, w szczególności przygotowano komponenty wykorzystujące Usługę Powszechnej Archiwizacji (U4) opracowywaną w ramach projektu PLATON.

Wstęp

Gwałtowny rozwój bibliotek cyfrowych w Polsce zapoczątkowany został uruchomieniem w 2002 roku Wielkopolskiej Biblioteki Cyfrowej, zrealizowanej z wykorzystaniem oprogramowania dLibra¹. W następnych latach pojawiały się kolejne biblioteki cyfrowe, których główną misją było zapewnienie łatwego dostępu do cyfrowej postaci zbiorów, często reprezentujących unikatowe kolekcje². Istotnym eta-

¹ A. Swędryński, M. Górny i in., Zastosowanie oprogramowania dLibra do budowy Wielkopolskiej Biblioteki Cyfrowej, [w:] *Internet w bibliotekach II. Łączność, współpraca, digitalizacja*, (Międzynarodowa Konferencja EBIB, Wrocław 23–26 września 2003), [online] Repozytorium Instytucjonalne PCSS [dostęp: 15.10.2012], dostępny w World Wide Web: <http://lib.psnc.pl/Content/8/ebib2003.pdf>.

² T. Parkoła, C. Mazurek, M. Werla, Rozwój bibliotek cyfrowych w Polsce, [w:] *Partnerzy bibliotek. Model komunikacji z otoczeniem*, red. I. Jurczak, E. Okularczyk, Łódź 2009, s. 239–246.

pem rozwoju tej infrastruktury było następnie powstanie Federacji Bibliotek Cyfrowych, która gromadzi i udostępnia obecnie dane opisujące zbiory (metadane) z ponad 80 różnych serwisów bibliotek cyfrowych³. Digitalizacja i udostępnienie tak olbrzymiego zasobu dziedzictwa online to efekt ogromnego wysiłku kilkuset polskich instytucji nauki i kultury. Udostępnianie zbiorów w postaci cyfrowej w Internecie jest niezwykle istotne dla odbiorców końcowych, ale nie należy zapominać też o kluczowej, dokumentacyjnej roli cyfrowych postaci zbiorów. Zarówno ze względu na udostępnianie, jak i na aspekty dokumentacyjne, nieodzowne jest przechowywanie wysokiej jakości postaci cyfrowych obiektów w długiej perspektywie czasu, bez względu na zachodzące zmiany formatów plików, sprzętu czy oprogramowania. Choć odpowiedzialne za masową digitalizację instytucje podejmują działania w kontekście bezpiecznej archiwizacji danych, badania ankietowe przeprowadzone w ramach projektu SYNAT⁴ pokazują, że nie ma obecnie w Polsce stosowanych rozwiązań, które wspomagają realizację idei długoterminowego przechowywania danych źródłowych zgodnie z uznanym na świecie modelem OAIS⁵.

Celem niniejszego artykułu jest przybliżenie efektów prac badawczo-rozwojowych, które realizowane są przez Poznańskie Centrum Superkomputerowo-Sieciowe (PCSS) w ramach projektu SYNAT i stanowią odpowiedź na oczekiwania związane z długoterminowym przechowywaniem danych źródłowych. Opracowywane usługi oferują szeroki zbiór funkcji, włączając w to migrację danych w zgodzie z modelem OAIS na potrzeby długoterminowego przechowywania, konwersję danych w celu budowania bibliotek i repozytoriów cyfrowych oraz zaawansowane dostarczanie danych źródłowych na potrzeby użytkowników końcowych. Opracowane rozwiązanie oferuje również dodatkowe mechanizmy zarządzania przechowywanymi obiektami. Są to przede wszystkim usługi związane z przechowywaniem danych źródłowych oraz ich monitorowaniem (w tym automatyczną okresową weryfikacją ich spójności i oceną ryzyka utraty możliwości odczytu danych na skutek wyjścia z użycia danego formatu zapisu), przekształcaniem danych (w tym również możliwość współdzielenia usług przekształcania danych pomiędzy instytucjami), a także wykonywaniem złożonych planów migracji i konwersji danych.

³ A. Lewandowska, C. Mazurek, M. Werla, Federacja Bibliotek Cyfrowych w sieci PIONIER. Dostęp do otwartych bibliotek cyfrowych i repozytoriów, [w:] *Open Access. Internet w bibliotekach*, (IV Ogólnopolska Konferencja EBIB, Toruń 7–8 grudnia 2007), [online] EBIB 2007 [dostęp: 15.10.2012], dostępny w World Wide Web: http://www.ebib.info/publikacje/mat-konf/mat18/lewandowska_mazurek_werla.php.

⁴ Projekt SYNAT finansowany jest przez Narodowe Centrum Badań i Rozwoju (nr umowy: SP/1/1/77065/10).

⁵ Referencyjny model OAIS (ang. Open Archival Information System) [online] Consultative Committee for Space Data Systems [dostęp: 15.10.2012], dostępny w World Wide Web: <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

Wyżej wymienione usługi uwzględniają uznane i powszechnie wykorzystywane formaty, standardy i narzędzia (w tym: METS, PREMIS, PRONOM/UDFR, DROID, FITS oraz FFMpeg). W zakresie składowania danych uwzględniona została możliwość zapewnienia wsparcia dla różnych systemów i technologii, w szczególności przygotowano komponenty wykorzystujące Usługę Powszechnej Archiwizacji (U4) opracowywaną w ramach projektu PLATON⁶. Usługa ta zapewnia bezpieczną archiwizację dużych ilości danych, włączając w to automatyczną replikację danych pomiędzy rozproszonymi geograficznie repozytoriami, i przeznaczona jest przede wszystkim dla instytucji naukowych przyłączonych do sieci Polski Internet Optyczny PIONIER.

Opisany powyżej pakiet oprogramowania został udostępniony pod nazwą dArceo. Oferuje on kompleksowe rozwiązanie w zakresie długoterminowego przechowywania obiektów cyfrowych, ze szczególnym uwzględnieniem zbiorów dziedzictwa kulturowego, w tym dokumentów tekstowych, obrazów oraz dokumentów audiowizualnych⁷. Pakiet ten jest ponadto przygotowany do współpracy z innymi modułami oprogramowania PCSS, takimi jak: dLibra, dMuseion i dLab, dzięki czemu możliwe jest wsparcie pełnego przepływu pracy nawet w dużych instytucjach realizujących złożone projekty digitalizacyjne.

Poniżej opisano usługi dArceo – ich funkcje i podstawowe charakterystyki. W ostatniej części artykułu przedstawiono mechanizm współdzielenia publicznie dostępnych usług przekształcania danych w dArceo. Artykuł zakończony jest podsumowaniem.

Charakterystyka i funkcje usług dArceo

System dArceo złożony jest z szeregu usług współpracujących ze sobą w celu zapewnienia opisanej wcześniej funkcjonalności długoterminowego przechowywania danych źródłowych. W skład usług dArceo wchodzi (w nawiasach podano oznaczenie kodowe usług używane w dalszej części artykułu): Zarządca Magazynu Danych (ZMD), Usługi Przekształcania Danych (UPD), Rejestr Usług (RU), Monitor Danych Źródłowych (MDZ), Monitor Systemu (MS), Zarządca Migracji i Konwersji Danych (ZMKD), Zarządca Uprawnień (ZU) oraz Zarządca Powiadomień (ZP). Rysunek 1 przedstawia schemat architektury logicznej pojedynczego wdrożenia dArceo, obrazując powiązania i współdziałanie poszczególnych usług.

⁶ Projekt PLATON finansowany jest w ramach Programu Operacyjnego Innowacyjna Gospodarka, lata 2007–2013; Projekt PLATON [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.

⁷ T. Kuipers, J. van der Hoeven, *Insight into digital preservation of research output in Europe. Survey report* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.



Rysunek 1. Schemat architektury logicznej pojedynczego wdrożenia dArceo

Źródło: Poznańskie Centrum Superkomputerowo-Sieciowe.

W ramach przedstawionej na diagramie architektury, dwie usługi mają charakter bazowy i są niezbędne do uruchomienia każdej instancji dArceo. Są to usługi ZU oraz ZP. Usługi te odpowiedzialne są za uwierzytelnianie i autoryzację w dostępie do przechowywanych zasobów oraz poszczególnych usług systemu, a także umożliwiają asynchroniczną komunikację pomiędzy usługami dArceo. Pozostałe usługi mają charakter funkcjonalny, tzn. ich zadaniem jest realizacja poszczególnych funkcji dArceo.

Każda usługa dArceo posiada odpowiedni interfejs komunikacyjny w postaci interfejsu programistycznego REST i/lub interfejsu użytkownika, a także opcjonalnie interfejsu służącego do jej konfiguracji (np. pliki konfiguracyjne, baza danych). W ramach wyszczególnionych usług o charakterze funkcjonalnym, umiejscowienie jednej usługi nad drugą oznacza jej zależność od usługi znajdującej się poniżej, przy czym relacja ta jest przechodnia, tj. usługa będąca wyżej w hierarchii zależy od wszystkich poniżej, jednak nie zależy od usług powyżej lub usług na tym samym poziomie. Zatem pojedyncza instancja systemu dArceo nie musi posiadać wszystkich usług z grupy usług o charakterze funkcjonalnym, natomiast ich wybór musi być zgodny z ograniczeniami nałożonymi przez zależności. Poniżej scharakteryzowano poszczególne usługi dArceo.

Zarządca Magazynu Danych

ZMD jest kluczowym elementem systemu dArceo, który przede wszystkim zapewnia funkcjonalność dostępu do danych źródłowych, ich modyfikacji, składowania oraz wersjonowania. Dane przechowywane w ZMD składowane są w tzw. magazynie danych, który jest elementem wykorzystywanym przez system dArceo. W tym zakresie ZMD umożliwia integrację z różnymi magazynami danych, takimi jak dyski sieciowe czy specjalizowane systemy archiwizacji danych (systemy taśmowe, systemy BlueRay). Na rysunku 1 w bloku ZMD wyodrębniono tzw. adapter magazynu danych, który jest elementem pośredniczącym w komunikacji pomiędzy dArceo a magazynem danych, w którym faktycznie składowane są obiekty poddane długoterminowemu przechowywaniu. Warto podkreślić, że mechanizmy ZMD przygotowano w taki sposób, aby mogły współpracować z magazynami danych, których czas odpowiedzi na żądanie zapisu lub odczytu danych może być długi (np. kilkadziesiąt minut), takimi jak wspomniana wcześniej Usługa Powszechnej Archiwizacji (U4) projektu PLATON⁸. Z tego względu komunikacja z ZMD następuje w sposób asynchroniczny.

Poza składowaniem i dostępem do danych źródłowych, ZMD zapewnia ekstrakcję metadanych technicznych z poszczególnych jego elementów. W tym celu usługa ZMD wykorzystuje odpowiednio przygotowane narzędzie FITS, które ekstrahuje metadane techniczne do formatów TextMD, DocumentMD, MIX, AES-X098B oraz VideoMD, w zależności od formatu pliku źródłowego. Warto zaznaczyć, że moduł ZMD automatycznie konstruuje i przechowuje dla każdego obiektu cyfrowego metadane techniczno-administracyjne w formacie METS oraz pełną historię operacji i migracji przechowywanego obiektu w formacie PREMIS.

Moduł ZMD wyposażony jest również we wbudowane repozytorium OAI-PMH (dalej: RO), które odpowiada za dostarczenie funkcjonalności interfejsu OAI-PMH do metadanych opisujących przechowywane dane źródłowe (podstawowym formatem metadanych udostępnianym przez RO jest format METS). Takie podejście pozwala zewnętrznym systemom na monitorowanie w sposób ustandaryzowany zmian zachodzących w ramach dArceo oraz regularne i przyrostowe pobieranie informacji o przechowywanych zasobach.

⁸ M. Brzeźniak i in., *Popular Backup/Archival Service and its Application for the Archival of the Network Traffic in the PIONIER Academic Network*, „Computational Methods in Science and Technology” 2010, wydanie specjalne, s. 109–118.

Usługi Przekształcania Danych

UPD są kluczowym elementem systemu dArceo w kontekście długoterminowego przechowywania danych źródłowych. UPD odpowiadają za dostarczenie funkcjonalności przekształcania danych, w tym migracji, konwersji i zaawansowanego dostarczania danych. Usługi te oferują spójny, jednolity interfejs komunikacyjny, dzięki czemu mogą być współdzielone między różnymi instancjami dArceo. Omówiono to szerzej w dalszej części artykułu. Ponadto możliwość składania poszczególnych usług w ciąg wywołań pozwala na dostarczenie bardziej zaawansowanych i kompleksowych przekształceń. dArceo definiuje trzy rodzaje usług należących do UPD:

- Usługi Migracji Danych (UMD) udostępniają funkcjonalność migracji danych, tj. zmiany bieżącego formatu danych/metadanych na inny format bez utraty informacji (konwersja bezstratna, migracja typu transformacja). Ta kategoria usług została wyodrębniona ze względu na jej charakter – dotyczy danych źródłowych (treści oraz metadanych kopii *master* danego obiektu) w perspektywie długoterminowego przechowywania.
- Usługi Konwersji Danych (UKD) udostępniają funkcjonalność dokonywania konwersji danych, tj. przekształcania danych/metadanych, mogącego wiązać się z utratą informacji (np. konwersja stratna, zmiana rozdzielczości, ekstrakcja fragmentu obiektu). Pojęcie konwersji odnosi się m.in. do przygotowywania wersji prezentacyjnych obiektów cyfrowych (lub ich fragmentów) w celu ich efektywnej prezentacji w Internecie.
- Usługi Zaawansowanego Dostarczania Danych (UZDD) udostępniają funkcjonalność dostarczania użytkownikowi końcowemu danych źródłowych w efektywny sposób. Usługi te są więc swoistym pośrednikiem pomiędzy użytkownikiem a magazynem danych. UZDD mogą np. pozwalać na serwowanie fragmentów większych obrazów, przysyłać materiały audiowizualne (np. przy wykorzystaniu właściwości pobierania progresywnego) lub wyświetlać obrazy na urządzeniach mobilnych.

Warto podkreślić, że system dArceo jest przygotowany w taki sposób, aby nowe usługi UPD mogły być dodawane do systemu. Dzięki temu z biegiem czasu możliwe jest wprowadzanie nowych migracji/konwersji (np. do formatów, które obecnie nie zostały jeszcze opracowane) czy usług zaawansowanego dostarczania danych. Możliwość rozszerzenia systemu dArceo w tym kontekście jest kluczowa z perspektywy zapewnienia długoterminowej dostępności przechowywanych obiektów.

Rejestr Usług

RU odpowiada za dostarczenie funkcjonalności rejestrowania usług UMD, UKD i UZDD w systemie dArceo. RU przechowuje i udostępnia informacje o możliwych do wykorzystania usługach przekształcania danych. Każda usługa w ramach RU opisana jest z wykorzystaniem technologii semantycznych, za pomocą ontologii OWL-S. Ponadto RU pochodzące z różnych instancji dArceo mogą się ze sobą synchronizować, tzn. wymieniać informacjami na temat publicznie dostępnych usług migracji, konwersji czy zaawansowanego dostarczania danych źródłowych. Efektem tego może być współdzielona przestrzeń (chmura) usług przekształcania danych, opisana szczegółowo w dalszej części artykułu.

Monitor Danych Źródłowych

MDZ odpowiedzialny jest za bieżące monitorowanie danych źródłowych w celu dokonania:

- oceny zdatności obiektów cyfrowych do użytku (integralność i spójność danych),
- oceny zdatności obiektów cyfrowych do długotrwałego przechowywania (ryzyko utraty informacji ze względu na wygaśnięcie wsparcia dla danego formatu).

Integralność i spójność danych weryfikowane są z wykorzystaniem sum kontrolnych SHA-512. W przypadku magazynów danych, które natywnie wspierają tego typu mechanizmy (np. PLATON U4), istnieje możliwość wyłączenia weryfikacji sum kontrolnych na poziomie dArceo, aby nie powielać operacji dokonywanych na niższych warstwach złożonego systemu. Zdatność obiektów cyfrowych określana jest na bazie informacji pochodzących z międzynarodowego rejestru UDFR, przede wszystkim jest to informacja o wycofaniu formatu z użytkowania lub istnieniu następczej wersji danego formatu w dłuższej perspektywie czasowej.

Monitor Systemu

MS odpowiada za dostarczanie funkcjonalności raportowania o stanie systemu i składowanych w nim danych. Dostępne są statystyki na poziomie całego systemu oraz w kontekście poszczególnych użytkowników dArceo. Statystyki podstawowe obejmują m.in. informacje o liczbie obiektów oraz

plików (danych i metadanych) wraz z ich rozmiarami. Statystyki dotyczące formatów metadanych i formatów plików obejmują m.in. informacje o liczbie obiektów lub plików danych, które w danym formacie są zapisane lub posiadają informacje. Ponadto MS udostępnia informacje o aktualnie funkcjonujących w ramach danej instancji dArceo usługach przekształcania danych oraz bieżącej konfiguracji MDZ.

Zarządca Migracji i Konwersji Danych

ZMKD odpowiada za dostarczenie funkcjonalności zaawansowanego migrowania danych i ich konwersji. Pozwala on na określenie planu migracji/konwersji, który zawiera reguły dotyczące zakresu migrowanych/konwertowanych danych oraz sposobu ich migracji/konwersji. Przyjmuje zlecenie wykonania planu i zarządza jego wykonaniem, tj. zleca innym usługom wykonanie odpowiednich części procesu i zbiera informacje o stanie wykonania zadania. Przykładem planu jest zlecenie migracji wszystkich plików TIFF-8 na pliki o formacie TIFF-10.

Migracja oraz konwersja danych mogą być realizowane przez usługę ZMKD, przy założeniu, że usługa ta została odpowiednio skonfigurowana, poprzez dodanie adekwatnego planu migracji lub konwersji. W celu przeprowadzenia migracji lub konwersji danych usługa ZMKD dobiera z RU konkretną usługę lub kilka usług migracji/konwersji danych. Następnie ZMKD steruje dobranymi usługami w celu wykonania zleconego planu migracji lub konwersji.

Zarządca Uprawnień

ZU jest odpowiedzialny za dostarczenie funkcjonalności zarządzania uprawnieniami użytkowników do poszczególnych elementów systemu, w szczególności obiektów źródłowych, czy poszczególnych usług dArceo.

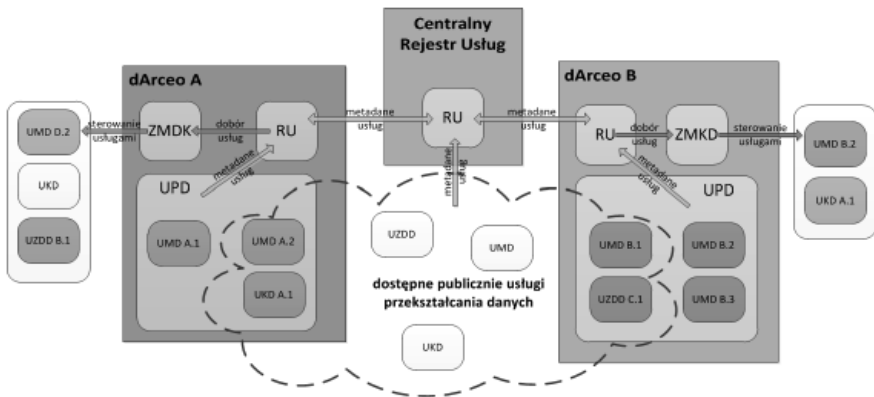
Zarządca Powiadomień

ZP odpowiada za dostarczenie funkcjonalności niezawodnego rejestrowania oraz przekazywania powiadomień pomiędzy modułami systemu dArceo oraz innymi systemami współpracującymi z dArceo.

Współdzielenie usług przekształcania danych

Usługa RU systemu dArceo może komunikować się z innymi usługami RU funkcjonującymi w niezależnych instancjach dArceo. Dzięki tej możliwości różne instancje systemu dArceo mogą wymieniać się między sobą informacjami na temat publicznie dostępnych usług przekształcania danych.

Zamieszczony poniżej diagram (rysunek 2) przedstawia mechanizm zarządzania i migracji danych w środowisku wielu instancji dArceo. Instancje te przedstawiono w uproszczeniu, koncentrując się na tych usługach, które współpracują ze sobą, a pomijając inne nieistotne z punktu widzenia omawianego aspektu systemu. Publiczne usługi przekształcania danych określono jako zbiór, do którego przynależność jest jednoznaczna z udostępnieniem publicznym usługi działającej w ramach danej instancji. Należy zaznaczyć, że w zbiorze tym znajdować się mogą zarówno usługi będące częścią różnych instancji dArceo, jak i niezależne usługi, upublicznione poza instancjami dArceo, ale zgodne ze specyfikacją usług przekształcania danych. Stąd część usług w zbiorze publicznych usług przekształcania danych pochodzi z poszczególnych instancji dArceo, a część jest dostępna niezależnie od systemów dArceo. Przykładowy scenariusz praktycznej realizacji takiej konfiguracji może obejmować usługi udostępniane przez dwa wdrożenia dArceo w instytucjach A i B oraz zbiór usług przekształcania oferowanych przez komercyjnych dostawców.



Rysunek 2. Zarządzanie konwersją i migracją danych w środowisku wielu instancji dArceo

Źródło: Poznańskie Centrum Superkomputerowo-Sieciowe.

Propagacja informacji o dostępnych usługach została przedstawiona w postaci zielonych strzałek, których grot wskazuje kierunek przepływu informacji o dostępności usługi. Usługi będące częścią instalacji dArceo udostępniane są Centralnemu Rejestrowi Usług przez rejestry lokalne. Usługi niebędące częścią instalacji dArceo muszą zostać zarejestrowane bezpośrednio w Centralnym Rejestrze Usług. Strzałki czerwone oznaczają wykorzystanie usługi poprzez odwołanie do RU. Możliwa jest również konfiguracja bez Centralnego Rejestru Usług, w której rejestry poszczególnych instancji dArceo synchronizują się pomiędzy sobą w modelu komunikacji „każdy z każdym” (ang. *peer-to-peer*, P2P).

Podsumowanie

System dArceo składa się z szeregu usług sieciowych, odpowiedzialnych za realizację funkcji długoterminowego przechowywania danych źródłowych, przede wszystkim w kontekście dokumentów tekstowych, graficznych oraz audiowizualnych. Podstawową ideą w ramach dArceo jest podejście transformacji typu migracja z modelu OAIS, uzupełnione dodatkowymi usługami konwersji danych i zaawansowanego dostarczania danych.

Obecnie dArceo jest testowane w ramach wdrożenia oprogramowania dLab⁹ w projekcie Repozytorium Cyfrowego Instytutów Naukowych (RCIN). RCIN to projekt zrzeszający 16 instytutów Polskiej Akademii Nauk w celu udostępnienia w Internecie interesujących zabytków piśmiennictwa. W ramach RCIN istnieje kilka niezależnych centrów digitalizacji, zewnętrzne narzędzia dedykowane do przygotowania wersji prezentacyjnych obiektów oraz szeregu raportów na temat postępu bieżących prac digitalizacyjnych. dLab jest narzędziem, które ma za zadanie ułatwić i usprawnić przebieg procesu digitalizacji poprzez automatyzację i monitorowanie poszczególnych jego etapów. Jednym z nich jest długoterminowe przechowywanie danych (archiwizacja). Na tym etapie procesu digitalizacji wykorzystywane jest oprogramowanie dArceo.

⁹ System do zarządzania procesem digitalizacji, rozwijany przez Poznańskie Centrum Superkomputerowo-Sieciowe; strona domowa PCSS [online], [dostęp: 10.10.2012], dostępny w World Wide Web: <http://dlab.pscn.pl>.

Bibliografia

1. Brzeźniak M. i in., *Popular Backup/Archival Service and its Application for the Archival of the Network Traffic in the PIONIER Academic Network*, „Computational Methods in Science and Technology” 2010, wydanie specjalne, s. 109–118.
2. Kuipers T., Hoeven van der J., *Insight into digital preservation of research output in Europe. Survey report* [online], [dostęp: 24.10.2012], dostępny w World Wide Web: http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf.
3. A. Lewandowska, C. Mazurek, M. Werla, Federacja Bibliotek Cyfrowych w sieci PIONIER. Dostęp do otwartych bibliotek cyfrowych i repozytoriów, [w:] *Open Access. Internet w bibliotekach*, (IV Ogólnopolska Konferencja EBIB, Toruń 7–8 grudnia 2007), [online], [dostęp: 15.10.2012], dostępny w World Wide Web: http://www.ebib.info/publikacje/matkonf/mat18/lewandowska_mazurek_werla.php
4. Parkoła T., Mazurek C., Werla M., Rozwój bibliotek cyfrowych w Polsce, [w:] *Partnerzy bibliotek. Model komunikacji z otoczeniem*, red. I. Jurczak, E. Okularczyk, Firma Księgarska Wiesława Juszczyka, Łódź 2009, s. 239–246.
5. Projekt PLATON [online], [dostęp: 24.10.2012], dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.
6. A. Swędryński, M. Górny i in., Zastosowanie oprogramowania dLibra do budowy Wielkopolskiej Biblioteki Cyfrowej, [w:] *Internet w bibliotekach II. Łączność, współpraca, digitalizacja*, (Międzynarodowa Konferencja EBIB, Wrocław 23–26 września 2003), [online], [dostęp: 15.10.2012], dostępny w World Wide Web: <http://lib.psnc.pl/Content/8/ebib2003.pdf>.

Justyna Maczuga, Jolanta Przyłuska

Institut Medycyny Pracy w Łodzi

Czy publikacje naukowe powinny mieć Digital Object Identifier?

Słowa kluczowe: Digital Object Identifier – DOI, identyfikator obiektów cyfrowych, identyfikator dokumentu elektronicznego, zasoby cyfrowe, publikacje elektroniczne, archiwizacja

Abstrakt: ilość dostępnych danych, tempo i dynamika wymiany informacji sprawiły, że Internet stał się podstawowym źródłem wiedzy. Zaistniała zatem silna potrzeba dokładnej identyfikacji jego zasobów, szczególnie tych naukowych. W artykule przedstawiono Digital Object Identifier (DOI) jako przykład identyfikatora obiektów cyfrowych, używanego do wyszukiwania dokumentów elektronicznych (zarówno artykułów, jak i całych numerów czasopism). Przeanalizowano proces nadania numeru DOI oraz korzyści z jego posiadania tak dla użytkowników, jak i wydawców. Wyszczególniono bazy danych oraz bazy artykułów, w których możliwe jest wyszukiwanie poprzez DOI (m.in.: Scopus, Web of Knowledge, OvidSP, Oxford Journals, ACS Publications). Wskazano, że długoterminowa archiwizacja publikacji elektronicznych wymaga nadania im stałego symbolu identyfikacyjnego, który umożliwi jego poprawne cytowanie i linkowanie.

Identyfikatory obiektów cyfrowych

Zwiększająca się liczba dokumentów elektronicznych w sieci wymusiła potrzebę ich dokładnego opisu i ustalenia sposobu jednoznacznej i szybkiej ich identyfikacji. Najczęściej stosowanymi identyfikatorami dokumentów drukowanych i dzieł audiowizualnych są m.in.:

- ISBN (*International Serial Book Number*), który określa kraj pochodzenia publikacji, wydawcę i wydanie,
- ISAN (*International Standard Audiovisual Number*) służący identyfikacji dzieł audiowizualnych,
- ISSN (*International Standard Serial Number*) to identyfikator wydawnictw ciągłych (zarówno tych tradycyjnych, jak i elektronicznych).

Stale identyfikatory obiektów cyfrowych pozwalają na zarządzanie prawem autorskim, dokładne cytowanie publikacji, sporządzanie elektronicznych spisów treści, tworzenie bibliografii, a przede wszystkim na szybkie przeszukiwanie zasobów Internetu (w tym repozytoriów, bibliotek cyfrowych i baz danych).

Jednym z najważniejszych identyfikatorów cyfrowych publikacji jest DOI (*Digital Object Identifier*). Identyfikator obiektu własności intelektualnej składa się z prefiksu (który określa wydawcę rejestrującego dokument) i sufixu (który bezpośrednio definiuje dokument) oddzielonych od siebie ukośnikiem (/), np. publikacja o numerze: 10.1366/11-06576, gdzie:

- 10 oznacza identyfikator DOI,
- 1366 to numer wydawcy, tzw. publisher ID (w tym przypadku Society for Applied Spectroscopy),
- 11-06576 identyfikuje obiekt cyfrowy, tzw. item ID (jest to artykuł *Pattern Recognition Assisted Infrared Library Searching*).

Na oficjalnej stronie internetowej systemu DOI (www.doi.org) znajduje się wyszukiwarka identyfikatorów, do której można wpisać dowolny numer poszukiwanej przez nas publikacji, np.:

Resolve a DOI Name

Type or paste a DOI name (e.g., 10.1000/182) into the text box below.

Rysunek 1. Wyszukiwarka DOI

Źródło: The DOI System [online], [dostęp: 23.08.2012], dostępny w World Wide Web: <http://www.doi.org>.

Po wybraniu „submit” ukaze się nam dokument, którego nazwę chcieliśmy ustalić. W tym przypadku nr 10.1007/s11192-012-0619-7 dotyczy publikacji *A comparative study of interdisciplinary changes between information science and library science*. Wyszukiwarka przeniesie nas do miejsca, z którego możemy pobrać dany dokument, jeśli tylko mamy uprawnienia do korzystania z niego:

Czy publikacje naukowe powinny mieć Digital Object Identifier?



Rysunek 2. Artykuł *A comparative study of interdisciplinary changes between information science and library science* w bazie SpringerLink

Źródło: SpringerLink [online], [dostęp: 7.08.2012], dostępny w World Wide Web: <http://www.springerlink.com/content/x12116113528lu17>.

Każdy numer DOI jest niepowtarzalny i przypisuje się go na stałe do jednostki cyfrowej, niezależnie od jej lokalizacji. Dokument zachowuje swój identyfikator DOI przez cały okres istnienia, a jeżeli zostanie kiedykolwiek usunięty, jego numer nie może być ponownie wykorzystany. DOI może być utworzony dla każdej jednostki cyfrowej, tj.: rozdziału z książki, artykułu, numeru czasopisma, wykresu, oprogramowania czy utworu muzycznego. Korzysta z opracowanego przez [Corporation](#) for National Research Initiatives standardu *Handle System*¹.

Rejestracja DOI

Kontrolę nad instytucjami i firmami, które uzyskały prawo do pełnienia roli agencji rejestrujących (*DOI Registration Agency*), sprawuje Międzynarodowa Fundacja DOI (*International DOI Foundation*). Zadaniem agencji rejestrujących jest przydzielanie numeru wydawcy (*Publisher ID*) i zapewnianie infrastruktury umożliwiającej im tworzenie swoich numerów obiektów (*Item ID*) oraz zarządzanie metadanymi przypisanymi identyfikatorom DOI. Międzynarodowa Fundacja DOI oczekuje też od agencji rejestrujących, że będą

¹ M. Nahotko, Identyfikacja obiektów w sieciach rozległych, [w:] *Internet w bibliotekach II. Łączność, współpraca, digitalizacja* [online], [dostęp: 6.08.2012], dostępny w World Wide Web: <http://ebib.oss.wroc.pl/matkonf/iwb2/nahotko.php>; A. Łozowska, Technologie informacyjne. Między DOI a Open Access, [w:] *Biblioteki naukowe w kulturze i cywilizacji. Działania i codzienność* [online], [dostęp: 16.08.2012], dostępny w World Wide Web: http://library.put.poznan.pl/2005/pdf/4_4.pdf.

one aktywnie promować ten system i stale współpracować nad jego rozwojem. Obecnie istnieje dziewięć (dane z sierpnia 2012 r.) agencji rejestrujących numery DOI na świecie. Są to:

- Airiti – dla dokumentów elektronicznych w języku chińskim,
- The Institute of Scientific and Technical Information of China (ISTIC) – agencja rejestrująca chińskie publikacje naukowe,
- Bowker – oferuje dodatkowo usługi marketingowe przeznaczone do promocji książek i czasopism,
- CrossRef – rejestruje tylko publikacje naukowe,
- DataCite – poza nadawaniem numerów DOI, koncentruje się na modelu biznesowym i poprawie infrastruktury naukowej wokół zbiorów z danymi,
- EIDR – rejestruje filmy i programy telewizyjne,
- Japan Link Center (JaLC) – promuje naukę i rozwój nowych technologii w Japonii, współpracuje z CrossRef w sprawie przydzielania DOI japońskim artykułom w języku angielskim,
- mEDRA – poza nadawaniem DOI, zajmuje się śledzeniem relacji między podmiotami własności intelektualnej,
- OPOCE (*Office des publications EU*) – przypisuje DOI publikacjom wszystkich europejskich podmiotów unijnych².

Koszt rejestracji obiektu w tych agencjach zależy od liczby identyfikatorów. Polscy wydawcy czasopism publikujący do 25 artykułów rocznie mogą skorzystać z agencji mEDRA (roczna opłata wynosi 150 €). Dla większych wydawców bardziej korzystna jest propozycja amerykańskiej agencji Bowker. Wydawnictwo, które publikuje do 999 artykułów/rok, zapłaci 500 \$ rocznie³.

Korzyści wynikające z zastosowania DOI

Wydawcy, użytkownicy, jak i sami bibliotekarze uzyskują korzyści z nadawania coraz większej liczbie publikacji elektronicznych numerów DOI. Wydawcom zapewnia to możliwość zarządzania prawami wydawniczymi, a administratorom stron www niezmienny identyfikator, który działa jak stały URL. Zmiana adresu serwera, na którym umieszczono dokument, uaktualnia wszystkie przypisy i lokalizacje stron. W przypadku bibliotekarzy ważne

² Factsheets. DOI System and Internet Identifier Specifications [online], [dostęp: 23.08.2012], dostępny w World Wide Web: <http://www.doi.org/factsheets/DOIIdentifierSpecs.html>.

³ E. Kulczycki, Identyfikator DOI dla czasopisma naukowego [online], [dostęp: 24.08.2012], dostępny w World Wide Web: http://ekulczycki.pl/warsztat_badacza/identyfikator-doi-dla-czasopisma-naukowego.

jest, by polecali DOI jako standardową część opisu bibliograficznego, gdyż zapewnia to niezmiennosc lokalizacji dokumentu (brak konieczności aktualizacji w opisie bibliograficznym adresów internetowych), usprawnia elektroniczną ich dostawę, zakup pojedynczych artykułów, identyfikację, wypożyczanie międzybiblioteczne. Użytkownicy Internetu, bibliotek i baz danych zyskują wygodę w poszukiwaniu i dostępie do interesujących ich publikacji⁴.

Do maja 2012 roku DOI otrzymało ponad 55 mln obiektów cyfrowych. Używany jest obecnie przez ponad 5 tys. wydawców, firm nagraniowych, towarzystw i instytucji, m.in.: Elsevier, John Wiley & Sons, Springer, Wolters Kluwer, China National Knowledge Infrastructure (CNKI), International Health & Science, International Federation of Reproduction Rights Organisations (IFRRO), STM International Association, Copyright Agency Limited, Publishers Licensing Society⁵.

Coraz więcej baz danych i baz artykułów naukowych wprowadza do wyszukiwarki umieszczonej na swojej stronie pozycję umożliwiającą wyszukiwanie poprzez numer DOI. Należą do nich: ACS Publications, OvidSP, Oxford Journals, Scopus, SpringerLink, Web of Knowledge, Wiley Online Library.

Długoterminowa archiwizacja

Pojęcie *długoterminowej archiwizacji zasobów cyfrowych* oznacza przechowywanie publikacji elektronicznych i zapewnienie ich długoterminowej użyteczności (dostępności, możliwości odczytania, prezentacji w formie zrozumiałej dla odbiorcy) zarówno obecnie, jak i w odległej przyszłości. Ważne jest również, aby zapewnić dostęp do nieprzetworzonej wersji dokumentu i umożliwić jego wykorzystanie zgodnie z pierwotnym celem. Ochrona użyteczności dokumentu cyfrowego jest natomiast zapewniona wtedy, gdy odbiorca ma pewność, że dokument, z którego korzysta, jest zgodny z wersją oryginalną i od dnia opublikowania przedstawia dokładnie to, co było zamierzeniem twórcy. Do najważniejszych cech długoterminowej archiwizacji zbiorów należy więc: ochrona dostępności, użyteczność, autentyczność (wiarygodność), integralność dokumentów, poufność (jeśli istnieje taka potrzeba, to treść dokumentu nie może być ujawniana osobom nieupoważnionym)⁶.

Długoterminowa archiwizacja zasobów cyfrowych wymaga zatem stabilności, autentyczności i ochrony, czyli cech, które zapewnia DOI. Stały iden-

⁴ A. Łozowska, *op. cit.*

⁵ Factsheets. DOI System..., *op. cit.*

⁶ A. Januszko-Szakiel, *Nowy wymiar zabezpieczania elektronicznych zasobów* [online], [dostęp: 5.09.2012], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?januszko>.

tyfikator przypisywany wszystkim dokumentom elektronicznym i obiektom cyfrowym jednoznacznie potwierdza ich autentyczność i jest przydatnym elementem, jeśli chodzi o zbiory archiwalne, muzealne i biblioteczne.

Podsumowanie

Współczesna archiwizacja, szczególnie ta dotycząca zasobów dziedzictwa cyfrowego (a więc tych podstawowych, niezwykle ważnych dla historii i kultury danego kraju), potrzebuje stałych odnośników i identyfikatorów, które je zdefiniują oraz umożliwią w przyszłości poprawne cytowanie oraz sprawne wyszukiwanie w bazach i repozytoriach dokumentów elektronicznych i obiektów cyfrowych.

Bibliografia

1. Factsheets. DOI System and Internet Identifier Specifications [online], [dostęp: 23.08.2012], dostępny w World Wide Web: <http://www.doi.org/factsheets/DOIIdentifierSpecs.html>.
2. Januszko-Szakiel A., *Nowy wymiar zabezpieczania elektronicznych zasobów* [online], [dostęp: 5.09.2012], dostępny w World Wide Web: <http://www.ebib.info/2010/115/a.php?januszko>.
3. Kulczycki E., *Identyfikator DOI dla czasopisma naukowego* [online], [dostęp: 24.08.2012], dostępny w World Wide Web: http://ekulczycki.pl/warsztat_badacza/identyfikator-doi-dla-czasopisma-naukowego.
4. Łozowska A., Technologie informacyjne. Między DOI a Open Access, [w:] *Biblioteki naukowe w kulturze i cywilizacji. Działania i codzienność* [online], [dostęp: 16.08.2012], dostępny w World Wide Web: http://library.put.poznan.pl/2005/pdf/4_4.pdf.
5. Nahotko M., Identyfikacja obiektów w sieciach rozległych, [w:] *Internet w bibliotekach II. Łączność, współpraca, digitalizacja* [online], [dostęp: 6.08.2012], dostępny w World Wide Web: <http://ebib.oss.wroc.pl/mat-konf/iwb2/nahotko.php>.

Zdzisław Pietrzyk

Biblioteka Jagiellońska

Archiwizacja zasobów cyfrowych Jagiellońskiej Biblioteki Cyfrowej

Słowa kluczowe: Biblioteka Jagiellońska, Jagiellońska Biblioteka Cyfrowa, JBC, trwała archiwizacja zasobów cyfrowych, konserwacja zbiorów

Abstrakt: ze względu na rodzaj przechowywanych zbiorów, współczesne biblioteki są w większości bibliotekami hybrydowymi. Zasady ochrony, konserwacji i zabezpieczania tradycyjnych zbiorów były wypracowywane przez wieki. Natomiast biblioteki cyfrowe to zjawisko znane na świecie od ćwierćwiecza. Standardy tworzenia, gromadzenia, opracowania i udostępniania zbiorów elektronicznych powstają w codziennej praktyce bibliotekarskiej. W Bibliotece Jagiellońskiej utworzono Jagiellońską Bibliotekę Cyfrową [JBC]. Stworzenie biblioteki cyfrowej nastąpiło z pewnym opóźnieniem w porównaniu do innych dużych bibliotek. Powodem był brak odpowiedniego sprzętu, zapewniającego wysoką jakość skanowania, a co za tym idzie, niemożliwe było udostępnianie czytelnikom zbiorów w wysokiej jakości. Na etapie planowania JBC istotnym zagadnieniem stało się zabezpieczanie i przechowywanie zbiorów elektronicznych. W zależności od wartości i znaczenia dla dziedzictwa narodowego, zbiory wytworzone w procesie skanowania, jak i te *born digital*, są przechowywane w trzech niezależnych systemach i środowiskach. Tak zaprojektowany system umożliwi wieloletnie przechowywanie zbiorów, które są udostępniane czytelnikom w postaci elektronicznej w JBC.

Wstęp

Utworzenie Jagiellońskiej Biblioteki Cyfrowej było celem, który zaczęliśmy sobie stawiać wiele lat temu. Cyfryzacja Biblioteki Jagiellońskiej zaczęła się w 1994 r. Od tego momentu wszystkie procesy biblioteczne, począwszy od gromadzenia zbiorów, po wypożyczenie książek, są obsługiwane przez zintegrowany system biblioteczny. Komputery od 1994 r. stały się podstawowym narzędziem pracy w Bibliotece Jagiellońskiej. Był to wielki skok technologiczny. Pracownicy posługujący się dotąd ołówkami, długopisami i maszynami do pisania poznali skomplikowany system kompleksowej obsługi procesów bibliotecznych. Stanowiło to wielkie wyzwanie, ponieważ był to dopiero początek cyfryzacji kraju. Decyzja dyrektora Biblioteki Jagiellońskiej prof. Krzysztofa Zamorskiego o katalogowaniu wyłącznie w systemie komputerowym VTLS (*Virginia Tech Library System*) była bardzo odważna i krytyko-

wana przez osoby sceptycznie podchodzące do powodzenia cyfryzacji. Rezygnacja z katalogowania na kartkach katalogowych, które były podstawą systemu informacyjnego w bibliotekach, było jednoznaczne ze zmianą podejścia do przechowywania danych. Kartki katalogowe były przechowywane w szafach i posługiwanie się nimi było nieodłącznym elementem korzystania ze zbiorów. Karty katalogowe służyły zatem do gromadzenia danych, które dziś nazywamy metadanymi. Informacje na nich zawarte stanowiły podstawę do poszukiwania książek i bez nich, jako nośnika informacji, trudno było wyobrazić sobie działanie biblioteki. Karty katalogowe były w miarę trwałe, a przechowywano je w szafach katalogowych. Nikt wówczas nie myślał o ich ochronie i specjalnym zabezpieczeniu. Zdarzały się kradzieże kart katalogowych, by nikt nie mógł zamówić już raz wypożyczonej książki. Zabezpieczeniem podstawowym informacji zawartych na kartach katalogowych był inwentarz. Obecnie zbiory katalogowane są wyłącznie w wersji elektronicznej, równocześnie sporządza się także inwentarz, który jest generowany przez system i drukowany na papierze.

Zabezpieczanie danych katalogowych

System VTLS, a następnie jego nowsza wersja VIRTUA, jest posadowiony na serwerze i codziennie robiona jest jego kopia. Wytwarzane każdego dnia dane są przechowywane na serwerach. Wraz z rozwojem softwaru nośniki informacji i odtwarzacze zmieniały się; pierwszy serwer miał aż 4 gigabity, a jedną z wielu trosk ówczesnych pracowników BJ było zabezpieczenie informacji, które zawierały rekordy bibliograficzne, hasła wzorcowe i przedmiotowe związane z katalogowaniem. Drugą formą zabezpieczenia katalogu komputerowego jest nagrywanie danych katalogowych na taśmę i przechowywanie ich w bibliotece taśmowej. W pierwszych kilkunastu latach funkcjonowania katalogu i elektronicznej pracy bibliotecznych mieliśmy bardzo skromne możliwości zabezpieczania danych na taśmach. Sprzęt, którym dysponowaliśmy, pozwalał wyłącznie na zabezpieczenie rekordów bibliograficznych w bibliotece taśmowej. Ostateczne zabezpieczenie podstawowych informacji jest sporządzane jako wydruk na papierze. Stanowi to jednocześnie spełnienie wymogu formalnego związanego z zabezpieczeniem majątku biblioteki.

Jagiellońska Biblioteka Cyfrowa

Następnym krokiem w procesie elektronicznej bibliotek było przechodzenie na etap tworzenia bibliotek cyfrowych. Pojawienie się Internetu i rozproszonych w nim informacji było kolejnym przełomem w dziejach komunikacji społecznej. Dane implementowane do sieci są rozproszone, chaotyczne i niestety w wielu miejscach niezetelne. Informacje te umieszczane są w chmurze informacji, w niebywałym chaosie informacyjnym – mimo istnienia wyszukiwarek internetowych – trudne do odnalezienia. Jedynym uporządkowanym zakątkiem Internetu są biblioteki cyfrowe. Są one bowiem zbudowane na identycznej zasadzie, jak biblioteki tradycyjne, w których jednym z najważniejszych wyznaczników prawnego działania jest ustrukturyzowanie wszelkich procesów. Dbano o ten czynnik sprawia, że biblioteki cyfrowe są uporządkowanym i wiarygodnym źródłem informacji.

Osobom niezającym zasad powstawania biblioteki cyfrowej wydaje się, że budowana jest ona w zasadzie bez większych nakładów finansowych. Wszak w zdecydowanej większości przypadków korzystanie z bibliotek zarówno tradycyjnych, jak i cyfrowych jest bezpłatne. Z praktyki budowy biblioteki cyfrowej wiemy, że przy tego typu przedsięwzięciach nakłady są, i to znaczne. Profesjonalnie zaplanowana i budowana biblioteka cyfrowa musi opierać się na odpowiednim sprzęcie oraz specjalistach, którzy sprzęt i systemy są w stanie obsłużyć.

Przy pierwszych próbach tworzenia wersji elektronicznych najcenniejszych obiektów w Polsce zbiory Biblioteki Jagiellońskiej były brane pod uwagę jako godne upowszechnienia. W 1994 r. powstał pierwszy projekt zdigitalizowania zbiorów BJ. Zdecydowano wówczas, że procesowi temu zostanie poddany najcenniejszy obiekt przechowywany w Bibliotece Jagiellońskiej – autograf rękopisu Mikołaja Kopernika *De Revolutionibus*. Proces technologiczny polegał na sporządzeniu diapozytywów, a następnie wykonano z nich kopię cyfrową. Proces sporządzania kopii autografu Kopernika przypominał bardziej wykonanie faksymile niż obiektu do biblioteki cyfrowej. Po sporządzeniu kopii cyfrowej nagrano ją na płytę CD, w ten sposób zaczęto udostępniać na tym nośniku wersję cyfrową rękopisu wielkiego astronoma. Przez wiele lat zdigitalizowany rękopis Kopernika był przechowywany na płycie CD, a na skopiowanej płycie był udostępniany użytkownikom. Do momentu powstania JBC kopia cyfrowa tego najcenniejszego polskiego rękopisu była przechowywana wyłącznie na płycie CD. Rękopis ten ma oczywiście, poza kopią cyfrową, inne kopie, w tym faksymile papierowe wydane w 1973 r. Obecnie udostępnianie oryginału działa Kopernika sprowadza się w zasadzie do rzadkiego prezentowania go na wystawach.

Brak odpowiedniego sprzętu uniemożliwił budowę profesjonalnej biblioteki cyfrowej, a byłem przeciwny budowaniu biblioteki cyfrowej bez tworzenia plików archiwalnych o dużej rozdzielczości. Wykonywanie skanów na sprzęcie, który zagrażałby zbiorom, uważałem za niedopuszczalne – sporządzanie skanów na skanerach płaskich, jakimi wówczas dysponowaliśmy, zagrażało bowiem zbiorom, szczególnie zbiorom specjalnym. Należy pamiętać, że jesteśmy opiekunami zbiorów, które mają przetrwać dla następnych pokoleń.

Pierwszym większym programem, stanowiącym załączek budowy biblioteki cyfrowej, był projekt „Cyfrowe skryptorium Biblioteki Jagiellońskiej”. Celem tego projektu miało być przygotowanie warsztatu pracy dla mediawistów w szerokim zakresie badań nad kulturą średniowiecza. Chcieliśmy zdigitalizować wszystkie rękopisy średniowieczne i inkunabuły z naszych zbiorów – chodziło bowiem o ochronę najcenniejszego zasobu – ochronę fizycznego zachowania oryginałów. Wystąpiliśmy do ówczesnego KBN o dotację. Projekt ten oceniono jako zbyt kosztowny i niemający charakteru badawczego.

Dzięki zapałowi pracowników BJ, udało się nam zeskanować 14 rękopisów BJ i 2 rękopisy z biblioteki klasztoru oo. Paulinów na Skałce wraz z ich opisami katalogowymi. Jednym z problemów było miejsce na serwerze, by posadzić tam rękopisy i oczywiście kłopot z przechowywaniem ich wersji cyfrowej. Płyty CD nie spełniały i nie spełniają tego zadania. Ta minibiblioteczka cyfrowa była namiastką tego, co chcieliśmy osiągnąć i była jednocześnie sprawdzianem umiejętności oraz możliwości pracowników BJ.

Początkowo stawialiśmy w Bibliotece Jagiellońskiej na utworzenie centralnej narodowej biblioteki cyfrowej i sieci bibliotek cyfrowych digitalizujących zbiory specjalne oraz regionalia. Taką nieudaną próbą stworzenia centralnej państwowej biblioteki cyfrowej była Polska Biblioteka Internetowa. Mimo zaimplementowania ponad 32 tys. pozycji, Polska Biblioteka Internetowa upadła. Zbiory PBI są nadal dostępne w sieci. Nie pora tu na analizę niepowodzenia tego przedsięwzięcia; było ono bowiem niezwykle złożone, poczynając od spraw finansowych, ambicjonalnych, a na politycznych kończąc. Pierwsza faza powstawania tej biblioteki to pasmo wielu fatalnych błędów, ponieważ bibliotekę cyfrową zaczęli tworzyć informatycy i urzędnicy – wówczas to dowiedzieliśmy się z metadanych, że autorem Biblii jest Gutenberg. Mimo tych błędów, PBI miała pełną akceptację i poparcie władz państwowych. Po odejściu ze stanowiska wiceministra ds. informatyzacji kraju Wojciecha Szewki, orędownika PBI, dalsze działanie tej biblioteki cyfrowej odbywało się już tylko siłą rozpędu, aż do całkowitego jej zatrzymania. W Bibliotece Jagiellońskiej przy okazji budowy PBI został wypracowany kanon literatury polskiej, który powinien zostać zdigitalizowany. Z ponad 16 tys.

pozycji, jakie wskazaliśmy jako kanon literatury, w BJ zostało zeskanowanych ponad 4 tys. starych druków i pozycji pochodzących z XIX w.

Biblioteka Jagiellońska, wobec zawirowań wokół PBI, rozpoczęła starania o środki na rozwój profesjonalnej biblioteki cyfrowej. Kolejne wystąpienia o środki europejskie w latach 2005–2007 kończyły się niepowodzeniem. Przez kilka lat odstawaliśmy od dużych bibliotek w zakresie udostępniania zbiorów cyfrowych, poza kupowanymi bazami danych i czasopismami elektronicznymi. Powodem było trzymanie się zasady kompleksowego rozwiązania kwestii związanych z kopiowaniem cyfrowym, udostępnianiem cyfrowym oraz archiwizacją zasobów cyfrowych. W międzyczasie tworzyliśmy podstawy pod przyszłą bibliotekę cyfrową Uniwersytetu Jagiellońskiego. Były to: wspomniane skrytorium zawierające kilkanaście rękopisów średniowiecznych, Wirtualny Gabinet Rycin, Cyfrowa Baza Ikonografii Teatralnej, skany książek itd.

Kolejny wniosek o środki europejskie na budowę biblioteki cyfrowej został zaakceptowany w 2009 r. jako projekt kluczowy Ministerstwa Kultury i Dziedzictwa Narodowego. Projekt JBC został skonstruowany w oparciu o inne kryteria niż nasze wcześniejsze założenia odnośnie do budowy biblioteki cyfrowej. W zasadzie odstąpiliśmy od masowej digitalizacji zasobów z tzw. zbiorów specjalnych, w większości spisanych i wydrukowanych na dobrym zasadowym papierze. W projekcie JBC priorytetem jest zabezpieczenie treści druków wydanych w pierwszej połowie XIX w., a przede wszystkim gazet wydrukowanych na kwaśnym, bardzo złej jakości papierze, który ulega samodegradacji. W projekcie JBC zamierzamy i w większej części już zdigitalizowaliśmy 225 tytułów czasopism, w tym „Gazetę Lwowską”, „Nową Reformę”, „Gazetę Narodową”, „Dziennik Polski”, które w całości występują tylko w naszej bibliotece, blisko tysiąc pozycji z Oddziału Rękopisów i Oddziału Starych Druków, ok. tysiąca jednostek ze zbiorów muzycznych i kartograficznych oraz kilkaset tytułów książek. Mimo problemów formalnych, realizacja projektu postępuje zgodnie z planem, a do zbiorów JBC dodawane są pozycje, na które uzyskaliśmy licencje od autorów, m.in. 22 maja br. zorganizowaliśmy spotkanie, podczas którego kilkunastu autorów przekazało do JBC licencje na udostępnianie wybranych swoich książek w JBC.

W JBC rozpoczęliśmy realizację odpowiedniego zabezpieczenia naszych zbiorów w różnych formatach i na różnych nośnikach. Nośnikach, które, z jednej strony, szybko się zmieniają i wymagają odpowiednich odtwarzaczy, a z drugiej strony, tradycyjnych, gdzie odtworzenie obrazu nie wymaga sprzętu, którego parametry techniczne wciąż się zmieniają – mam tu na myśli mikrofilm. Wszyscy przekonaliśmy się, że takie nośniki, jak taśmy magnetofonowe magnetyczne w swojej popularnej postaci są bardzo zawodne i ich odtwarzacze są już praktycznie niedostępne, podobnie dzieje się

z różnymi formatami dyskietek i płyt, na których dane zapisywano cyfrowo. Formaty dyskietek z komputerów Amstrad, dyskietki 5¼ cala, dyskietki 3½ cala są trudne do odczytania, a także wymagają specjalistycznego sprzętu i konwersji oprogramowania.

Jedną z zasadniczych kwestii, podobnie jak przy cyfrowym katalogu zbiorów, było zabezpieczenie sporządzonych kopii cyfrowych i ich archiwizacja. Pierwszy i podstawowy sposób to zgrywanie plików cyfrowych na macierze dyskowe na serwerach, drugim rozwiązaniem jest rozbudowana biblioteka taśmowa, gdzie na taśmę magnetyczną nagrywane są pliki cyfrowe, oraz trzecia metoda to wykonywanie kopii mikrofilmowych z plików cyfrowych. Co do ostatniego punktu – do niedawna, by przyspieszyć digitalizację, z mikrofilmów uzyskiwano pliki cyfrowe, były one jednak w zdecydowanej większości bardzo złej jakości. Obecnie konwersja pliku cyfrowego na taśmę celuloidową daje mikrofilm o bardzo dobrej jakości.

Archiwizacja, jak można określić – klasyczna, na serwerze jest stosunkowo prosta, bardziej skomplikowany jest proces archiwizacji mikrofilmowej oraz taśmowej (patrz slajdy).

Archiwizacja mikrofilmowa

Archiwizacja mikrofilmowa w JBC

- W projekcie Jagiellońskiej Biblioteki Cyfrowej przewidziano wykonanie dodatkowej archiwalnej kopii mikrofilmowej
- W tym celu zakupiono konwerter wykonujący mikrofilm z kopii cyfrowej



Przygotowanie archiwizacji mikrofilmowej w JBC

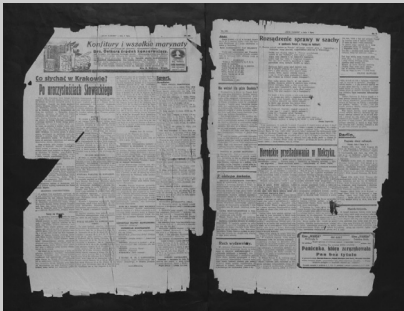
- Archiwizacja mikrofilmowa w JBC odbywa się równolegle z procesem digitalizacji
- W związku z tym bibliotekarze cyfrowi wspomagają proces mikrofilmowania zbiorów
- Kopia mikrofilmowa jest wykonywana z przygotowanej przez bibliotekarzy kopii cyfrowej

Zalety mikrofilmowania z kopii cyfrowej

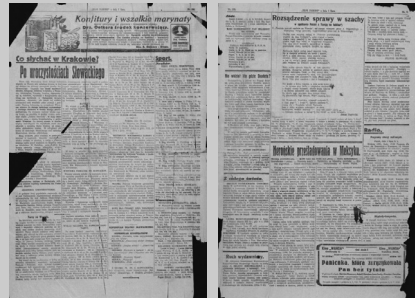
- Skład czasopism lub książek często zawiera błędy w postaci poprzestawianych stron
- Czasem w woluminie czasopisma brakuje numeru bądź strony, a są one zawarte w innym woluminie
- Bibliotekarze cyfrowi poprawiają tego typu przestawienia wykonując kopię cyfrową
- Dzięki temu te poprawki można wykorzystać przy wykonywaniu kopii mikrofilmowej
- Dodatkowo każda klatka poddana została korekcie graficznej, takiej jak kopie cyfrowe zasobów

Cyfrowa kopia archiwalna i kopia mikrofilmowa

Cyfrowa kopia archiwalna



Pliki mikrofilmowane



Zalety korekty graficznej przy wykonywaniu kopii mikrofilmowej



- Odpowiednie wykadrowanie klatki
- Wyrównanie kolumny (strony w dokumentach często są przekrzywione)
- Wyostrenie obrazu

Dwie kopie archiwalne



- Poza kopią archiwalną przechowywaną na taśmach streamingowych w formacie TIFF, pliki przeznaczone do konwersji na mikrofilm również zostają zachowywane
- Archiwalna kopia cyfrowa jest obrazem zeskanowanym
- Kopia mikrofilmowa jest obrazem dokumentu po wykonaniu korekty graficznej

Format archiwizowania



- Jako format archiwizowania drugiej kopii cyfrowej wybrano JPEG2000
- Ten format w archiwizowaniu używa się w Bayerische Staatsbibliothek, ostatnio promuje go również Biblioteka Kongresu
- Jego zaletą jest najlepsza kompresja bezstratna
- Wadą – brak metadanych pliku (np. rozdzielczość)
- Wszystkie pliki mają swoje źródła w formacie TIFF, gdzie są zapisane metadane

Nagranie taśm

- Przygotowany katalog eksportowy zostaje nagrany na taśmy w dwóch egzemplarzach
- Obsługa dużej liczby taśm jest ułatwiona przez korzystanie z robota taśmowego

Nagranie taśm



Przechowanie taśm

- ◆ Egzemplarz podręczny taśmy jest przechowywany w Oddziale Komputeryzacji
- ◆ Egzemplarz archiwalny – w odrębnej bezpiecznej lokacji

Przechowanie taśm



Dorota Olejnik, Elżbieta Gongąła,
Zofia Kukurowska

Biblioteka Główna Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy

Repozytorium Cyfrowe Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy jako przykład opracowania, udostępniania i archiwizacji dorobku naukowego wykładowców uczelni

Słowa kluczowe: Repozytorium Cyfrowe Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy, RC UTP, długoterminowa archiwizacja danych cyfrowych, PLATON

Abstrakt: w artykule opisano proces organizowania Repozytorium Cyfrowego Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy (RC UTP) jako przykład wykorzystania środków unijnych do budowy zasobu cyfrowego biblioteki. Omówiono programy unijne realizowane przez Dział Współpracy z Gospodarką, dzięki którym wyposażono Pracownię Digitalizacji i dofinansowano utworzenie repozytorium. Przedstawiono zespół zaangażowany w digitalizację oraz przebieg prac związanych z funkcjonowaniem RC UTP. Uzasadniono kryteria tworzenia kolekcji oraz zaprezentowano potencjalnego odbiorcę treści zasobów repozytoryjnych. Omówiono problematykę stosowanych licencji, przeanalizowano statystykę wykorzystania poszczególnych publikacji i repozytorium w ogóle. Opisano dotychczasowe doświadczenia związane z uruchomieniem modułu autoarchiwizacji. Odniesiono się do zagadnienia archiwizacji plików powstałych w toku digitalizacji oraz do planów długoterminowej archiwizacji danych na platformie PLATON.

Zgodnie z założeniami projektu, Repozytorium Cyfrowe UTP ma być adresowane do przedsiębiorców z regionu, jednak po ukończeniu projektu unijnego zakłada się jego większe wykorzystanie dla potrzeb środowiska uczelnianego.

Wstęp

Nowe możliwości w jakości komunikacji, jakie daje Internet zmieniły formy publikowania naukowego. Internet stał się odpowiedzią na potrzebę szybkiego publikowania. Specyfika publikowania tradycyjnego, np. czasopism naukowych, sprawiała że określone dane w chwili ukazania się drukiem były już często nieaktualne. Jednocześnie przy rosnących cenach takich publikacji, forma ta przestała odpowiadać

potrzebom środowiska naukowego. Zrozumiano, że aby wyniki badań naukowych przynosiły oczekiwane korzyści, muszą one być dostępne dla szerszego grona odbiorców. W wyniku potrzeby powszechnego dostępu zintensyfikowano działania na rzecz wykorzystania sieci dla potrzeb świata nauki. Ich wynikiem stały się czasopisma elektroniczne, następnie biblioteki cyfrowe i repozytoria. Właśnie te ostatnie, obok szybkości publikowania wyników badań, zapewniają także długoterminowe przechowywanie zgromadzonego materiału. Repozytoria spełniają też rolę systematyzacji i porządkowania wiedzy, co w dobie rozwoju Internetu i zalewu informacji stanowi wyzwanie, a zarazem konieczność¹.

Dobre funkcjonowanie repozytorium może zapewnić współpraca między naukowcami, bibliotekarzami, informatykami oraz pracownikami administracyjnymi uczelni². Świadomość, jakie korzyści daje repozytorium, powinna motywować każdą z tych grup do pracy na jego rzecz. Repozytorium uczelniane jest również szansą dla wzrostu prestiżu biblioteki i uczelni. To też sposób na odnalezienie się bibliotek akademickich w nowej rzeczywistości. Przygotowanie prezentacji na temat repozytorium, organizowanie szkoleń w celu rozpropagowania tejże idei, zazwyczaj należy do zadań biblioteki. Obok zaangażowania bibliotekarzy ideałem jest pełne poparcie dla repozytorium ze strony władz uczelni. Decyzje w dziedzinie finansów przeznaczanych na ten cel oraz konkretne zarządzenia bardzo ułatwiają pracę.

Repozytorium uczelniane to cenne narzędzie nie tylko dla kadry naukowej, ale także dla środowiska studenckiego. Umieszczanie w nich obok prac doktorskich, publikacji pokonferencyjnych, skryptów czy materiałów edukacyjnych sprawia, że odbiorcami są również studenci danej uczelni. Wykłady i prezentacje umieszczane często na prywatnych stronach wykładowców mogą być z powodzeniem udostępniane za pomocą repozytorium.

Środki unijne

Bezdyskusyjne korzyści z posiadania repozytorium uczelnianego wywołują często entuzjazm w organach zarządczych. Mija on, gdy przechodzi się do strony finansowej przedsięwzięcia. Jego koszty uzależnione są od przyjętej strategii budowy repozytorium.

¹ M. Rychlik, E. Karwasińska, Repozytorium instytucjonalne jako czynnik wspomagający rozwój nauki w środowisku akademickim [online], [dostęp: 29.08.2012], dostępny w World Wide Web: https://repozytorium.amu.edu.pl/jspui/bitstream/10593/1074/1/Biblioteka_07.pdf.

² *Ibid.*, s. 6.

W przypadku Repozytorium Cyfrowego Uniwersytetu Technologiczno-Przyrodniczego (RC UTP) droga ta przebiegała w kierunku od środków finansowych do planowania repozytorium. Pierwszy projekt, który w znaczącym stopniu wspomógł powstanie i realizację repozytorium nosił nazwę „Realizacja II etapu Regionalnego Centrum Innowacyjności”.³ Był to jeden z wielu projektów strukturalnych wdrażanych w latach 2007–2013 na Uniwersytecie Technologiczno-Przyrodniczym w Bydgoszczy. UTP jako jedyna w województwie kujawsko-pomorskim uczelnia publiczna o profilu technicznym, odgrywa ważną rolę w rozwoju technologicznym regionu. Podstawą jest efektywna współpraca uczelni z przedsiębiorstwami. W ramach „Realizacji II etapu Regionalnego Centrum Innowacyjności” stworzono sieć współpracujących ze sobą specjalistycznych laboratoriów badawczych. Programem wsparcia zostały objęte wszystkie jednostki uczelniane. Mają one wspomagać rozwój regionu, zwłaszcza małych i średnich przedsiębiorstw. W roku 2010 powstało Centrum Digitalizacji i Przetwarzania Informacji Naukowo-Technicznej oraz działająca przy Dziale Informacji Naukowej Biblioteki Głównej UTP, Pracownia Digitalizacji. Miało to umożliwić dostęp do zbiorów bibliotecznych czytelnikom spoza uczelni, ze szczególnym uwzględnieniem mieszkańców regionu kujawsko-pomorskiego oraz wspierać proces dydaktyczny uczelni i kształcenie na odległość. Pracownia wyposażona została w skaner płaski A2 WideTek 25, skaner Plustek A3 Optibook A300, aparat fotograficzny Canon 550D wraz z potrzebnym oprzyrządowaniem, a także oprogramowanie służące obróbce skanów (Document Express oraz Photoshop). Dzięki temu możliwe było rozpoczęcie w roku 2010 współpracy naszej Biblioteki z Kujawsko-Pomorską Biblioteką Cyfrową.

Biblioteka Główna UTP uzyskała środki finansowe na utworzenie Repozytorium Cyfrowego UTP w ramach projektu „Naukowcy UTP dla gospodarki regionu”. Przygotował go i realizuje Dział Współpracy z Gospodarką. Fundusze pochodzą z Programu Operacyjnego Kapitał Ludzki, jednego z programów operacyjnych wdrażanych w latach 2007–2013. Jest on częścią Europejskiego Funduszu Społecznego wspierającego i finansującego politykę społeczną państw członkowskich na płaszczyźnie przeciwdziałania bezrobociu i rozwoju zasobów ludzkich⁴. Sam Program Kapitał Ludzki składa się z 10 Priorytetów. Projekt „Naukowcy UTP dla gospodarki regionu” stanowi Priorytet VII: Regionalne kadry gospodarki, Działanie 8.2: Transfer wiedzy, Poddziałanie 8.2.1: Wsparcie dla współpracy sfery nauki i przedsiębiorstw.

³ Realizacja II etapu Regionalnego Centrum Innowacyjności. Mój region w Europie [online]. [dostęp 29.08.2012], dostępny w World Wide Web: <http://www.rci-2etap.utp.edu.pl>.

⁴ Ministerstwo Pracy i Polityki Społecznej – Departament Wdrażania Europejskiego Funduszu Społecznego. O programie [online], [dostęp: 4.07.2012], dostępny w World Wide Web: <http://www.kapitalludzki.gov.pl/o-programie>.

Głównym celem ww. projektu „jest stworzenie warunków do zwiększenia i intensyfikacji współpracy między środowiskiem naukowym UTP a regionalnymi przedsiębiorcami”⁵. Cel ten ma zostać osiągnięty m.in. przez udostępnianie przedsiębiorcom zbiorów bibliotecznych, tzn. digitalizację i upowszechnianie w sieci za pomocą repozytorium, najnowszych publikacji pracowników naukowych uczelni. Projekt „Naukowcy UTP dla gospodarki regionu” przewiduje realizację zadań w latach 2011–2013.

Przygotowania do uruchomienia Repozytorium

Biblioteka Główna UTP chętnie podjęła się uczestnictwa w projekcie unijnym „Naukowcy UTP dla gospodarki regionu”, dostrzegając w nim szansę na utworzenie repozytorium uczelnianego. Zakup potrzebnego sprzętu, wybór odpowiedniego oprogramowania, jak i budowę zespołu osób pracujących nad repozytorium, powierzono Bibliotece Głównej. Doposażono Pracownię Digitalizacji, zakupując wysokiej klasy skaner dzielowy Bookeye 4. Dzięki niemu proces digitalizacji nabrał przyspieszenia. Poprawie uległa też jakość skanów.

Budowa repozytorium na oprogramowaniu bezpłatnym, tzw. open source, pozwala obniżyć koszty. Jednak trzeba pamiętać, że wymaga ono profesjonalnego zaplecza informatycznego. Implementacja i serwisowanie takiego oprogramowania spoczywa całkowicie na administratorze repozytorium⁶. Niestety BG UTP nie dysponuje odpowiednim wsparciem informatycznym, dlatego zdecydowano się na wybór oprogramowania komercyjnego. Wybór padł na poznaną wcześniej, dzięki pracy na rzecz KPBC, dLibrę. Na początku roku 2011 zorganizowano szkolenie dla zespołu tworzącego RC UTP. W ramach szkolenia zapoznano się szczegółowo z działaniem oprogramowania. Prześledzono drogę deponowania dokumentu. Dzięki zakupowi takiego oprogramowania, zapewniono serwis i pomoc ze strony Poznańskiego Centrum Superkomputerowo-Sieciowego, twórcy i dystrybutora dLibry.

Zespół pracujący nad uruchomieniem repozytorium powstał na bazie osób zaangażowanych już w przygotowanie i publikację materiałów do KPBC. Członkowie zespołu na co dzień pracują w różnych Działach BG UTP. Oddział Gromadzenia – 3 osoby, Czytelnia Główna – 5 osób, Wypożyczal-

⁵ Nauka i gospodarka. Katalog ofert współpracy [online], [dostęp: 7.07.012], dostępny w World Wide Web: <http://www.wspolpraca.utp.edu.pl>.

⁶ E. Karwasińska, M. Rychlik, W kierunku uczelnianego repozytorium cyfrowego [online], [dostęp: 29.08.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui/handle/10593/1797>.

nia – 2 osoby, Dział Informacji Naukowej – 2 osoby. Doświadczenie i praktyka, zdobyte podczas współpracy przy tworzeniu KPBC okazały się bardzo pomocne przy budowie uczelnianego repozytorium.

Następnie opracowano informacje o Projekcie, zasady gromadzenia i udostępniania publikacji, regulamin i wytyczne dla autorów oraz instrukcje dla użytkowników. Wszystkie informacje umieszczono na stronie www RC UTP.

Wykonywanie potrzebnych prac odbywa się na kilku poziomach jednocześnie. Opracowano kolejne etapy procesu od wyboru konkretnego dokumentu, aż po jego publikację w Repozytorium. Wyszczególniono sześć bloków prac, które należy wykonać w tym procesie:

1. selekcja, redakcja, publikacja i rekatalogowanie prac,
2. negocjacje, podpisywanie i archiwizacja umów,
3. skanowanie,
4. przetwarzanie i archiwizacja zeskanowanych plików,
5. obsługa techniczna i informatyczna,
6. zarządzanie i koordynacja wszystkich prac.

Do każdego zadania przypisano poszczególne osoby. Każda osoba otrzymała pisemną instrukcję zleconych prac do przygotowania i ilość obiektów, i godzin do wypracowania w danym roku. W pierwszym roku zespół pracujący przy RC UTP miał do przepracowania ok. 20 godzin (na osobę) miesięcznie. W kolejnych latach realizacji projektu liczba ta zmalała o połowę. Wykonane prace są rejestrowane w „Kartach pracy”, które przekazuje się do Działu Współpracy z Gospodarką. „Karty” stanowią podstawę rozliczenia finansowego.

Repozytorium Cyfrowe UTP

Jak już wspomniano, głównymi adresatami RC są przedsiębiorcy, dlatego pierwotnie skupiono się na materiałach, które mogłyby być przydatne dla tej grupy. Zajęto się selekcją prac doktorskich uwzględniając aspekt aktualności. Decydujący był rok wydania pracy. Skupiono uwagę na najnowszych rozprawach habilitacyjnych UTP. Powołując się na aktualność wyników prac naukowych, selekcji poddano również materiały konferencyjne, które ukazały się w UTP oraz artykuły pracowników opublikowane w czasopiśmie naukowych. Ograniczenia czasowe sprawiły, że w pierwszej kolejności selekcji poddano prace, które już były w magazynach i katalogu Biblioteki Głównej oraz te, których dysponentem majątkowych praw autorskich jest Uczelnia. W przypadku rozpraw habilitacyjnych czy materiałów konferencyjnych skróciło to czas oczekiwania na otrzymanie licencji, z racji „Porozumienia” z Wydawnictwami Uczelnianymi. Treść „Porozumienia UTP” (informacja, która

pojawia się w metadanych, polu prawa) wypracowano już na etapie przygotowywania publikacji do KPBC. Wtedy to uzgodniono z Wydawnictwami Uczelnianymi zasady przekazywania rozpraw do biblioteki cyfrowej. Wraz z przystąpieniem Biblioteki Głównej do budowy własnego repozytorium, Wydawnictwa wyraziły zgodę na powtórne udostępnienie tych publikacji, aneksując wcześniejsze „Porozumienie”. Tak więc materiały, których dysponentem majątkowych praw autorskich jest Uczelnia, są umieszczane w Repozytorium na mocy „Porozumienia UTP”.

Inaczej przedstawiała się kwestia udostępniania w RC UTP prac doktorskich. Właścicielami i dysponentami praw autorskich są w tym przypadku autorzy. Opracowano więc „Licencję UTP”, którą zatwierdził uczelniany Rada Prawny. Treść formularza była wzorowana na licencji Creative Commons (CC). W licencji niewyłącznej autor wyraża zgodę (lub nie) na udostępnianie swojej pracy w Repozytorium UTP i określa jego zakres.

W ramach wyznaczonych zadań, nawiązano kontakt z autorami wybranych prac z kilku ostatnich lat, wysłano do nich formularze licencji i przeprowadzono negocjacje. Część osób entuzjastycznie przyjęła informację o powstałym RC UTP. Świadomość korzyści, jakie płyną z deponowania prac w Repozytorium sprawiła, że autorzy szybko zdecydowali się na podpisanie licencji. Mały procent osób z tej grupy zainteresowany był autoarchiwizacją. Część osób odpowiedziała odmownie na propozycję zdeponowania swojej pracy doktorskiej w RC UTP. W kilku przypadkach tłumaczono to niechęcią do dzielenia się swoją pracą. Najwięcej osób było pozytywnie nastawionych, ale ze względu na słabą orientację w zagadnieniach prawa autorskiego, przyjęło postawę oczekującą.

Efekty pracy redaktorów są pierwszymi z jakimi styka się użytkownik repozytorium. Redaktorzy dokładają wszelkich starań, by informacje o danej publikacji były czytelne, a jednocześnie zgodne z wymogami schematu Dublin Core. Publikacja, która została zakwalifikowana do RC UTP jest opracowywana w dLibrze jako publikacja planowana. Przekazuje się ją następnie do skanowania. Po procesie skanowania publikacja wraca do redaktora, w celu sprawdzenia jakości plików tiff. Po sprawdzeniu poprawności osoba odpowiedzialna za przetwarzanie, zmienia format plików tiff na DjVu. Ona też umieszcza je w specjalnym katalogu, skąd są pobierane i dołączane do treści dokumentu, a następnie publikowane w RC UTP. Proces kończy czynność rekatalogowania.

Pliki tiff powstałe w trakcie skanowania archiwizowane są w wydzielonym miejscu na serwerze.

Wszystkie prace są na bieżąco monitorowane i analizowane. Po uruchomieniu Repozytorium harmonogram prac ulegał zmianom. Na spotkaniach zespołu omawiano i rozwiązywano pojawiające się problemy. Przeważnie

dotyczyły one obiegu dokumentu przeznaczonego do umieszczenia w Repozytorium, czy wypełniania konkretnych pól metadanymi. Obecnie prace przebiegają sprawnie. W przyszłości dla usprawnienia przepływu informacji o pracach nad digitalizacją i publikowaniem dokumentów, planuje się wdrożenie oprogramowania, które umożliwi każdej osobie z zespołu monitorowanie stopnia zaawansowania prac. Rozwiązanie takie proponują autorzy pracy „Tworzenie i rozwój bibliotek cyfrowych”⁷. Dotychczas każdy samodzielnie rejestrował realizację swoich zadań. W miarę potrzeby sporządzał sprawozdania.

W marcu 2011 roku Repozytorium zaistniało w Internecie. Można się do niego „dostać” na trzy sposoby.

- Sposób pierwszy: bezpośrednio, wpisując adres w wyszukiwarce <http://dlibra.utp.edu.pl>;
- Sposób drugi: przez stronę www Biblioteki Głównej UTP <http://www.bg.utp.edu.pl>;
- Sposób trzeci: przez stronę www Uniwersytetu Technologiczno-Przyrodniczego <http://www.utp.edu.pl>.

Rozwój Repozytorium Cyfrowego UTP

Dotychczas wzbogacanie zbiorów repozytorium spoczywało na zespole redakcyjnym. Mamy nadzieję, że uruchomienie i spopularyzowanie modułu „Autoarchiwizacja” zaangażuje samych naukowców w tworzenie zasobu RC UTP.

Autoarchiwizacja jest kolejną zaletą repozytorium. Umożliwia samodzielne deponowanie pracy przez autora. Aby spełniła ona swoje zadanie wymaga konkretnego zaangażowania kadry naukowej. Pomoc zespołu redakcyjnego i administratora jest jak najbardziej wskazana. Według badań ankietowych, pracownicy naukowi oczekują wsparcia z zakresu prawa autorskiego czy wprowadzania metadanych. Autoarchiwizacja, wbrew temu co deklarują naukowcy⁸, powinna mieć charakter obligatoryjny. Odgórny nakaz publikowania w repozytorium przyczynia się do wzrostu jego zasobu. Pracownicy naukowi obawiają się przede wszystkim kolejnych obowiązków, a tym samym dodatkowego obciążenia. Ukazywanie korzyści płynących z „obecności” w repozytorium dla nich jako naukowców i ich studentów, mogą ten opór przełamać.

⁷ C. Mazurek, T. Parkoła, M. Werla, *Tworzenie i rozwój bibliotek cyfrowych*, [w:] *Biblioteki cyfrowe*, red. M. Janiak, M. Krakowska, M. Próchnicka, Warszawa 2012, s. 357.

⁸ M. Rychlik, E. Karwasińska, Projekt powołania repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu. Badanie ankietowe postaw pracowników naukowych [online], [dostęp: 29.08.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui/handle/10593/434>.

Samodzielne deponowanie prac przez autorów z poziomu aplikacji www, wymagało poznania modułu self-archiving i przetestowania go przez samych bibliotekarzy. Po ustaleniu konfiguracji tego mechanizmu, opracowano treść regulaminu dla autorów. Dokumentacja oprogramowania Libra 5.0 opisuje i ilustruje przykładowy scenariusz przekazywania obiektów cyfrowych przez autorów i moderowania procesu publikowania przez redaktora.

Procedura autoarchiwizacji została szczegółowo omówiona w instrukcji umieszczonej na stronie Repozytorium. Zamieszczono tu notatkę o Serwisie Sherpa, w którym autorzy mogą znaleźć informacje dotyczące polityki zamieszczania publikacji w repozytoriach.

Informację o możliwości samodzielnego deponowania prac w Repozytorium umieszczono na stronie www Repozytorium Cyfrowego UTP. Jednocześnie wysłano tę informację do wszystkich pracowników Uczelni drogą mailową. Odnotowano większą (niż zwykle) ilość czytelników. Spodziewano się licznych telefonów, wiadomości mailowych, a tym samym intensywnej współpracy. Tymczasem było kilka zapytań i tylko jeden autor wykonał samodzielną archiwizację, a inny wyraził chęć współpracy.

Mamy świadomość, że promocja Repozytorium i idei autoarchiwizacji w szczególności, wymaga jeszcze wielu zabiegów. Na nich (obok bieżących prac związanych z prowadzeniem RC) planujemy skupić się w najbliższej przyszłości.

W procesie dydaktycznym potrzebne są publikacje dostępne dotychczas w jednym egzemplarzu. Liczba tego typu publikacji nie jest na razie imponująca, ale ich wykorzystanie wskazuje na słuszność podjętego działania. W przyszłości po zakończeniu projektu unijnego planuje się umieszczać w Repozytorium skrypty i materiały dydaktyczne.

Uruchomienie środków na realizację projektu wiązało się z ograniczeniami czasowymi. Z tego też powodu szerszą promocję RC wśród pracowników naukowych postanowiono przeprowadzić dopiero w nowym roku akademickim (2012/2013). Studentom zaczynającym naukę w UTP, Repozytorium zostanie zaprezentowane podczas przysposobienia bibliotecznego. Planuje się przeprowadzenie szkoleń dla wszystkich zainteresowanych współpracą na rzecz Repozytorium. Informację o samym Repozytorium i sposobie korzystania przesłano za pośrednictwem Działu Informacji i Promocji drogą elektroniczną do pracowników uczelni, choć jak już wspomniano, nie wpłynęło to na intensyfikację współpracy. Dodatkowo w biuletynie uczelnianym „Format UTP Kwartalnik Uniwersytetu Technologiczno-Przyrodniczego” pojawił się artykuł na temat RC UTP⁹. Kolejny artykuł poszerzający informacje o Repozytorium został już wysłany do Wydawnictwa.

⁹ *Format UTP*, „Kwartalnik Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy” 2011, nr 3(58).

Ukaże się on w kolejnym numerze biuletynu. Jednocześnie z promocją i popularyzacją RC wśród społeczności uczelnianej czeka nas szereg spotkań z władzami uczelni. Obowiązek deponowania w BG rozpraw doktorskich w postaci drukowanej i elektronicznej wymaga wydania odpowiedniego rozporządzenia. Liczymy na to, że rozporządzenie będzie obligowało autora pracy do dołączenia opisu pracy na przygotowanym formularzu, jak i do podpisania zgody (licencji) na umieszczenie pracy w RC UTP. Mamy nadzieję, iż tą drogą usprawnimy pozyskiwanie publikacji, a tym samym zwiększymy zasób Repozytorium. Dodatkową mobilizacją dla autorów ma być, zgodnie z planami, uwzględnianie ich aktywności na rzecz repozytorium przy okresowej ocenie dokonywanej przez Władze uczelni. Zdajemy sobie sprawę z ogromu pracy jaki jeszcze przed nami. Jesteśmy jednak pewni, że jest to wysiłek, który należy podjąć.

Mimo że promocja Repozytorium nie „ruszyła” jeszcze w pełni, to statystyki wykorzystania pokazują, że jest ono często odwiedzane przez użytkowników. Od marca 2011 roku łączna liczba użytkowników wyniosła ponad 20 tys. W skali miesiąca notujemy ok. 2 tys. odwiedzin. Niestety nie mając możliwości rozróżnienia użytkowników (przedsiębiorców i pracowników UTP czy studentów) nie jesteśmy w stanie określić stopnia wykorzystania repozytorium przez każdą z tych grup.

W pierwszym roku działalności umieściliśmy w RC UTP łącznie 250 publikacji. Obecnie zasób obejmuje 348 pozycji, w tym 48 rozpraw habilitacyjnych, 119 prac doktorskich, 81 materiałów konferencyjnych i 94 artykuły z czasopism naukowych oraz publikacja zawierająca wyniki badań. Najczęściej przeglądaną kolekcją są właśnie prace doktorskie (1034 wyświetlenia). Dużą popularnością cieszą się rozprawy habilitacyjne (974 wyświetlenia). O połowę mniej wyświetleń (504) mają materiały konferencyjne. Najmniejszą popularnością cieszą się artykuły z czasopism (293)¹⁰. Co więcej, umieszczając tę samą rozprawę habilitacyjną w KPBC, gdzie tylko kilku użytkowników po nią sięgnęło, w RC UTP zanotowano 400 odsłon tego tytułu. Największą liczbę odwiedzin odnotowano w styczniu 2012 roku – ok. 750 wyszukiwań.

Archiwizacja

Budowaniu RC UTP towarzyszyło ciągłe pytanie, jak zabezpieczyć pliki cyfrowe, powstające w procesie digitalizacji. Ich przechowywanie jest jednym z istotnych zadań repozytorium. Ze względu na postęp w technologii kom-

¹⁰ Stan na 28 sierpnia 2012 r.

puterowej i oprogramowaniu ważne jest, aby długotrwale przechowywane były pliki jak najmniej przetworzone, tzw. „pliki pierwotne”. W przypadku prac nad materiałami publikowanymi w KPBC, pliki te na początku zgrywano na płyty CD. Ich przechowywaniem i inwentaryzacją zajmowali się pracownicy Działu Informacji Naukowej. Obecnie pliki publikacji z KPBC i RC UTP są przechowywane na serwerze Biblioteki Głównej UTP. Kolejnym krokiem w rozwoju bibliotek cyfrowych, a tym samym repozytoriów, jest rozwiązanie problemu długoterminowej archiwizacji zgromadzonych danych. Całkowita ich utrata może nastąpić na skutek klęski żywiołowej, takiej jak powódź czy pożar, lub awarii, kradzieży sprzętu, aktów wandalizmu. Przechowywanie danych na znanych nośnikach, też nie jest wyjściem „na dłuższą metę”. Trwałość dysku twardego i CD/DVD szacuje się w zależności od obciążenia na 5–10 lat, pamięci flash – 10 lat, taśmy LTO – 15–30 lat¹¹. Informacje należy również chronić przed przestępstwami informatycznymi, np.: usunięciem danych, czy ich modyfikacją. Wysoki poziom bezpieczeństwa informacji może dać jedynie przechowywanie ich kopii w odległych geograficznie miejscach. Wymaga to jednak zastosowania zaawansowanego oprogramowania zarządzającego danymi.

Ważnym projektem o zasięgu krajowym, związanym z rozwojem infrastruktury nauki (sieć PIONIER), a szczególnie z rozwojem zaawansowanych aplikacji i usług teleinformatycznych jest projekt Platforma Obsługi Nauki PLATON¹². Projekt jest realizowany przez 22 jednostki, w tym Uniwersytet Technologiczno-Przyrodniczy jako jednostkę koordynującą i wiodącą w Miejskiej Akademickiej Sieci Komputerowej BYDMAN. Bezpośrednim celem projektu PLATON jest wdrożenie nowoczesnych usług teleinformatycznych: wideokonferencji, eduroam, usług kampusowych, powszechnej archiwizacji, naukowej interaktywnej telewizji, dostępnych dla środowiska naukowego w Polsce. Na platformie PLATON naukowiec otrzymuje prosty i bezpośredni dostęp do świata wirtualnego obejmującego aplikacje, laboratoria i Internet w trybie „zawsze i wszędzie”. Usługa powszechnej archiwizacji natomiast daje możliwość przechowywania, ochrony danych i zabezpieczania wyników badań, w takiej lokalizacji, która jest najwygodniejsza dla użytkownika¹³. Wszelkonostronną ochronę danych uzyskuje się przez stosowanie odpowiednich protokołów przesyłania

¹¹ N. Meyer, M. Brzeźniak, M. Jankowski, Wprowadzenie do usługi. Warsztaty promocyjne dla użytkowników Usługi Powszechnej Archiwizacji, [online], [dostęp: 29.08.2012], dostępny w World Wide Web: https://www.storage.pionier.net.pl/wiki/images/a/a0/Prezentacja_wprowadzenie.pdf.

¹² Polski Internet Optyczny PIONIER. Projekt PLATON [online], [dostęp: 19.07.2012], dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.

¹³ *Pionier Magazine* [online], [dostęp: 29.08.2012], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1365/PLATON_specjalnie.html.

danych, mechanizmów szyfrowania informacji i autoryzacji użytkowników oraz zabezpieczeń programowych i sprzętowych (np. kontrola dostępu, firewall), precyzyjnie skonfigurowanych przez specjalistów zajmujących się bezpieczeństwem. Usługa powszechnej archiwizacji może stać się bardzo przydatna dla wielu instytucji nauki i edukacji. Nie posiadając rozbudowanej w tym względzie infrastruktury i specjalistycznego oprogramowania, będą one zmuszone zlecić tę usługę instytucji zewnętrznej. Biblioteka Główna korzysta z wszelkich możliwych sposobów poszerzenia wiedzy na w/w temat. W marcu tego roku pracownicy BG UTP wzięli udział w szkoleniu, organizowanym przez PCSS w Toruniu na temat usługi powszechnej archiwizacji systemu PLATON (U4).

Usługi Powszechnej Archiwizacji

Jednostką, która zajmuje się koordynacją wszelkich zagadnień informatycznych w skali całego Uniwersytetu Technologiczno-Przyrodniczego jest Uczelniany Ośrodek Rozległych Sieci Komputerowych (UORSK).

Biorąc po uwagę charakter i miejsce BG UTP w strukturze uczelni, wydaje się rzeczą naturalną, że swoje plany w kwestii archiwizacji danych musi ona rozpatrywać na podstawie strategii działania UTP. Właściwym rozwiązaniem jest raczej zintensyfikowanie wysiłków, by potrzeby BG były uwzględniane w strategii rozwoju sieci dla całej uczelni, niż poszukiwanie odrębnych rozwiązań.

Biblioteka Główna podjęła kroki porządkowania i przygotowania danych z RC UTP w celu archiwizowania w systemie PLATON.

RC UTP nie przestanie działać z chwilą zakończenia realizacji projektu „Naukowcy UTP dla Gospodarki Regionu”. Biblioteka Główna UTP nadal będzie powiększać zasób materiałów dostępnych w Repozytorium, publikować prace doktorskie i habilitacyjne, materiały konferencyjne, czy artykuły naukowców UTP. Planowane jest powiększenie zasobu o skrypty uczelniane, których nakład został wyczerpany, a jest na nie zapotrzebowanie oraz umożliwienie udostępniania studentom materiałów edukacyjnych przez wykładowców.

Podsumowanie

Środki unijne pozwoliły na stworzenie i uruchomienie Repozytorium. W przyszłości planuje się wykorzystać to narzędzie bardziej efektywnie, by zaspokoić potrzeby w dziedzinie informacji naukowej nie tylko regionalnego przemysłu, ale rodzimej kadry naukowo-dydaktycznej i społeczności studenckiej.

Repozytorium ma stanowić powód zacieśnienia współpracy środowiska uczelnianego z Biblioteką Główną, a tym samym być w centrum procesu naukowo-dydaktycznego uczelni.

Bibliografia

1. Brzeźniak M., PLATON, Usługi powszechnej archiwizacji, [online], [dostęp: 29.08.2012], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1374/PLATON_Uslugi_powszechej_archiwizacji.html/
2. *Format UTP*, „Kwartalnik Uniwersytetu Technologiczno-Przyrodniczego w Bydgoszczy” 2011, nr 3(58).
3. Karwasińska E., Rychlik M., *W kierunku uczelnianego repozytorium cyfrowego* [online], [dostęp: 29.08.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui/handle/10593/1797>.
4. Mazurek C., Parkoła T., Werla M., *Tworzenie i rozwój bibliotek cyfrowych*, [w:] *Biblioteki cyfrowe*, red. M. Janiak, M. Krakowska, M. Próchnicka, Warszawa 2012, s. 341–361.
5. Meyer N., Brzeźniak M., Jankowski M., Wprowadzenie do usługi. Warsztaty promocyjne dla użytkowników Usługi Powszechnej Archiwizacji [online], [dostęp: 29.08.2012], dostępny w World Wide Web: https://www.storage.pionier.net.pl/wiki/images/a/a0/Prezentacja_wprowadzenie.pdf.
6. Ministerstwo Pracy i Polityki Społecznej – Departament Wdrażania Europejskiego Funduszu Społecznego, O programie [online], [dostęp: 4.07.2012], dostępny w World Wide Web: <http://www.kapitalludzki.gov.pl/o-programie>.
7. Nauka i gospodarka. Katalog ofert współpracy [online], [dostęp: 7.07.2012], dostępny w World Wide Web: <http://www.wspolpraca.utp.edu.pl>.
8. Pionier Magazine [online], [dostęp: 29.08.2012], dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1365/PLATON_specjalnie.html.
9. Polski Internet Optyczny PIONIER. Projekt PLATON [online], [dostęp 19.07.2012], dostępny w World Wide Web: <http://www.platon.pionier.net.pl/online>.
10. Realizacja II etapu Regionalnego Centrum Innowacyjności. Mój region w Europie [online], [dostęp 29.08.2012], dostępny w World Wide Web: <http://www.rci-2etap.utp.edu.pl>.
11. Rychlik M., Karwasińska E., Projekt powołania repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu. Badanie ankietowe postaw pracowników naukowych [online], [dostęp: 29.08.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui/handle/10593/434>.
12. Rychlik M., Karwasińska E., Repozytorium instytucjonalne jako czynnik wspomagający rozwój nauki w środowisku akademickim [online], [dostęp: 29.08.2012]. Dostępny w World Wide Web: https://repozytorium.amu.edu.pl/jspui/bitstream/10593/1074/1/Biblioteka_07.pdf.

Marta Stąporek

Biblioteka Politechniki Krakowskiej

Organizacja i funkcjonalność wiarygodnego archiwum cyfrowego uczelni wyższej na przykładzie Repozytorium Politechniki Krakowskiej

Słowa kluczowe: komunikacja naukowa, repozytorium instytucjonalne, Repozytorium Politechniki Krakowskiej, SUW, wiarygodność archiwum cyfrowego

Abstrakt: w artykule zaprezentowano Repozytorium Politechniki Krakowskiej (RPK), które jest częścią Zintegrowanego Systemu Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych (SUW). Przedstawiono zagadnienia dotyczące struktury organizacyjnej RPK, sposobów gromadzenia i deponowania w nim dokumentów, sposobu prezentacji danych o zamieszczanych dokumentach, możliwości przeszukiwania i wyszukiwania jego zasobów oraz zasad ich udostępniania. Omówiono działania podejmowane przez Bibliotekę PK na rzecz promocji RPK w lokalnym środowisku. Scharakteryzowano Internet i narzędzia zapewniające szybszy przepływ informacji i usprawniające komunikację naukową. Podjęto próbę określenia kryteriów wiarygodności archiwum cyfrowego. Oceniono, w jakim zakresie spełnia je RPK.

Internet jako medium wspomagające i upowszechniające naukę

Narzędzia informatyczne i tworzone dzięki nim nowe usługi internetowe wykorzystywane były i nadal są przez różne grupy społeczne i zawodowe, również przez szeroko rozumianą społeczność badaczy i studentów. Ta bowiem, mając świadomość znaczenia informacji i komunikacji, czyni starania ku efektywnemu zastosowaniu i wykorzystaniu istniejących możliwości Internetu do własnych, specyficznych potrzeb.

Już w 1996 r. w Stanach Zjednoczonych powstało konsorcjum Internet2, rozwijające i wdrażające technologie sieciowe, wykorzystujące szybki transfer danych. Obecnie zrzesza 221 uniwersytetów, agencje rządowe, laboratoria badawcze, regionalne i stanowe organizacje oraz partnerów branży informatycznej, reprezentowane przez

ponad 50 państw. Główne założenia projektowe konsorcjum skupiają się na edukacji i badaniach naukowych¹. To dzięki pracy konsorcjum nad praktycznym zastosowaniem Internetu, możemy korzystać dziś z bibliotek cyfrowych, wirtualnych laboratoriów, wideokonferencji czy też platform zdalnego nauczania.

Obecne rozwiązania technologiczne Internetu oferują naukowcom, badaczom i pozostałym uczestnikom komunikacji naukowej szereg możliwości ułatwiających i usprawniających: wyszukiwanie, gromadzenie i porządkowanie informacji; uaktualnianie własnej wiedzy; publikowanie i recenzowanie artykułów naukowych; uczestniczenie w projektach i eksperymentach badawczych; współtworzenie przez wielu autorów równocześnie tego samego dokumentu; komunikację interpersonalną; przekazywanie i pozyskiwanie informacji o nadchodzących wydarzeniach².

Możliwości te realizowane są za pośrednictwem platform internetowych, dających możliwość personalizacji (profil użytkownika) oraz komunikacji. Dobrym przykładem społecznościowej platformy naukowej, integrującej w sobie równocześnie wiele funkcji, jest BioInfoBank Library³. Platforma pozwala na publikowanie artykułów naukowych online, pełni funkcję repozytorium materiałów dydaktycznych, platformy e-learningowej, narzędzia do komunikacji, zawiera bazy danych, udostępnia też najświeższe informacje ze świata nauki⁴.

Przykładów tego typu projektów, realizowanych w środowisku sieciowym i dedykowanych nauce jest wiele (np. Academia.edu⁵, ResearchGate⁶, Nature Network⁷, Peer Evaluation⁸). Zachętą do ich tworzenia jest nie tylko rozwój nauki i ruchu Open Access (OA), ale również zalecenia międzynaro-

¹ Internet2 [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.internet2.edu>.

² E. Kulczycki, *Wykorzystanie mediów społecznościowych przez akademickie uczelnie wyższe w Polsce. Badania w formule otwartego notatnika* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://amu.academia.edu/EmanuelKulczycki/Papers/1766754/Wykorzystanie_mediow_spolecznosciowych_przez_akademickie_uczelnie_wyzsze_w_Polsce._Badania_w_formule_otwartego_notatnika.

³ L. Rychlewski, R. Dudek, E. Rodakowska, *BioInfoBank Library. Narzędzie dla naukowców i uczelni wyższych* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://lib.psnc.pl/Content/311/BC-16-Rychlewski.pdf>.

⁴ BioInfoBank Library [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://lib.bioinfo.pl>.

⁵ Akademia.edu [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://academia.edu>.

⁶ ResearchGate [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.researchgate.net>.

⁷ Nature Network [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://network.nature.com>.

⁸ Peer Evaluation [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.peerevaluation.org>.

dowych organizacji i rządów poszczególnych państw, chętnie finansujących tego typu projekty.

Rosnąca liczba narzędzi i zasobów internetowych – baz danych, portali społecznościowych, serwisów internetowych, blogów, stron www, forów dyskusyjnych⁹ – oferuje niezliczone ilości informacji, pochodzących ze stron www tysięcy organizacji i od milionów użytkowników indywidualnych.

Internet jest środowiskiem dynamicznym, zmieniającym się codziennie nie tylko co do ilości zasobów, ale również co do ich zawartości i struktury. W analizach poświęconych zamieszczanym treściom i informacjom w Internecie spotkać można wiele zarzutów dotyczących ich wiarygodności. Ta budowana jest przez szereg czynników, takich jak choćby dokładność, kompletność, spójność, aktualność, uporządkowanie bądź odpowiednie sklasyfikowanie treści/informacji¹⁰.

Wiarygodność informacji zależy przede wszystkim od wiarygodności jej nadawcy. Nadawca wiarygodny to taki, który jest kompetentny w swojej dziedzinie, jest ekspertem i autorytetem, daje gwarancje rzetelności, wysokiej jakości informacji oraz wiedzy, jaką posiada. Nadawca wiarygodny to również ktoś, kto potrafi odpowiednio zaprezentować posiadaną wiedzę w sposób uporządkowany, logiczny, spójny, ale również obiektywny. Przesądza ona o tym, czy odbiorca skorzysta z oferowanych przez nadawcę informacji¹¹.

Biblioteki naukowe i ich bibliotekarze, wykorzystując różnego rodzaju narzędzia internetowe, powinni zdawać sobie sprawę z odpowiedzialności, jaką ponoszą w procesie komunikacji naukowej, jako wiarygodni nadawcy. Na nich ciąży odpowiedzialność udostępniania i upowszechniania wysokiej jakości informacji i zasobów. Odpowiednie prezentowanie i świadczenie usług w Internecie, szybka i skuteczna komunikacja z użytkownikami za pośrednictwem sieci, sprzyja kształtowaniu pozytywnego wizerunku biblioteki, a opracowane, opisane i odpowiednio uporządkowane elektroniczne zasoby to atut, który umacnia jej autorytet, jako rzetelnego i wiarygodnego nadawcy.

Warto przeglądać i oceniać serwisy biblioteczne, biblioteki cyfrowe, repozytoria czy też większe systemy udostępniania nauki, czyniąc to w oparciu o opracowane dotychczas kryteria oceny tego typu narzędzi. Analiza czynni-

⁹ S. Cisek, *Nauka 2.0. Nowe narzędzia komunikacji naukowej* [online], [dostęp: 24.09. 2012], dostępny w World Wide Web: <http://informacjacyfrowa.wsb.edu.pl/pdfs/nauka%202.0.pdf>.

¹⁰ J. Boruszewski, *Jakość i wiarygodność informacji w infobrokerstwie* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://lingua.amu.edu.pl/Lingua_22/13_Jaroslawa%20Boruszewski.pdf.

¹¹ K. Stankiewicz, *Wpływ Internetu na percepcję wiarygodności informacji*, [w:] *Spółczesność informacyjna. Wizja czy rzeczywistość?*, II ogólnopolska konferencja naukowa: Kraków 30 maja 2003, t. II [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://winntbg.bg.agh.edu.pl/skrypty2/0096/407-415.pdf>.

ków świadczących o wysokiej jakości i wiarygodności ma decydujący wpływ na podniesienie jakości własnych narzędzi oferowanych przez biblioteki.

Metody i kryteria oceny serwisów internetowych opracowywane były przez różne osoby (Helge Clausen, Alistair Smith) i instytucje (np. Komisja Europejska w 2001 r. zaproponowała kryteria oceny stron o tematyce zdrowotnej). W literaturze przedmiotu wyróżniono kilka metod służących ocenie serwisów internetowych: 1) metoda stosowania automatycznych procedur, polegająca na wykorzystaniu programów komputerowych, oceniających mechanicznie strony; 2) metoda statystyczna analizująca typ i ilość informacji (liczenie odnośników); 3) metoda jakościowo-heurystyczna polegająca na ocenie wybranej grupy cech formalnych czy treściowych (wg Helge Clausena); 4) metoda kognitywna (eksperti prowadzą test „na użytkownika”, na podstawie którego oceniają, czy dane źródło spełnia wymagania zdefiniowane wcześniej przez użytkowników)¹².

Chętnie wykorzystywaną metodą jest metoda jakościowo-heurystyczna Helge Clausena. Jej wykorzystanie nie wymaga zakupu bądź tworzenia programu do analizy i opracowania jej wyników. Główne kryteria oceny według Helge Clausena zastosowane przy użyciu metody jakościowo-heurystycznej to: struktura i wygląd stron (organizacja treści, racjonalne użycie grafiki), jakość informacji (instytucja sprawcza, aktualizacja, poprawność, ilość informacji), linki i nawigacja (odpowiedniość, adnotacje, liczba odsyłaczy, łatwość wertowania), wrażenie estetyczne (kolory, czcionki, format), miscellanea (pomoc, użytkownik), ogólne (oryginalność, przyjazność serwisu).

Niniejszy artykuł jest próbą scharakteryzowania czynników wpływających na wiarygodność serwisu Repozytorium Politechniki Krakowskiej (RPK) w oparciu o jego cechy formalne. Główną uwagę zwrócono na dwa aspekty serwisu – jego jakość użytkową oraz jakość zawartości informacyjnej. Wpływają one bowiem na wizerunek wiarygodnego medium nauki, zachęcają do korzystania ze zarchiwizowanych wcześniej w repozytorium zasobów, wykorzystania go jako przyjaznego, łatwego w obsłudze narzędzia do deponowania własnego dorobku, dzielenia się nim ze społecznością sieciową; wreszcie czynniki te mają wpływ na budowanie zaufania społecznego do projektu.

Przez jakość użytkową rozumiana jest tu możliwość zastosowania i wykorzystania RPK przez jego użytkowników końcowych; natomiast jakość zawartości informacyjnej określa sposób przedstawienia serwisu użytkownikom.

¹² B. Bednarek-Michalska, *Oceń jakość informacji elektronicznej!* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://biblioteka.oeiizk.waw.pl/konferencyjne/michalska.pdf>.

Repozytorium Politechniki Krakowskiej – pomysł i etapy realizacji w obrębie projektu SUW

Ogólna informacja o RPK

W styczniu 2012 r. Rektor PK, prof. dr hab. inż. Kazimierz Furtak, dokonał uroczystego otwarcia RPK. Realizacja instytucjonalnego repozytorium rozpoczęła się wraz z realizacją projektu SUW: Zintegrowanego Systemu Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych¹³. Do opracowywania wstępnej koncepcji projektu SUW przystąpiono w listopadzie 2008 r. W Bibliotece PK (BPK) powołano kilkunastoosobowy zespół projektowy, którego zadaniem było rozeznanie się w temacie, a następnie realizacja zadań szczegółowych skoncentrowanych na tworzeniu systemu (infrastruktura informatyczna, aspekty prawne i merytoryczne systemu, promocja i szkolenia). W kwietniu 2009 r. w Ministerstwie Nauki i Szkolnictwa Wyższego Politechnika Krakowska złożyła wniosek opracowany przez BPK o dofinansowanie projektu SUW (rozpatrzony pozytywnie, 25 września 2009 r. MNiSW podpisało umowę z PK o dofinansowanie).

Projekt SUW w ogólnym zamierzeniu służyć ma udostępnianiu w Internecie informacji naukowej, w tym informacji o charakterze bibliograficznym, oraz pełnotekstowych materiałów naukowych (dydaktycznych i edukacyjnych), których autorami są pracownicy i studenci Uczelni. System obsługiwany jest za pomocą opracowanego w BPK oprogramowania, które pozwala zintegrować poszczególne elementy systemu w spójną całość. Dzięki temu użytkownik będzie mógł przeszukiwać wszystkie zbiory i zasoby BPK z jednego miejsca. W skład systemu SUW wchodzi następujące elementy: 1. Repozytorium Politechniki Krakowskiej (w tym zasoby BC PK), 2. Bibliografia Publikacji Pracowników PK, 3. Komputerowy Katalog BPK, 4. Bazy własne BPK (Architektura i Sztuka Krakowa, Konferencje Krynickie – Referaty, Niepublikowane Prace Naukowe Pracowników PK), 5. Baza Podręczników, 6. Portal społecznościowy¹⁴.

W 2009 r. prace zespołu SUW skupiły się przede wszystkim na realizacji instytucjonalnego repozytorium. Wersja testowa udostępniona została

¹³ Zintegrowany System Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.biblos.pk.edu.pl/suw>.

¹⁴ H. Juszczak, M. Stąporek, *The Technical Sciences Knowledge Exchange and Academic Publications Sharing Integrated System (ABBR, SUW) as an example of academic system to share knowledge and communicate between the university environment* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.bg.pw.edu.pl/iatul2011/proceedings/ft/Staporek_M.pdf.

użytkownikom w listopadzie 2011 r. Podstawę zasobów RPK we wstępnej fazie tworzenia stanowiły wchłonięte zasoby i metadane Biblioteki Cyfrowej Politechniki Krakowskiej (BC PK). BC PK tworzona od 2006 r., archiwizowała i udostępniała online pełne teksty publikacji wydanych na PK – prace doktorskie, prace habilitacyjne, skrypty oraz artykuły opublikowane w „Czasopiśmie Technicznym”.

Postanowienia ogólne dotyczące funkcjonowania RPK (zasady gromadzenia, deponowania, udostępniania i przechowywania jego zasobów) określa zatwierdzony przez Senat PK „Regulamin Repozytorium PK”¹⁵. Dostęp do zasobów RPK możliwy jest z portalu BPK, z katalogu BPK, z bazy „Bibliografia Publikacji Pracowników PK” oraz poprzez wyszukiwarki internetowe.

Jakość użytkowa RPK

Udostępnianie i promowanie wyników działalności naukowo-badawczej, główny cel RPK, realizowany jest dzięki jego funkcjonalności, którą zapewniają:

- zgromadzenie w jednym miejscu dokumentów i publikacji z zakresu nauk technicznych,
- możliwość dokonywania samodzielnej archiwizacji dokumentów przez autorów,
- możliwość zarządzania przez użytkownika dokumentami zgromadzonymi w RPK (opcja: „Moje zasoby” i „Ulubione zasoby”),
- uporządkowanie i opracowanie zgromadzonych zbiorów w sposób logiczny i konsekwentny (dzięki odpowiednio przygotowanemu formularzowi do deponowania dokumentów),
- zapewnienie trwałego i długoterminowego przechowywania dokumentów,
- zapewnienie swobodnego dostępu do zasobów zdeponowanych przez studentów i pracowników PK,
- zapewnienie możliwości wyszukiwania i przeglądania materiałów według różnych kryteriów,
- możliwość śledzenia statystyk dla zdeponowanych dokumentów (dane dotyczące: liczby pobrań pojedynczych dokumentów, liczby zgromadzonych dokumentów w poszczególnych kolekcjach RPK, liczby nowo zdeponowanych w RPK dokumentów w ciągu minionych dni),

¹⁵ Regulamin Repozytorium Politechniki Krakowskiej. Repozytorium Politechniki Krakowskiej [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://suw.biblos.pk.edu.pl/regulamin>.

- możliwość śledzenia dokumentów i zmian dokonywanych w poszczególnych kolekcjach (opcja: „Dodaj do ulubionych”, kanał RSS),
- zapewnienie komunikacji między zalogowanymi do systemu użytkownikami (system umożliwia rankingowanie zdeponowanych zasobów, ich komentowanie, wyrażanie własnych opinii o zasobach RPK),
- zapewnienie dokumentom RPK widoczności w sieci (metadane rozpoznawalne przez internetowe wyszukiwarki).

Gromadzenie i opracowanie zasobów RPK

RPK gromadzi cyfrowe wersje dokumentów, których autorami są pracownicy i studenci Uczelni lub też ich wydawcą jest PK. Gromadzone są również dokumenty związane z działalnością PK (kolekcja „Materiały informacyjne PK”) oraz cyfrowe wersje zbiorów BPK. Główne warunki stawiane deponowanym dokumentom dotyczą: 1) ich zawartości (preferowane treści o charakterze naukowym bądź dydaktycznym), 2) gwarancji autora, że deponowanie w archiwum zasobu odbywa się zgodnie z obowiązującym prawem autorskim oraz 3) decyzji autora odnoszącej się do stopnia udostępnienia zdeponowanego dokumentu użytkownikom Internetu.

Autorzy za pośrednictwem formularza mogą deponować dokumenty różnego rodzaju (artykuły z czasopism, książki, rozdziały lub fragmenty książek, wyniki badań i raporty naukowe, sprawozdania naukowe, prace dyplomowe, patenty, materiały i prezentacje konferencyjne, materiały do zajęć dydaktycznych, materiały dotyczące jednostek PK, filmy, rękopisy, preprinty), zapisane w powszechnie używanych formatach (pdf, html, jpg).

Deponowania dokumentów autorzy mogą dokonywać samodzielnie w dogodnym dla nich czasie i z dowolnego miejsca, pod warunkiem że mają dostęp do Internetu (deponowanie bezpośrednio). Pomocą w zakresie deponowania dokumentów służą bibliotekarze BPK oraz grupa redaktorów wydziałowych (osób odpowiedzialnych za upowszechnianie informacji o systemie SUW i RPK na swoich wydziałach, wskazanych przez dziekanów PK, współpracujących bezpośrednio w tym zakresie z BPK). Deponowanie dokumentów odbywa się z platformy RPK (za pośrednictwem systemu SUW), która nie wymaga od użytkownika dodatkowej rejestracji. Logowanie do RPK odbywa się na takich samych zasadach, jak logowanie do systemu bibliotecznego. Każda osoba, będąca użytkownikiem BPK, dysponująca ważną kartą biblioteczną (dla studentów – elektroniczna legitymacja studencka, dla pracowników – uczelniana karta identyfikacyjna), może natychmiast zalogować się bez konieczności rejestrowania się w RPK i zakładania konta z nowym identyfikatorem i nowo zdefiniowanym

hasłem (podwójna korzyść dla użytkownika: oszczędność czasu i odciążenie jego pamięci).

Zespół ds. merytorycznych projektu SUW zdecydował, że opracowanie dokumentów odbywać się będzie we współdziałaniu z ich autorami. Przygotowany został odpowiedni formularz, za pośrednictwem którego dokonywany jest opis bibliograficzny dokumentu deponowanego w RPK. Im więcej elementów opisu bibliograficznego zostanie wypełnionych, tym pełniejszą informację o dokumencie otrzymują użytkownicy końcowi. Szczegółowy opis (metadane) wpływa również na widoczność zasobu w Internecie – wyszukiwalność dokumentów za pomocą wyszukiwarek ogólnych i naukowych.

Proces deponowania dokumentów odbywa się w kilku etapach:

1. autor wybiera rodzaj deponowanego dokumentu oraz kolekcję, do której chce przypisać dokument,
2. autor wypełnia formularz, wprowadzając metadane dokumentu,
3. autor określa licencję i prawa dostępu do deponowanego dokumentu,
4. autor przesyła metadane wraz z dołączonym dokumentem do akceptacji redaktorom RPK,
5. akceptacja redaktorów i włączenie dokumentu do zasobów RPK.

Wg „Regulaminu Repozytorium PK” zdeponowane dokumenty nie ulegają usunięciu z archiwum. Jeśli dokument przeszedł weryfikację pozytywnie u redaktora RPK, niezwłocznie włączany jest do zasobów RPK; od tej pory autor nie ma możliwości samodzielnego dokonywania w nim zmian, ani usunięcia go z cyfrowego archiwum. Prawa takie zachowuje jednak redaktor RPK, uprawniony do usunięcia dokumentów naruszających w jakikolwiek sposób prawa autorskie bez konieczności uzyskania na to zgody przez autora.

Na opis zasobu składają się: 1) elementy wymagane, które autor uzupełnia obowiązkowo (tytuł, autor, data wydania, wydawca, klasyfikacja tematyczna, jednostka PK), 2) elementy zalecane (abstrakt, format, słowa kluczowe, język, link do katalogu BPK) oraz 3) elementy opcjonalne (współtwórca, wariant tytułu, uwagi, uwagi dla redaktora RPK).

Udostępnianie i wyszukiwanie zasobów RPK

Zasoby RPK udostępniane są bezpłatnie w Internecie, jednakże nie do wszystkich użytkowników mają pełny, nieograniczony dostęp. Wynika to z półotwartego charakteru archiwum, które tworzone jest przede wszystkim z myślą o społeczności macierzystej Uczelni oraz w oparciu o poszanowanie i respektowanie praw autorskich. Decyzję o udostępnianiu deponowanych dokumentów w RPK jego twórcy pozostawili autorom.

Pełne opisy dokumentów (metadane) dostępne są dla wszystkich zainteresowanych użytkowników sieci Internet. Opisy prezentowane są w formie skróconej (autor dokumentu, tytuł, rodzaj), użytkownik natomiast w miarę własnych potrzeb informacyjnych ma możliwość rozwinięcia opisu o kolejne jego elementy, uporządkowane w poszczególnych sekcjach:

- tematyka (abstrakt, klasyfikacja, jednostka PK),
- zasób (format, link do pełnego tekstu, rodzaj licencji, prawa dostępu dla użytkowników, liczba pobrań),
- odnośniki (linki zewnętrzne: do katalogu BPK, do „Bibliografii Publikacji Pracowników PK”, do pozostałych dokumentów kolekcji, do której został dany dokument przypisany, opinie użytkowników),
- dyskusje o zasobie (możliwość prześledzenia wątków na forum dyskusyjnym, poświęconych dokumentowi).

Decyzja o dostępie do pełnej wersji dokumentu pozostaje w gestii jego autora, który na etapie deponowania przypisuje mu rodzaj licencji (Creative Commons, Domena publiczna, Licencja PK). W przypadku wyboru licencji PK, autor może zawęzić grupę użytkowników mających prawo otworzyć pełny tekst dokumentu – 1) zalogowani użytkownicy w RPK, 2) użytkownicy korzystający z: a) komputerów w domenie PK, b) komputerów BPK, 3) pracownicy BPK.

Przeglądanie dokumentów możliwe jest według nazwiska autora, kolekcji, w których zamieszczono dokument, według wydziału, z którego pochodzi autor deponowanego zasobu oraz według ostatnio opublikowanych w repozytorium dokumentów (z możliwością ograniczenia liczby dni). Wyszukiwanie natomiast realizowane jest poprzez wpisanie w pole wyszukiwania słowa/słów kluczowych, frazy wyszukiwawczej bądź nazwiska autora.

Istnieje również możliwość zawężania wyników wyszukiwania według kategorii: główne kolekcje RPK (1. Artykuły i czasopisma, 2. Publikacje książkowe, 3. Materiały konferencyjne, 4. Prace dyplomowe, 5. Materiały informacyjne PK), rodzaje zasobów (książki, artykuły, prace dyplomowe, multimedia), jednostki PK (wydziały Uczelni, jednostki międzywydziałowe, Biblioteka PK, Administracja, Zbiory BPK), data wydania, format dokumentu (pdf, html, jpg), język dokumentu (oferowanych dwanaście).

Przechowywanie dokumentów i ich bezpieczeństwo

W „Regulaminie Repozytorium PK” zawarte zostały również postanowienia dotyczące przechowywania i bezpieczeństwa zdeponowanych dokumentów. Zagwarantowano w nich poszanowanie praw własności intelektualnej, bezterminową archiwizację dokumentów przy zachowaniu polityki bezpieczeństwa danych. Znajduje się tu również zapis o niemożności ingerencji autorów

w treść opublikowanych już w RPK dokumentów, jak i o niemożności samowolnego ich usunięcia przez autorów. Opublikowany dokument staje się niejako własnością RPK, podobnie jak dokumenty opublikowane jako artykuły w czasopismach lub jako książki stają się własnością ich wydawców, przy czym RPK nie posiada praw majątkowych do zdeponowanych dokumentów i nie czerpie z ich publikowania i udostępniania żadnych korzyści finansowych.

Jakość zawartości informacyjnej RPK

Nawigację w obrębie serwisu RPK zapewnia stałe menu główne, powtarzające się na wszystkich stronach serwisu, które ułatwia odnalezienie i zrozumienie informacji umieszczonych na stronach oraz płynne poruszanie się w serwisie w celu odszukania żądanych informacji o repozytorium, serwisie i dokumentach.

Wygląd strony głównej serwisu zbliżony jest do wyglądu strony Google (jedno okienko), w które użytkownik wpisuje zapytanie. Jest to zabieg celowy – sugerujący, że podobnie jak wyszukiwarka najwyższej punktowana w rankingach – również RPK zapewnia użytkownikowi szybki i łatwy sposób wyszukania właściwej dla niego informacji („od jednego kliknięcia”).

Stałymi elementami serwisu, powtarzającymi się na wszystkich stronach, które może przeglądać użytkownik, są również loga instytucji, dzięki którym repozytorium powstało i funkcjonuje – Politechniki Krakowskiej, Biblioteki Politechniki Krakowskiej, Unii Europejskiej, Programu Operacyjnego Innowacyjna Gospodarka. Loga mają nie tylko znaczenie informacyjne; dają użytkownikowi przede wszystkim gwarancję jakości i wiarygodności serwisu, świadczą o jego renomie.

Zwyczajowo w stopce serwisu umieszczono informację o instytucji, której przysługują prawa autorskie do serwisu (BPK) oraz przedział czasowy, w którym był on tworzony (2009–2012). Informacja ta daje użytkownikowi pewność, że nie trafił na stronę przestarzałą, lecz aktualizowaną, a zatem spełnia jedno z kryteriów wiarygodnego źródła informacji – aktualność (weryfikacja aktualności serwisu możliwa jest również dzięki przeglądaniu zasobów według kryterium „ostatnio opublikowane”).

Menu główne skonstruowano w sposób, który odzwierciedla porządek i logikę serwisu. W osobnych blokach zamieszczono: 1) informacje ogólne dotyczące projektu („Informacje o RPK”), 2) instrukcje dla potencjalnych autorów zamierzających deponować zasoby („Dla autorów”), 3) sekcje dla użytkowników zainteresowanych zasobami repozytorium i ewentualnie ich metadany („Przełóż” i „Szukaj”), 4) sekcję „Pomocy”, 5) strefę dla użytkowników uprawnionych do logowania – użytkowników Biblioteki PK – umożliwiającą

gromadzenie zasobów własnych i ulubionych („Zaloguj”). Ze względu na jakość informacyjną, sprzyjającą budowaniu wiarygodności serwisu, renomy i społecznego zaufania, znaczenia nabierają z pewnością pkt. 1), 2) i 4).

W zakładce „Informacje o RPK” użytkownik znajdzie dane ogólne o repozytorium, regulamin, informacje o osobach odpowiedzialnych za repozytorium w BPK oraz na poszczególnych wydziałach Uczelni, wreszcie odesłanie do podserwisu projektu głównego SUW.

Zakładka „Dla autorów” oferuje osobom zainteresowanym publikowaniem własnych zasobów w RPK, instrukcję ich deponowania, a także ogólne informacje o licencjach nadawanych dokumentom w repozytorium (wraz z odesłaniem do stron organizacji Creative Commons Polska).

Dla użytkownika korzystającego z RPK po raz pierwszy, zgrupowano przydatne informacje w zakładce „Pomoc”. FAQ, najczęściej zadawane przez użytkowników pytania (1), krótka prezentacja (TurboDemo Tutorial), przeprowadzająca użytkownika przez proces logowania się na platformę (2) oraz film przedstawiający możliwości RPK (3) prezentują w miarę kompleksowo serwis oraz zapewniają, przede wszystkim nowym użytkownikom, wsparcie w zakresie podejmowania samodzielnych prób korzystania z jego możliwości.

Polityka prezentowania informacji o repozytorium i o zasobach wydaje się być oparta na zasadzie „minimum formy na rzecz eksponowania treści/informacji”, stąd konsekwentnie w całym serwisie zrezygnowano z elementów graficznych (z wyłączeniem logo). Zastosowano białe tło stron, które „odbija” zawarte na nich informacje, ułatwiając użytkownikowi skupienie uwagi na istotnych elementach. Pomarańczowy kolor linków w serwisie (stosowane do opisu dokumentu, opisu kolekcji, linków zewnętrznych) stanowi poprawną estetykę, będąc jednocześnie świadomym nawiązaniem do idei OA, a przy okazji skromnym akcentem, że serwis jest budowany na jego zasadach. Bardziej świadomi użytkownicy mogą uznać wymienione elementy za przejaw staranności, pewnej fachowości.

Scharakteryzowane ogólnie jakość zawartości informacyjnej oraz jakość użytkowa RPK przyczynią się, mamy nadzieję, do zachęcenia użytkowników do korzystania z możliwości, jakie daje serwis w zakresie publikowania i trwałego archiwizowania własnych dokumentów, jak i przeglądania bądź wyszukiwania zgromadzonych już zasobów. Aktywne, częste wykorzystywanie platformy być może zachęci dodatkowo użytkowników do kolejnego kroku – wykorzystania jej jako narzędzia do komunikacji naukowej. Możliwość komentowania, recenzowania, opiniowania, rekomendowania zdeponowanych w RPK zasobów, a taką gwarantuje projekt, nada mu nową jakość i nowy sens informacyjny, jako że użytkownicy będą kształtowali jego wygląd i zasoby¹⁶.

¹⁶ M. Górski i in., Znaczenie Zintegrowanego Systemu Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych (SUW) w organizacji komunikacji naukowej

Promocja projektu SUW i RPK

Od początku realizowania projektu SUW i tworzenia RPK zespół ds. szkoleń i promocji organizował spotkania informacyjne i szkolenia dedykowane społeczności PK. W 2009 r. zrealizowano szkolenia z zakresu idei OA wraz z przykładami światowych i europejskich rozwiązań – czasopisma OA, instytucjonalne i dziedzinowe repozytoria, kursy OA prestiżowych uczelni. Zespół ds. merytorycznych i zespół ds. szkoleń i promocji przygotowały serwis internetowy poświęcony OA, dostępny ze strony głównej BPK (aktualizowany na bieżąco)¹⁷. Prezentacje na temat OA oraz realizowanego na Uczelni projektu SUW przedstawiono pracownikom BPK, członkom Rady Bibliotecznej, uczestnikom Rad Wydziałowych oraz redaktorom-konsultantom RPK. Od października 2011 r. do czerwca 2012 r. dla ww. grup, dla przedstawicieli jednostek międzyuczelnianych, a także dla przedstawicieli studentów poszczególnych wydziałów przeprowadzono szereg szkoleń poświęconych funkcjom i zasadom organizacji uczelnianego repozytorium oraz praktycznym aspektem jego wykorzystania.

Promocja projektu SUW i RPK odbywała się również w trakcie imprez organizowanych przez BPK w ramach Międzynarodowych Dni OA oraz Ogólnopolskiego Tygodnia Bibliotek (od 2010 r.). Bibliotekarze zaangażowani w prace projektowe SUW i RPK wystąpili na kilku konferencjach, referując projekt i poszczególne etapy jego realizacji. Jednym z elementów kampanii promocyjnej projektu SUW jest strona internetowa, na której znajdują się informacje poświęcone jego idei i realizacji w poszczególnych latach (2009–2012) oraz kalendarium.

Podsumowanie

Dyskusja na temat modelu otwartej nauki oraz repozytoriów uczelnianych wydaje się być w Polsce w początkowej fazie rozwoju. Projekty tego typu powstały w kilku polskich ośrodkach akademickich: Repozytorium Politechniki Krakowskiej¹⁸, Repozytorium Uniwersytetu im. Adama Mickiewicza w Pozna-

na Politechnice Krakowskiej, [w:] *Otwarte zasoby wiedzy – nowe zadania uczelni i bibliotek w rozwoju komunikacji naukowej* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://suw.biblos.pk.edu.pl/resources/i5/i7/i9/i4/r5794/GorskiM_JuszczykH_LipinskaD_PietrukowiczM_ZnaczenieZintegrowanego.pdf.

¹⁷ Open Access [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.biblos.pk.edu.pl/open_access.

¹⁸ Repozytorium Politechniki Krakowskiej [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://suw.biblos.pk.edu.pl>.

niu (AMUR)¹⁹, Repozytorium Ery Politechniki Wrocławskiej²⁰, Repozytorium Uniwersytetu Warszawskiego²¹.

Publikowanie dokumentów w repozytoriach daje wiele korzyści zarówno samym autorom, jak i uczelniom. Warto przytoczyć tu argumenty przemawiające za repozytorium, jakie przedstawiły realizatorki poznańskiego repozytorium AMUR – Małgorzaty Rychlik i Emilii Karwasińskiej: „Korzyści, które uzyskuje autor deponujący swoje prace w repozytorium, to przede wszystkim: 1) powszechność i szybkość dostępu do prac naukowych; 2) zwiększenie cytawalności dorobku naukowego (w zależności od dziedziny: od 36 proc. w biologii do 250 proc. w fizyce); 3) zabezpieczenie dokumentu (prace zdeponowane w repozytorium mają swój stały, indywidualny identyfikator); 4) repozytorium gwarantuje długotrwałe przechowywanie prac, w odróżnieniu od stron domowych wydziałów lub zakładów, które często zmieniają swoje adresy; 5) z prac zdeponowanych w repozytorium można generować osobiste bazy danych, które stanowią rejestr dorobku naukowego pracowników naukowych; 6) istnieje możliwość śledzenia statystyk dotyczących zdeponowanych prac, świadczących o zainteresowaniu artykułem. Uczelnia zyskuje następujące korzyści: 1) wzrost prestiżu uczelni poprzez wzrost widoczności prac naukowych; 2) promowanie kadry naukowej; 3) ułatwienie zadań administracyjnych np. w przypadku sprawozdawczości; 4) wspieranie edukacji studentów poprzez ułatwienie dostępu do prac naukowych”²².

Można prognozować, że w Polsce będą powstawać nowe projekty repozytoriów dziedzinowych i instytucjonalnych, nie tylko ze względu na korzyści, jakie one dają, ale również ze względu na zalecenia Komisji Europejskiej, sformułowane w oficjalnym komunikacie prasowym 17 lipca 2012 r. Według zaleceń Komisji, od 2014 r. wszystkie artykuły opracowywane dzięki dofinansowaniu ze środków „Horyzont 2020” muszą być publicznie udostępniane: „artykuły będą niezwłocznie udostępniane on-line przez wydawcę („złoty” dostęp) [...]; lub naukowcy będą udostępniać swoje artykuły w otwartym archiwum nie później niż sześć miesięcy (12 miesięcy dla artykułów z dziedziny nauk społecznych i humanistycznych) po publikacji („zielony” dostęp)”²³.

¹⁹AMUR. Adam Mickiewicz University Repository [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui>.

²⁰Repozytorium Ery Politechniki Wrocławskiej [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://zet10.ipee.pwr.wroc.pl>.

²¹Repozytorium Uniwersytetu Warszawskiego [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <https://depotuw.ceon.pl>.

²²M. Rychlik, E. Karwasińska, *Open Access po polsku* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.forumakad.pl/archiwum/2010/05/64_open_access_po_polsku.html.

²³Komisja Europejska, *Dane naukowe. Otwarty dostęp do wyników badań przyczyni się do zwiększenia potencjału innowacyjnego Europy* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/790&format=HTML&aged=0&language=PL&guiLanguage=en>.

Bibliografia

1. Akademia.edu [online], 2012 [dostęp: 24.09.2012], dostępny w World Wide Web: <http://akademia.edu>.
2. Bednarek-Michalska B., *Oceń jakość informacji elektronicznej!* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://biblioteka.oeiizk.waw.pl/konferencyjne/michalska.pdf>.
3. Open Access [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.biblos.pk.edu.pl/open_access.
4. Regulamin Repozytorium Politechniki Krakowskiej. Repozytorium Politechniki Krakowskiej [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://suw.biblos.pk.edu.pl/regulamin>.
5. Repozytorium Politechniki Krakowskiej [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://suw.biblos.pk.edu.pl>.
6. Zintegrowany System Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.biblos.pk.edu.pl/suw>.
7. BioInfoBank Library [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://lib.bioinfo.pl>.
8. Boruszewski J., *Jakość i wiarygodność informacji w infobrokerstwie* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://lingua.amu.edu.pl/Lingua_22/13_Jaroslav%20Boruszewski.pdf.
9. Cisek S., *Nauka 2.0. Nowe narzędzia komunikacji naukowej* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://informacjacyfrowa.wsb.edu.pl/pdfs/nauka%202.0.pdf>.
10. Peer Evaluation [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.peerevaluation.org>.
11. Komisja Europejska, *Dane naukowe. Otwarty dostęp do wyników badań przyczyni się do zwiększenia potencjału innowacyjnego Europy* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/790&format=HTML&aged=0&language=PL&guiLanguage=en>.
12. Górski M. i in., Znaczenie Zintegrowanego Systemu Wymiany Wiedzy i Udostępniania Akademickich Publikacji z Zakresu Nauk Technicznych (SUW) w organizacji komunikacji naukowej na Politechnice Krakowskiej, [w:] *Otwarte zasoby wiedzy – nowe zadania uczelni i bibliotek w rozwoju komunikacji naukowej* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://suw.biblos.pk.edu.pl/resources/i5/i7/i9/i4/r5794/GorskiM_JuszczykH_LipinskaD_PietrukowiczM_ZnaczenieZintegrowanego.pdf.
13. Internet2 [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.internet2.edu>.
14. Juszczyk H., Stąporek M., *The Technical Sciences Knowledge Exchange and Academic Publications Sharing Integrated System (ABBR, SUW) as an example of academic system to share knowledge and communicate between the university environment* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.bg.pw.edu.pl/iatul2011/proceedings/ft/Staporek_M.pdf.
15. Kulczycki E., *Wykorzystanie mediów społecznościowych przez akademickie uczelnie wyższe w Polsce. Badania w formule otwartego notatnika* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://amu.academia.edu/Emanuel>

- Kulczycki/Papers/1766754/Wykorzystanie_mediow_spoecznościowych_przez_akademickie_uczelnie_wyższe_w_Polsce._Badania_w_formule_otwartego_notatnika.
16. Nature Network [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://network.nature.com>.
 17. Repozytorium Eny Politechnika Wroclawska [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://zet10.ipee.pwr.wroc.pl>.
 18. ResearchGate [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://www.researchgate.net>.
 19. Rychlewski L., Dudek R., Rodakowska E., *BioInfoBank Library. Narzędzie dla naukowców i uczelni wyższych* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://lib.psnc.pl/Content/311/BC-16-Rychlewski.pdf>.
 20. Rychlik M., Karwasińska E., *Open Access po polsku* [online], [dostęp: 24.09.2012], dostępny w World Wide Web: http://www.forumakad.pl/archiwum/2010/05/64_open_access_po_polsku.html.
 21. Stankiewicz K., Wpływ Internetu na percepcję wiarygodności informacji, [w:] *Spoleczeństwo informacyjne. Wizja czy rzeczywistość?*, II ogólnopolska konferencja naukowa: Kraków 30 maja 2003, t. II, s. 407–415 [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <http://winntbg.bg.agh.edu.pl/skrypty2/0096/407-415.pdf>.
 22. AMUR. Adam Mickiewicz University Repository [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <https://repozytorium.amu.edu.pl/jspui>.
 23. Repozytorium Uniwersytetu Warszawskiego [online], [dostęp: 24.09.2012], dostępny w World Wide Web: <https://depotuw.ceon.pl>.

Marek Szepski

Krakowska Akademia im. Andrzeja Frycza Modrzewskiego

Biblioteka naukowa w Internecie – od formy do funkcji

Słowa kluczowe: biblioteka, funkcje biblioteki, digitalizacja

Abstrakt: w artykule podjęto próbę odpowiedzi na pytanie, co łączy książkę stojącą na półce z reprezentującym ją plikiem komputerowym; a w konsekwencji na pytanie, co tak naprawdę przechowuje biblioteka. Zanim zaczniemy cyfryzować zbiory biblioteczne oraz nim zadamy pytanie, jak przechowywać cyfrowe dokumenty, najpierw trzeba określić, co, a przede wszystkim po co, przechowywać. W epoce komunikacji sieciowej odpowiedzi na proste pytania mogą okazać się zaskakujące. Zmiana tradycyjnego spojrzenia na bibliotekę, jako miejsca gromadzenia kopii dokumentów, spowoduje, że zobaczymy organizację operującą informacją. Pojawiają się nowe cele, zadania i możliwości jej funkcjonowania.

Wstęp

Na potrzeby niniejszego artykułu, dokonano krytycznej analizy funkcjonowania bibliotek cyfrowych. Niestety wniosek podstawowy był negatywny; biblioteki nie spełniały pokładanych w nich nadziei. Kontynuując temat, autor chce znaleźć odpowiedź na pytanie, co trzeba zmienić, aby biblioteki (cyfrowe) rozwijały się prężnie ku zadowoleniu czytelników i miłośników książek.

Oczekiwania

Na wpisane w wyszukiwarkę Google hasło „biblioteka cyfrowa” pojawia się długa lista bibliotek udostępniających swoje zbiory w sieci. Są to zarówno biblioteki dużych instytucji i szkół wyższych, ale także biblioteki o istnieniu których nie mieliśmy pojęcia. Biblioteki cyfrowe stają się więc standardem. Tworzenie cyfrowych kolekcji i udostępnianie ich w sieci jest traktowane jako specjalne przedsięwzięcie realizowane obok normalnej działalności bibliotecznej. Podkreśla się znaczenie cyfrowej formy zbiorów. Postanowiono zatem przeanalizować, jak forma cyfrowa zbiorów może wpływać na działalność biblioteki.

W Wikipedii można znaleźć określenie biblioteki: „Biblioteka (od greckiego *βιβλιοθήκη* [bibliothekē]; *βιβλίον* [biblion] – książka) – instytucja społeczna, która gromadzi, przechowuje i udostępnia materiały biblioteczne oraz informuje o materiałach bibliotecznych (swoich i obcych). W innym znaczeniu jest to też nazwa samego budynku, pomieszczenia lub mebla zawierającego zbiory biblioteczne. Materiały biblioteczne to zbiór książek oraz innych materiałów źródłowych. Klasyczne zbiory biblioteczne są mocno związane z nośnikami fizycznymi. Materiały dostępne w postaci cyfrowej są gromadzone i udostępniane w tzw. bibliotekach cyfrowych”¹.

Ostatnie zdanie wzbudza duże wątpliwości. Wynika z niego, że biblioteka cyfrowa nie jest biblioteką, ale osobnym bytem. Jednak, poza wyjątkowymi sytuacjami, są one tworzone i prowadzone w ramach istniejących instytucji bibliotecznych. Powołanie specjalnych działów zajmujących się digitalizacją nie tworzy jeszcze nowych bytów. Można zrozumieć, że w początkowym okresie całe oprzyrządowanie techniczne było nowością, czymś obcym i tajemniczym w pracy bibliotekarzy, ale obecnie jest to sprzęt powszechnego użytku. Wyróżnianie cyfrowego charakteru zbioru dokumentów jest niczym nieuzasadnione. Biblioteka i sposób jej działania są mocno związane z tradycyjnymi formami zbiorów. W czasach Internetu wypada na nowo zdefiniować zadania i funkcjonowanie bibliotek.

Cyfryzacja w bibliotece

O bibliotekach cyfrowych można przeczytać, że rozwijają się dynamicznie, są przyszłością, zwiększają dostęp do wiedzy i wiele podobnych optymistycznych opinii. Zwykle jednak formułowane są dwa główne zadania tego typu podmiotów.

Każda biblioteka działa w dwóch obszarach: archiwizacji i bezpiecznego przechowywania zbiorów oraz udostępniania zbiorów, które są potrzebne czytelnikom w ich aktualnej pracy. Postać cyfrowa dokumentów jest rozwiązaniem, które może ułatwić realizację zadań w obu ww. obszarach.

Jako pierwsze zadanie dla biblioteki cyfrowej wymienia się zabezpieczenie cennych wydawnictw i starodruków, które biblioteka posiada w swoich zbiorach. Ten obszar jest niewątpliwie ważny, ale nie jest zupełnie nowy. Podobną rolę spełniały inne formy tworzenia kopii dokumentów, choćby w postaci mikrofilmów. Ten obszar działania nie wzbudza wątpliwości, z większymi lub mniejszymi kłopotami udaje się znaleźć potrzebne na to środki.

¹ Biblioteka [online], [dostęp: 20.10.2012], dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Biblioteka>.

Jedyne zastrzeżenia można mieć do faktu, że zainteresowanie tymi zbiorami ograniczone jest do wąskiego grona specjalistów.

Drugim celem stawianym bibliotekom cyfrowym jest ułatwienie dostępu do posiadanych zasobów. Niestety dokonane analizy pokazują, że tego zadania biblioteki cyfrowe nie realizują w sposób zadowalający². Pierwotna przyczyna tego problemu leży w nieprecyzyjnym sformułowaniu zadania. Zwykle nie jest powiedziane, co i komu ma być udostępniane i co to znaczy, że dostęp ma być ułatwiony. Można na przykład zapytać, czy ułatwieniem są osobne katalogi zbiorów klasycznych i cyfrowych. Taka sytuacja jest niestety typowa.

Z niewiadomych powodów pomija się powszechny efekt cyfryzacji bibliotek, którym są katalogi komputerowe. Nawet te biblioteki, które nie cyfryzują zbiorów, są w cyfrowym świecie. Udostępnienie katalogu w Internecie jest najpowszechniejszą usługą. Wiele osób w poszukiwaniu bibliografii, przegląda katalogi. W naturalny sposób dotyczy to bibliotek narodowych gromadzących „z urzędu” wszystkie publikacje. Ogromne znaczenie ma tu obszerność i kompletność katalogu. Być może bibliotekarze nie zdają sobie w pełni sprawy ze znaczenia tej usługi.

Pliki cyfrowe – komunikaty nowej epoki

Powszechną sytuacją jest, że na półce stoją obok siebie dwie książki. Wyglądają bardzo różnie. Jedna jest starannie wydana, na dobrym papierze, wydrukowana czytelną czcionką, ładnie oprawiona. Druga jest trudna do czytania: papier pożółkły, druk ze starego powielacza, zbyt mała czcionka. Te dwie książki mają tego samego autora, ten sam tytuł i taką samą treść. Pytanie – co łączy te książki i czym się różnią, w sytuacji powszechnej cyfryzacji nabiera nowego znaczenia. Prowadzi wprost do postawienia pytania o to, co tak naprawdę przechowuje biblioteka. Skoro wszystko będziemy teraz przechowywać w postaci plików cyfrowych, to co te pliki reprezentują. Nim włączymy skanery, nim zadamy pytanie, jak przechowywać, najpierw wypada zastanowić się, co przechowywać i dla kogo. Czy cyfryzacja to tylko cyfrowe zakodowanie obrazu książki stojącej na półce, czy może coś więcej? Czy powinniśmy cyfryzować obie książki stojące na półce, czy wystarczy tylko zrobić to z jedną? A gdy zeskanowaną książkę (przy okazji można trochę

² Zob. też.: A. Januszko-Szakiel, M. Szepski, *Internet – niewykorzystana szansa promocji polskiej nauki*, [w:] *Polska w mediach, media w Polsce*, red. nauk. Z. Pucek, J. Bierówka, Kraków 2012, s. 123–137; M. Szepski, *Polskie biblioteki cyfrowe z punktu widzenia internauty*, [w:] *Tradycja i nowoczesność w bibliotece naukowej XXI wieku*, red. nauk. A. Januszko-Szakiel, Kraków 2012, s. 233–247.

poprawić jej wygląd) wydrukujemy i postawimy na półce obok dwóch poprzednich, to powstaje pytanie; ile mamy książek, nadal dwie, czy już trzy. A może prawidłową odpowiedzią jest, że mamy jedną książkę. Bez odpowiedzi na pytanie, po co cyfryzować, podejmowane działania są nieskoordynowane i przypadkowe.

Pierwsza nasuwająca się odpowiedź, że dwie przywołane na początku książki mają tę samą treść, a różnią się wydaniem, czyli wyglądem, jest niewystarczająca i nie wnosi nic nowego. Postaram się poszukać lepszej odpowiedzi na zadane pytania.

Marshall McLuhan jest autorem teorii determinizmu technologicznego³. Głosi ona, że rozwój ludzkości odbywa się skokowo i związany jest ze zmianą sposobu komunikowania się ludzi. Takimi przełomowymi wynalazkami były: alfabet fonetyczny, maszyna drukarska i telegraf (jako początek urządzeń komunikacji elektronicznej). Żyjemy w czasach, w których doświadczamy kolejnego przełomu. Zmianę zapoczątkowało wynalezienie mikroprocesora. Obecną epokę można określić jako epokę informacji w cyfrowej sieci. Rewolucja informacyjna zmieniła funkcjonowanie świata, pogłębiła jego globalizację, ale także wywołała przeobrażenia społeczne w skali grup czy jednostek. Wobec tak doniosłych zmian trzeba na nowo określić zadania stojące przed bibliotekami, zdefiniować, co to jest książka i zasady dostępu do zgromadzonych zasobów.

W pewnych obszarach dyskusja na te tematy toczy się burzliwie. Dyskusja o prawach autorskich jest tu dobrym przykładem. Duże grupy użytkowników Internetu nie akceptują obecnie obowiązujących zasad ochrony i uważają, że zasoby internetowe można wykorzystywać bez opłat. Walczą z tym przedstawiciele firm (przede wszystkim fonograficznych, filmowych), które sporą część zysków czerpią z opłat za prawa autorskie. Innym obszarem zmian jest sprzeciw największych uniwersytetów w USA wobec płatnych czasopism naukowych i zapowiedź, że wszystkie prace naukowe ich pracowników będą publikowane przez uniwersytety bezpośrednio w sieci. Na to wszystko nakłada się ruch Open Access, głoszący, że osiągnięcia naukowe są dobrem wspólnym ludzkości i wszelkie ograniczenia w dostępie do wyników prac naukowych nie powinny mieć miejsca. Naruszone zostało też prawo prasowe. Łatwość publikowania w sieci spowodowała, że niemożliwe staje się zdefiniowanie, co to jest gazeta i czasopismo⁴.

³O teorii McLuhana można przeczytać w: Griffin E., *Podstawy komunikacji społecznej*, GPW 2003, gdzie teorii McLuhana poświęcony jest jeden z rozdziałów. Szersze omówienie zawiera: Levinson P., *Miękkie ostrze, czyli historia i przyszłość rewolucji informacyjnej*, Muza 2006.

⁴A. Adamski, *Mieszane uczucia. O proponowanej nowelizacji ustawy Prawo prasowe w odniesieniu do Internetu*, [w:] *Polska w mediach, media w Polsce, op. cit.*, s. 139–153.

Plik cyfrowy – obraz czy tekst

Gdyby zadać pytanie, co gromadzi biblioteka, zapewne w odpowiedzi otrzymalibyśmy, że: książki, czasopisma, druki, zdjęcia, mapy, niektórzy dodaliby, że mikrofilmy, nagrania audio, filmy. Pojawiłyby się także płyty CD i DVD, kasety VHS i zapewne wiele innych podobnych odpowiedzi. Ale są to odpowiedzi powierzchowne. Wymienione określenia to jedynie formy utrwalenia i są istotne tylko w aspekcie technicznym. Jeśli weźmiemy dowolną książkę, to znaczenie ma jej treść, a fakt, że jest wydrukowana na papierze i oprawiona, jest drugorzędny.

Po pojawieniu się e-czytników widać już wyraźnie, że treść jest oddzielona od formy jej prezentacji. Ta sama treść prezentowana i sprzedawana bywa jednocześnie jako wydrukowany tom, plik cyfrowy (np. w formacie pdf), aplikacja dla e-czytnika, nagranie audio lub mp3. Cyfryzacja powoduje, że utożsamianie formy przechowywania z zawartością merytoryczną nie może dalej wyznaczać zakresu działania bibliotek. Jeżeli więc biblioteka nie przechowuje książek, to powstaje pytanie, co w takim razie znajduje się w bibliotece. Wydaje się, że w dobie cyfryzacji odpowiedź narzuca się sama: biblioteki przechowują (a także opracowują i udostępniają) informację. Niektóre informacje mają charakter źródłowy, jak np. treść książki, inne powstają po przetworzeniu innych informacji czego przykładem może być katalog. Coraz więcej informacji nie ma swojej materialnej wersji (dokumentu) i występuje jedynie jako czysta informacja. Taką sytuację mamy np. jeśli biblioteka ma dostęp internetowy do publikacji zgromadzonych w tematycznych bazach danych.

Biblioteka internetowa – forma odchodzi w niepamięć

Konsekwencje takiej definicji biblioteki są natychmiastowe. Traci sens przymiotnik „cyfrowa” dodany do „biblioteki”. Biblioteka jest po prostu biblioteką bez dodatkowych przymiotników, a to, że przechowuje część informacji w postaci cyfrowej, co wymaga urządzeń do jej zobrazowania, jest taką samą sytuacją, jak to, że trzeba mieć specjalistyczne urządzenia do odtworzenia informacji zapisanej na płytach, filmach itp. Jeżeli już chceć dodać jakiś przymiotnik, to mogłoby to być: internetowa. Biblioteka internetowa może różnić się od biblioteki *tradycyjnej* tym, że nie ma budynku, korzystanie z jej usług nie jest związane geograficznie z konkretnym miejscem. Klasyczna biblioteka ma dwa kanały dostępu do informacji: skorzystanie z czytelni i wypożyczenie. Obecnie dochodzi kolejny kanał, jakim jest Internet. W sieci

biblioteka udostępnia swoje zbiory i przede wszystkim katalog zbiorów. Często zapomina się, że możliwość przeglądania katalogu jest ważną usługą świadczoną przez bibliotekę. Aby było to efektywne, biblioteka musi mieć jeden katalog obejmujący zarówno zbiory cyfrowe, jak i tradycyjne.

Wspomniana definicja biblioteki koncentruje się na książkach. Spojrzenie na bibliotekę jako na miejsce gromadzenia informacji wywołuje natychmiastowe skojarzenie z technologiami informacyjnym. Mamy więc cyfrową postać informacji, serwery, Internet, całą infrastrukturę teleinformatyczną i w konsekwencji informatyków, którzy stają się istotnymi współpracownikami. Informatycy, tworząc swoje aplikacje, patrzą na świat, jak na pewną abstrakcję. Biblioteka jest traktowana jako tajemniczy obiekt. Istotna jest odpowiedź na kilka pytań. Po pierwsze, po co jest ten obiekt, czyli co ma robić biblioteka. Oczywiście, każdy wie, co jest zadaniem biblioteki, problem stanowią rozbieżne oczekiwania różnych osób względem tej instytucji. Konieczne jest więc ustalenie jednoznacznej, wspólnej odpowiedzi.

Kolejne pytanie dotyczy zakresu działania obiektu. Jeżeli już ustalimy, że biblioteka gromadzi np. książki, to trzeba określić, z jakiego zakresu (zwykle nie wszystkie). Z zakresem związana jest bowiem odpowiedzialność zawarta w idei: „jeśli coś robimy, to robimy to dobrze”. Można ten problem zobrazować pytaniem: jeśli w bazie danych nie ma jakiejś książki, to jaką mam pewność, że taka książka nie istnieje. Mówiąc dokładniej, jeśli zapytam o publikację jakiegoś autora, to wolno mi przyjąć (choć nigdy nie mam pewności), że są to wszystkie publikacje.

Na koniec trzeba jeszcze określić, z jakimi innymi obiektami ta biblioteka jest powiązana, czyli skąd trafiają do niej informacje i komu są dostarczane.

Informacja to o wiele więcej niż książka. Rysunek 1 ilustruje, jak wiele i jak różnorodnych informacji związanych jest z typowym dokumentem. W bibliotece, oprócz tradycyjnie gromadzonych zbiorów, czyli opublikowanych książek i czasopism, mogą znajdować się także niepublikowane wykłady, raporty, katalogi itp. Także treść dokumentu może być rozbudowana o różne elementy w taki sposób, że tekst podstawowy jest tylko jednym z nich. Dodatkowo z dokumentem mogą być związane materiały uzupełniające, zwiększające możliwości wykorzystania dokumentu. Mimo że jest to jeszcze rzadkość, to takie podejście bywa już stosowane. Autorzy umieszczają w sieci aktualizacje danych statystycznych, grafiki w dużej rozdzielczości, prezentacje do wykładów prowadzonych na podstawie publikacji, a nawet cały tekst. Dobre przykłady można znaleźć na stronach internetowych wydawnictw, czego ilustracją może być strona wydawnictwa Helion⁵.

⁵ Strona domowa wydawnictwa Helion. Księgarnia internetowa Helion.pl [online], [dostęp: 20.10.2012], dostępny w World Wide Web: <http://helion.pl>.

Typy dokumentu	Dokument	Treść	Materiały uzupełniające
<ul style="list-style-type: none"> - książka / skrypt - artykuł - dokument urzędowy - raport - wykład (publiczny) - referat (konferencja) - wykład (studenci) - dokumentacja - patent - norma - katalog produktów 	<ul style="list-style-type: none"> - kto umieścił - autorzy - tytuł - data publikacji - data dodania do bazy - miejsce druku – nota bibliograficzna - czy recenzja - słowa kluczowe - typ dokumentu 	<ul style="list-style-type: none"> - streszczenie - spis treści - tekst - biblioteka - autorskie materiały uzupełniające - komentarze podpisane - komentarze anonimowe - link do strony autora - recenzje 	<ul style="list-style-type: none"> - komentarze autorskie - rozwiązania zadań - prezentacja ppt - programy komputerowe - zdjęcia duże - dźwięk - animacje i filmy

Rysunek 1. Struktura informacji związanych z dokumentem tekstowym

Źródło: opracowanie własne.

Odpowiedzialność za informację

Jedną z kluczowych cech informacji – zwłaszcza w czasach Internetu – jest jej wiarygodność. Najczęściej kojarzy się to z wiarygodnością źródła informacji. Informatycy używają w tym kontekście tajemniczego określenia – integralność danych. Wiarygodności nie należy mylić z prawdziwością. Naukowcy zwykle spierają się, prezentując różne sprzeczne i wykluczające się poglądy. Oznacza to, że przynajmniej jedno ze stanowisk (a czasami wiele różnych koncepcji) nie jest prawdziwe.

Wiarygodność w przypadku biblioteki oznacza raczej, że informacja jest kompletna, nie zawiera sprzeczności i „białych plam”. Jako przykład rozważmy bibliotekę w Krakowskiej Akademii (KA). Zakres tego, co ma gromadzić biblioteka, może być określony w różny sposób. Można np. przyjąć kryterium tematycznie, skoro Krakowska Akademia ma wydział prawa, to projekt funkcjonowania biblioteki może wymagać aby była w niej gromadzona, informacja o wszystkich publikacjach dotyczących prawa. Podkreślimy, informacja o dokumentach, a nie koniecznie same dokumenty. Można zdefiniować zakres także inaczej. KA kształci studentów na wydziale prawa, w bibliotece powinny być wszystkie materiały potrzebne studentom w trakcie studiów, gromadzone duże liczby egzemplarzy podręczników, a preferowane jest

udostępnianie treści w Internecie. Innym kryterium może być określenie terytorialne, co w tym przypadku oznacza zbiór wszystkiego, co jest związane z KA, w szczególności wszystko to, co wytworzyli pracownicy KA, nie tylko wszystkie publikacje naukowe, materiały do zajęć dydaktycznych, ale także dokumenty o charakterze administracyjnym, jak np. uchwały senatu, a także wszystko, co o KA napisali inni lub co może dotyczyć KA, a więc publikacje o szkołach wyższych w Polsce i na świecie.

Brak zdefiniowanego zakresu zbiorów prowadzi do tego, że zawartość biblioteki jest przypadkowa, a w efekcie, mało który z użytkowników jest zadowolony.

Informacja kojarzy się z czymś niematerialnym. Nasuwa się więc pytanie, czy aby udostępnić jakąś informację, trzeba mieć jakiś jej tradycyjny nośnik. Konkretnie, czy aby umieścić w katalogu informacje o jakiejś właśnie wydanej książce, trzeba ją wcześniej kupić. Oczywiście nie! Prawo autorskie ogranicza możliwość korzystania z treści dokumentu, ale dokument to nie tylko treść, ale także różne informacje z nim związane. Jak wiele informacji jest dostępnych, widać na stronach księgarń internetowych.

Zbiory analogowe mają pewną bardzo istotną cechę, nie dają się same z siebie przetwarzać. Efektem tego jest wprowadzenie metadanych, które są przetworzeniem wybranych informacji zawartych w dokumencie. Informacja cyfrowa nie ma tego ograniczenia, a konsekwencje najlepiej widać w standardach obowiązujących w Internecie. Z uwagi na to, że metadane są integralną częścią treści dokumentu, podstawowym sposobem poszukiwania informacji jest wyszukiwanie pełnotekstowe. Takie wyszukiwanie nie wymaga znajomości zasad klasyfikacji dokumentów i rozwiązuje problemy wynikające z ograniczeń, które w sposób naturalny pojawiają się w każdym systemie zamkniętym. Inną konsekwencją cyfryzacji informacji są powiązania między dokumentami. Naturalne wydaje się więc, że katalog biblioteki (do którego mamy dostęp przez Internet) będzie zawierał linki do treści dokumentu, nawet gdy biblioteka nie dysponuje tym dokumentem. Tu otwiera się wielkie pole do działania. Trzeba skatalogować i zabezpieczyć dokumenty, które powstały i zostały upublicznione wyłącznie w postaci cyfrowej. Z takiego punktu widzenia odpowiedź na często zadawane pytanie – co skanować – traci sens. Nie trzeba skanować. Aby powstała dobra biblioteka cyfrowa, skaner nie jest potrzebny. Paradoks: skanujemy to, co wydrukowane, a pomijamy dokumenty, które pierwotnie miały postać cyfrową, poprzedzając ich druk.

Niestety wszyscy łatwo zauważą, że aby taka biblioteka była naprawdę dobra, trzeba w domenie publicznej, czyli w Internecie, publikować więcej wartościowych tekstów. Należy domagać się, aby to, co zostało napisane w wyniku otrzymania pieniędzy publicznych, też stawało się publiczne. Działania rządu mogą szybko zmienić obecne przyzwyczajenia. Wystarczy

formalnie docenić publikacje internetowe, co jest jedynie zmianą obecnej umowy urzędniczego oceniania.

Duży wpływ na niedostatek bieżących publikacji ma działalność firm, które bronią tradycyjnie rozumianych praw autorskich. Powszechny jest pogląd, że skanować i umieścić w domenie publicznej można jedynie prace, których autorzy zmarli ponad 70 lat temu. Jest to podwójnie nieprawdziwe.

Po pierwsze, skanując, możemy naruszyć prawa tłumacza, redaktora, grafika i wydawcy. Oznacza to praktycznie, że chcąc umieścić taki tekst w sieci, trzeba go na nowo w sieci wydać.

Podkreślmy, że skanowanie i udostępnianie zeskanowanego obrazu w sieci jest zwykle kiepskim rozwiązaniem. Może mieć sens jedynie wtedy, gdy cenny jest wygląd dokumentu. Jeśli zależy nam przede wszystkim na treści, to należałoby uzyskać plik tekstowy, który można edytować i opublikować go w sieci z uwzględnieniem zasad typografii internetowej. Dokumenty wydrukowane są czcionką, która źle wygląda na ekranie monitora, są za szerokie aby je wygodnie czytać i przeważnie w formacie, który nie zgadza się ani z ekranem, ani z typowym rozmiarem papieru drukarek komputerowych.

Po drugie, ustawa o prawie autorskim i prawach pokrewnych⁶ pozwala bibliotekom naukowym na bardzo wiele. Precyzują to artykuły 27 i 28 ustawy.

Artykuł 27 ustawy mówi wprost: „Instytucje naukowe i oświatowe mogą, w celach dydaktycznych lub prowadzenia własnych badań, korzystać z rozpowszechnionych utworów w oryginale i w tłumaczeniu oraz sporządzać w tym celu egzemplarze fragmentów rozpowszechnianego utworu”.

Artykuł 28 w punkcie 3 pozwala dokonać wprost przeniesienia do sieci. „Biblioteki, archiwa i szkoły mogą: [...] udostępniać zbiory dla celów badawczych lub poznawczych za pośrednictwem końcówek systemu informatycznego (terminali) znajdujących się na terenie tych jednostek”.

Oznacza to, że na potrzeby dydaktyki lub prac badawczych możemy umieścić w sieci, ogólnodostępnej lub ograniczonej do uczelni, potrzebne fragmenty lub nawet całe dokumenty wydane współcześnie.

Bez wątpienia biblioteki zaczęły korzystać z możliwości, jakie daje Internet, można więc oczekiwać, że doświadczenia wyniesione z funkcjonowania sieci zostaną przeniesione do bibliotek. Stąd można spodziewać się realizacji koncepcji Web 2.0 w kontekście funkcjonowania bibliotek. Grono osób odpowiedzialnych za umieszczanie pozycji w bibliotece zostanie znacznie powiększone, a rola bibliotekarzy ulegnie zmianie i będą oni odpowiedzialni za organizację tej pracy.

Cele postawione przed bibliotekami ulegną poszerzeniu o promocję właścicieli biblioteki i nowe usługi dostarczania informacji. Być może koncepcje

⁶ Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst ujednolicony: Dz. U. 1994 Nr 24 poz. 83).

inteligentnej sieci rozumiejącej przetwarzaną informację znajdą zastosowanie właśnie w sytuacji wiarygodnej informacji źródłowej, a taka ma szansę znaleźć się w bibliotekach.

Podsumowanie

Podsumowując powyższe rozważania, można sformułować najważniejsze wnioski.

1. Przestańmy traktować biblioteki cyfrowe jako coś nowego, tajemniczego i specjalnego. Biblioteka będzie przechowywać dokumenty przede wszystkim w postaci cyfrowej i traktować je tak samo jak wszystkie inne przechowywane w różnych postaciach.
2. Zadaniem biblioteki jest przetwarzanie informacji i zapewnienie jej wiarygodności.
3. Schowajmy skanery do szafy. Biblioteki mają koncentrować się na współczesnych dokumentach cyfrowych.
4. Można spojrzeć na nowo na prawa autorskie. To, co wolno zgodnie z prawem umieścić w sieci, jest o wiele obszerniejsze niż teksty sprzed 70 lat.
5. Trzeba budować świadomość nowych zachowań i modeli biznesowych, których efektem ma być dążenie przez autorów do publikowania ich dorobku w sieci.
6. Biblioteki będą działać zgodnie z ideą sieci Web 2.0 i szeroka społeczność będzie budować zasób biblioteki.

Można zatem w skrócie powiedzieć, że biblioteki cyfrowe przechowują cyfrowe kopie zgromadzonych dokumentów, a biblioteka działająca w Internecie będzie przetwarzać informacje.

Biblioteki (cyfrowe) koncentrują się na formie posiadanych zbiorów. Taki punkt widzenia mocno ogranicza. Plik cyfrowy w naturalny sposób jest częścią Internetu. Stwarza to nowe, być może jeszcze nieświadomione, możliwości działania. Biblioteki przenoszą się do sieci. Jest to rzeczywisty efekt cyfryzacji. Stąd, dla dalszego rozwoju tak ważne jest wykorzystanie doświadczeń wynikających z analizy zachowań użytkowników Internetu. Po pierwsze, należy wykorzystać dobre praktyki, takie jak wyszukiwanie pełnotekstowe, hipertekst, widoczność zasobów w wyszukiwarkach (np. Google). Następnie biblioteki przejmą funkcje społecznościowe, zgodnie z koncepcjami Web 2.0 i staną się miejscem łączącym autorów, wydawców i czytelników.

Bibliografia

1. Adamski A., *Mieszane uczucia. O proponowanej nowelizacji ustawy Prawo prasowe w odniesieniu do Internetu*, [w:] *Polska w mediach, media w Polsce*, red. nauk. Z. Pucek, J. Bierówka, Oficyna Wydawnicza AFM, Kraków 2012, s. 139–153.
2. Biblioteka [online], [dostęp: 20.10.2012], dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Biblioteka>.
3. Golat R., *Aspekty prawne digitalizacji*, [w:] *Cyfrowy świat dokumentu – wydawnictwa, biblioteki, muzea, archiwa*, red. H. Hollender, Centrum Promocji Informatyki, Warszawa 2011, s. 64–87.
4. Januszko-Szakiel A., Szepski M., *Internet – niewykorzystana szansa promocji polskiej nauki*, [w:] *Polska w mediach, media w Polsce*, red. nauk. Z. Pucek, J. Bierówka, Oficyna Wydawnicza AFM, Kraków 2012, s. 123–137.
5. Strona domowa wydawnictwa Helion. Księgarnia internetowa Helion.pl [online], [dostęp: 20.10.2012], dostępny w World Wide Web: <http://helion.pl>.
6. Szepski M., *Polskie biblioteki cyfrowe z punktu widzenia internauty*, [w:] *Tradycja i nowoczesność w bibliotece naukowej XXI wieku*, red. nauk. A. Januszko-Szakiel, Oficyna Wydawnicza AFM, Kraków 2012, s. 233–247.
7. Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst ujednolicony: Dz. U. 1994 Nr 24 poz. 83).

Aneta Januszko-Szakiel

Biblioteka Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego

Narodowy program długoterminowej archiwizacji cyfrowego zasobu nauki i kultury – propozycja dla Polski

Słowa kluczowe: polskie zasoby cyfrowe, narodowe dziedzictwo cyfrowe, trwała ochrona zasobów cyfrowych, archiwizacja publikacji elektronicznych, narodowa strategia archiwizacji, archiwistyka cyfrowa, Ministerstwo Kultury i Dziedzictwa Narodowego, Ponańskie Centrum Superkomputerowo-Sieciowe

Abstrakt: zgodnie z Zaleceniami Komisji Europejskiej¹ oraz treściami zawartymi w raporcie opracowanym na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego, pn. *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*², 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 artykule została zaprezentowana autorska propozycja hipotetycznego scenariusza działań na rzecz długotrwałej archiwizacji polskich zasobów cyfrowych, z uwzględnieniem najistotniejszych światowych norm i standardów archiwistyki cyfrowej oraz warunków i możliwości polskich instytucji pamięci. W programie uwzględniono założenia Ministerstwa Kultury i Dziedzictwa Narodowego RP dotyczące powołania narodowych centrów kompetencji odpowiedzialnych za monitorowanie prac digitalizacyjnych i wdrażanie zmian technologicznych w zakresie m.in. przechowywania danych cyfrowych w długim czasie, a także budowy sieci bezpiecznych magazynów danych oraz repozytoriów cyfrowych, trwale przechowujących obiekty cyfrowe.

Wstęp

Zaproponowany w niniejszym artykule program dotyczy trwałej archiwizacji cyfrowych materiałów zgromadzonych w polskich instytu-

¹ Zalecenie Komisji z 27.10.2011 w sprawie digitalizacji i udostępniania w Internecie dorobku kulturowego oraz w sprawie ochrony zasobów cyfrowych [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recom28nov_all_versions/pl.pdf.

² *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020* [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf>.

cjach nauki i kultury³. Podstawowe typy instytucji, których zasoby tworzą narodowy zasób cyfrowy podlegający trwałej ochronie, to wydawnictwa, biblioteki, archiwa, muzea, instytuty audiowizualne, instytuty naukowe, badawczo-rozwojowe oraz instytucje kształcące.

Program odnosi się do różnych typów cyfrowych dokumentów; zdigitalizowanych oraz *born digital*, zapisanych na przenośnych mediach fizycznych typu: dyskietki, CD, DVD, oraz umieszczanych na serwerach i udostępnianych w Internecie.

W programie uwzględniono założenie MKiDN o powołaniu merytorycznych centrów kompetencji⁴. Centra kompetencji zostały utworzone przy następujących instytucjach:

- Biblioteka Narodowa,
- Narodowy Instytut Dziedzictwa⁵,
- Narodowy Instytut Audiowizualny,
- Narodowe Archiwum Cyfrowe,
- Narodowy Instytut Muzealnictwa i Ochrony Zbiorów.

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 w zakresie zarządzania procesami digitalizacji, ochrony i udostępniania następujących typów materiałów:

- Biblioteka Narodowa – materiały biblioteczne,
- Narodowy Instytut Dziedzictwa – zabytki,

³ Zaproponowana wersja programu została opracowana na podstawie autorskiego scenariusza programu trwałej ochrony cyfrowych zasobów polskich bibliotek, opublikowanego w „Przeglądzie Bibliotecznym” w 2011 r. Niniejsza propozycja jest ujęciem rozszerzonym, w której, obok cyfrowych zasobów bibliotecznych, uwzględniono cyfrowe zasoby archiwów, muzeów, instytucji gromadzących i przechowujących dokumenty audiowizualne, instytutów badawczych i rozwojowych oraz innych typów podmiotów, które w toku swej działalności tworzą dokumenty, będące świadectwem polskiej działalności w obszarze nauki i kultury, zasługujące na trwałą ochronę z myślą o obecnych i przyszłych użytkownikach.

⁴ A. Duńczyk-Szulc, *Polskie i europejskie programy digitalizacyjne na przykładzie WPR Kultura+ i EuropeanPhotography*, wystąpienie podczas konferencji: Dziedzictwo w sieci – różne aspekty digitalizacji, Kraków 29–30 listopada 2012.

⁵ Narodowy Instytut Dziedzictwa, wcześniej Krajowy Ośrodek Badań i Dokumentacji Zabytków. Zmiana nazwy nastąpiła z dniem 1 stycznia 2011 r. i ma związek z nadaniem tej instytucji nowego statutu. Podstawę prawną zmian stanowi Zarządzenie nr 32 Ministra Kultury i Dziedzictwa Narodowego z dnia 23 grudnia 2010 r.

- Narodowy Instytut Audiowizualny – materiały audiowizualne i audialne,
- Narodowe Archiwum Cyfrowe – materiały archiwalne,
- Narodowy Instytut Muzealnictwa i Ochrony Zbiorów – muzealia.

Proponowany program zawiera siedemnaście postulatów, przyporządkowanych do trzech grup:

Przyjęcie roli koordynatora działań archiwizacyjnych w Polsce

1. Ministerstwo Kultury i Dziedzictwa Narodowego RP przyjmuje rolę koordynatora działań na rzecz opracowania narodowej strategii trwałej ochrony polskiego dziedzictwa nauki i kultury w cyfrowej postaci

Do zadań własnych Ministerstwa Kultury i Dziedzictwa Narodowego, w szczególności Departamentu Mecenatu Państwa oraz Departamentu Dziedzictwa Kulturowego, należy koordynowanie zadań związanych z ochroną, zabezpieczaniem i udostępnianiem zbiorów kultury i nauki m.in. poprzez ich digitalizację.

W celu zabezpieczenia kompletnej kolekcji polskiego dziedzictwa kultury i nauki MKiDN powołuje centra kompetencji odpowiedzialne za: Biblioteka Narodowa – ochronę materiałów bibliotecznych, Narodowy Instytut Dziedzictwa – ochronę zabytków, Narodowy Instytut Audiowizualny – ochronę materiałów audiowizualnych i audialnych, Narodowe Archiwum Cyfrowe – ochronę materiałów archiwalnych, Narodowy Instytut Muzealnictwa i Ochrony Zbiorów – ochronę muzealiów.

Statuty instytucji powołanych przez MKiDN do pełnienia roli centrów kompetencji wskazują jednoznacznie, że instytucje te zostały powołane do realizacji szeroko pojętych zadań ochrony dziedzictwa kulturowego Polski w celu jego zachowania dla przyszłych pokoleń⁶. Pod nadzorem minister-

⁶ Statut BN [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.bn.org.pl/o-bn/statut-bn>; Statut Narodowego Instytutu Dziedzictwa [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.nid.pl/idm,13,statut.html>; Statut Narodowego Instytutu Audiowizualnego [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://bip.mkidn.gov.pl/media/download_gallery/index6ba4.pdf; O NAC. Zadania [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.nac.gov.pl/pl/zadania>; Statut Narodowego Instytutu Muzealnictwa i Ochrony Zbiorów [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://bip.mkidn.gov.pl/media/download_gallery/20121231p-88_Z_NIMOZ.pdf.

stwa, ze swym doświadczeniem i dorobkiem, z dużym prawdopodobieństwem instytucje te są w stanie zorganizować i poprowadzić archiwa, które spełnią kryteria wiarygodności oraz zapewnią długoterminową użyteczność polskiego dziedzictwa cyfrowego.

2. MKiDN ustala ekonomiczne podstawy działań na rzecz trwałej ochrony polskiego zasobu cyfrowego

MKiDN tworzy solidne, pochodzące z wiarygodnych źródeł podstawy długoterminowego finansowania podjętych działań archiwizacyjnych. Oprócz środków pochodzących z budżetu państwa, MKiDN oraz centra kompetencji szukają źródeł finansowania w innych instytucjach rządowych oraz pozarządowych, państwowych i prywatnych.

3. MKiDN formułuje cele i założenia podejmowanego przedsięwzięcia

W celu zapewnienia zgodności działań archiwizacyjnych z zasadami działalności archiwów wiarygodnych, MKiDN definiuje cel i założenia podjętych działań, a także zasady, którymi będzie kierować się, realizując nowe zadania.

Celem MKiDN jest zorganizowanie sieci wiarygodnych, długoterminowych archiwów dla 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.

MKiDN zakłada współodpowiedzialność i współdziałanie polskich instytucji kultury i nauki na rzecz opracowania strategii długoterminowej archiwizacji polskiego zasobu cyfrowego.

MKiDN 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 pamięci na świecie.

MKiDN 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; MKiDN przekazuje do publicznej wiadomości dokumentację działań oraz zakłada staranną analizę i uwzględnienie wszelkich opinii wewnętrznych (pracowników) oraz zewnętrznych (opinii publicznej),
- 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 MKiDN radzi sobie z realizacją wytyczonych celów.

MKiDN zakłada ponadto, 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. MKiDN publikuje informacje o podejmowanym przedsięwzięciu

W celu zapewnienia przejrzystości działań archiwizacyjnych MKiDN publikuje informacje o podjętym przedsięwzięciu. Wszystkie instytucje nauki i kultury w Polsce, zwłaszcza te, których zbiory tworzą narodowe dziedzictwo podlegające szczególnej ochronie, otrzymują pełną informację o przedsięwzięciu ministerstwa.

5. MKiDN gromadzi wiedzę na temat trwałej ochrony zasobów cyfrowych

Zgodnie z zaleceniami prezentowanymi w piśmiennictwie przedmiotu, MKiDN rozpoczyna wszelkie prace związane z planowaniem działań archiwizacyjnych od starannej analizy literatury 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. MKiDN zapoznaje się ze sposobem organizowania prac archiwizacyjnych w instytucjach nauki i kultury innych krajów, zaawansowanych i mogących pełnić rolę doradcy. MKiDN zapoznaje się też z działalnością i zaleceniami międzynarodowych organizacji skupionych wokół zadań długoterminowej archiwizacji.

Z różnych źródeł i doświadczeń MKiDN tworzy zasób wiedzy na temat organizowania działań archiwizacyjnych w Polsce.

6. MKiDN identyfikuje ewentualne rodzime projekty oraz plany programów ochrony dziedzictwa cyfrowego

Z uwagi na zalecenia dotyczące unikania dublowania prac i środków potrzebnych na działania archiwizacyjne, MKiDN stara się ustalić, czy w polskich instytucjach pamięci, nauki i kultury realizuje się bądź planuje realizację programów długoterminowej ochrony zasobów cyfrowych. W przypadku istnienia takich programów lub planów, MKiDN konfrontuje ich cele, założenia i zakres z własnymi zamierzeniami. MKiDN rozważa połączenie projektów i kooperację.

Organizacja pracy w zakresie trwałej ochrony polskich zasobów cyfrowych

7. MKiDN zwołuje stałą konferencję na temat długoterminowej archiwizacji polskich zasobów cyfrowych

MKiDN organizuje ogólnopolską konferencję na temat trwałej ochrony polskich zasobów cyfrowych oraz zaprasza do udziału przedstawicieli polskich instytucji wydawniczych, instytucji pamięci, nauki i kultury.

MKiDN deklaruje, że zwołana konferencja będzie mieć charakter spotkań organizowanych cyklicznie, najlepiej w ustalonych odstępach czasu i określonym miejscu.

Na pierwszej konferencji przedstawiciel MKiDN wygłasza referat wprowadzający w tematykę ochrony zasobów cyfrowych, podkreśla rangę problemu, podaje przykłady inicjatyw światowych oraz informuje o rozpoczętym przedsięwzięciu, przytaczając cele, założenia i zasady działania.

MKiDN zgłasza potrzebę ukonstytuowania ogólnopolskiej grupy roboczej ds. opracowania strategii ochrony polskich zasobów cyfrowych i zaprasza w jej szereg osoby, które chcą współpracować i posiadają wiedzę w zakresie organizacyjnych, technicznych, prawnych oraz ekonomicznych zagadnień tworzenia strategii archiwizacji zasobów cyfrowych. Celem grupy roboczej będzie rozpoczęcie i prowadzenie działań mających na celu utworzenie systemu depozytowego polskiego zasobu cyfrowego.

8. MKiDN powołuje Ogólnopolską Grupę Roboczą ds. trwałej ochrony polskich zasobów cyfrowych

MKiDN formuje i powołuje, spośród chętnych osób, Ogólnopolską Grupę Roboczą do zadań długoterminowej archiwizacji polskich zasobów cyfrowych, ustala strukturę organizacyjną Grupy z zespołem koordynującym na czele. W skład zespołu koordynującego wchodzi Biblioteka Narodowa, Narodowy Instytut Dziedzictwa, Narodowy Instytut Audiowizualny, Narodowe Archiwum Cyfrowe, Narodowy Instytut Muzealnictwa i Ochrony Zbiorów oraz Poznańskie Centrum Superkomputerowo-Sieciowe, jako instytucja od lat zaangażowana w procesy organizacji i zarządzania cyfrowymi zasobami polskich bibliotek, muzeów i archiwów, a także z uwagi na znajomość zagadnień trwałej ochrony zasobów cyfrowych oraz potrzeb instytucji pamięci w tym zakresie, wreszcie z racji swych kompetencji technicznych, które są niezbędne do realizacji zadań związanych z długoterminową ochroną archiwów cyfrowych i ich zasobów.

Integralną częścią zespołu koordynującego jest podzespół ds. obsługi merytorycznej, złożony z ekspertów, znawców organizacyjnych, technicznych, prawnych oraz ekonomicznych zagadnień długoterminowej archiwizacji zasobów cyfrowych. W skład tego podzespołu wchodzi 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.

Podzespół ds. obsługi merytorycznej identyfikuje metody prac, rozwiązania, standardy, normy, akty prawne itp. dotyczące archiwizacji zasobów elektronicznych, stosowane w instytucjach nauki i kultury na świecie. Opracowuje plany działania i poddaje je ocenie wewnętrznej przez zespół koordynujący oraz zewnętrznej przez opinię publiczną. Wszelkie sugestie uwzględnia, modyfikując proponowany plan działania.

Przy zespole koordynującym działa również podzespół 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 kontrolę oraz ewaluacją wszystkie procesy realizowane w ramach działalności archiwizacyjnej.

W strukturze organizacyjnej, obok zespołu koordynującego, MKiDN powołuje cztery następujące zespoły:

- Zespół ds. obsługi organizacyjnej działań archiwizacyjnych, składający się z bibliotekarzy, bibliotekoznawców, informatologów, archiwistów, muzealników, menedżerów kultury i oświaty, organizatorów kolekcji dokumentów audiowizualnych, przedstawicieli instytucji na-

ukowych i badawczych tworzących i zarządzających kolekcjami dokumentów cyfrowych,

- Zespół ds. obsługi technicznej działań archiwizacyjnych, składający się z przedstawicieli bibliotek, muzeów, archiwów, centrów informacji naukowej, instytutów audiowizualnych, instytutów naukowych, instytutów badawczo-rozwojowych, uczelni wyższych oraz informatyków (fachowców branży IT),
- Zespół ds. obsługi prawnej działań archiwizacyjnych, składający się z przedstawicieli bibliotek, muzeów, archiwów, centrów informacji naukowej, instytutów audiowizualnych, instytutów naukowych, instytutów badawczo-rozwojowych, uczelni wyższych oraz prawników,
- Zespół ds. obsługi ekonomicznej działań archiwizacyjnych, tworzony przez przedstawicieli bibliotek, muzeów, archiwów, centrów informacji naukowej, instytutów audiowizualnych, instytutów naukowych, instytutów badawczo-rozwojowych, uczelni wyższych i ekonomistów.

Wymienione zespoły mają charakter zadaniowy (wykonawczy). Ich zadania definiuje zespół koordynujący. Poszczególne zespoły zadaniowe mają liderów, którzy organizują i odpowiadają za terminową i właściwą pracę zespołów przed zespołem koordynującym. Liderzy na bieżąco dokumentują działania zespołów, a dokumentację przekazują do podzespołu ds. zarządzania jakością i dokumentacją.

Przewiduje się, w zależności od zadania, współpracę osób przynależących do różnych zespołów.

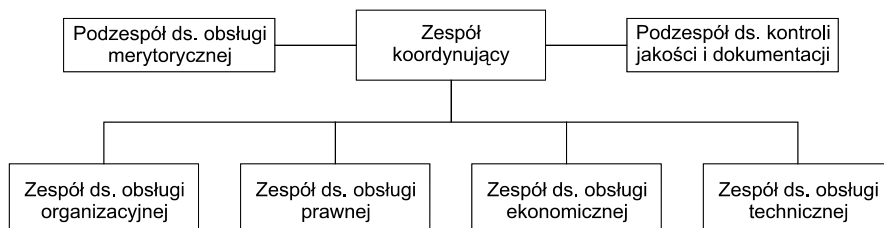
W skład poszczególnych zespołów wchodzi osoby z różnych polskich instytucji, głównie bibliotek, muzeów, archiwów, instytutów audiowizualnych, centrów informacji naukowej, instytutów naukowych, instytutów badawczo-rozwojowych, uczelni wyższych, firm informatycznych, kancelarii prawnych, instytucji sektora biznesu i finansów.

Praca w zespołach odbywa się w zależności od potrzeb, tradycyjnie bądź w trybie zdalnym.

Zespół koordynujący ustala sposób komunikowania się w obrębie poszczególnych zespołów oraz grupy roboczej.

Zespół koordynujący 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: opracowanie własne.

9. MKiDN definiuje ogólne cele Ogólnopolskiej Grupy Roboczej i poszczególnych zespołów

- Cel(e) Ogólnopolskiej Grupy Roboczej ds. trwałej archiwizacji polskich zasobów cyfrowych:
 - utworzenie długoterminowego, wiarygodnego i stabilnego systemu depozytowego dla polskiego zasobu cyfrowego,
 - zapewnienie długoterminowej użyteczności zasobów cyfrowych zgromadzonych w polskich instytucjach kultury i nauki, z myślą o obecnych i przyszłych użytkownikach,
 - wyeksponowanie elektronicznych zasobów polskiej nauki i kultury w cyfrowej kolekcji dziedzictwa światowego.
- Cel(e) zespołu koordynującego:
 - 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. MKiDN zakłada i prowadzi portal internetowy na temat trwałej ochrony polskich zasobów cyfrowych

MKiDN zakłada odrębny portal WWW bądź wydziela na swojej witrynie miejsce poświęcone sprawom długoterminowej archiwizacji polskich zasobów elektronicznych. Portal ten pełni funkcję wirtualnego informatorium o tym, co dzieje się w zakresie archiwizacji zasobów cyfrowych w Polsce; jest miejscem dyskusji, wymiany poglądów, publikowania informacji na temat rozwoju przedsięwzięcia.

Istnienie i sprawne funkcjonowanie takiego portalu jest świadectwem przejrzystości działań archiwizacyjnych, tym samym czynnikiem budującym wiarygodność systemu depozytowego polskich zasobów cyfrowych.

Elementem towarzyszącym serwisowi WWW jest elektroniczny biuletyn na temat archiwizacji, rozsyłany na zasadzie newslettera do zainteresowanych i współpracujących osób i instytucji, w celu bieżącego informowania np. o terminach spotkań oraz rezultatach działań.

MKiDN stara się uzyskać opinie i akceptację od przedstawicieli polskich instytucji bibliotecznych dla zarysowanej koncepcji rozpoczęcia 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.

Planowanie szczegółowych zadań Ogólnopolskiej Grupy Roboczej ds. trwałej ochrony polskich zasobów cyfrowych

11. MKiDN definiuje zadania szczegółowe zespołu koordynującego

Zespół koordynujący:

- wraz z podzespołem 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 cyfrowych zasobów polskich bibliotek,
- prezentuje działania rodzime na forach krajowym i międzynarodowym.

12. MKiDN definiuje zadania szczegółowe podzespołu ds. obsługi merytorycznej

Podzespół ds. obsługi merytorycznej:

- pełni funkcję merytorycznego zaplecza (źródła wiedzy) dla całości projektu,
- gromadzi wiedzę o działaniach archiwizacyjnych w instytucjach kultury i nauki innych krajów,
- identyfikuje i rejestruje istniejące rozwiązania, zalecenia, wzorce, normy, standardy itp. 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 polskiego zasobu cyfrowego.

13. MKiDN definiuje zadania szczegółowe podzespołu ds. kontroli jakości i dokumentacji

Podzespół ds. zarządzania jakością 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. MKiDN definiuje zadania szczegółowe zespołu ds. obsługi organizacyjnej

Zespół ds. obsługi organizacyjnej:

- organizuje badania dotyczące identyfikacji oraz stanu liczbowego i jakościowego materiałów cyfrowych przechowywanych w polskich instytucjach 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 itp.,
- 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 publikacji elektronicznych,

- wraz z zespołem ds. obsługi technicznej oraz prawnej opracowuje politykę tworzenia metadanych; definiuje format, zakres oraz poziom szczegółowości metadanych archiwizowanych publikacji,
- wraz z zespołem ds. obsługi technicznej ustala wytyczne dotyczące zgłaszania i przekazywania publikacji do instytucji archiwizujących,
- organizuje szkolenia z zakresu długoterminowej archiwizacji publikacji elektronicznych,
- 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 liczby 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 ze strony użytkowników.

W tym celu MKiDN przygotowuje rozporządzenie dotyczące przeprowadzenia badania oraz tworzy rejestr instytucji kultury i nauki 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 kolekcji dziedzictwa narodowego i na długoterminową ochronę.

Poszczególne instytucje na mocy zaleceń MKiDN powołują wewnętrzne zespoły ds. 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 rozporządzeniu dotyczącym badania określony zostaje 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 nauki i kultury, wytypować do określonych zabiegów konserwatorskich. Instytucje podejmują próbę odczytu i prezentacji treści dokumentów cyfrowych, aby oddzielić zasoby nieużyteczne (których treści nie można odczytać, z uwagi na brak odpowiedniego sprzętu i oprogramowania) od użytecznych (których treści można odczytać i zaprezentować, w formie zrozumiałej dla

użytkownika). Instytucje podejmują również próbę oceny wartości treści materiałów nieużytecznych i stopnia zapotrzebowania na nie zgłaszanego przez użytkowników. Instytucje ustalają, czy nieużyteczne dokumenty cyfrowe mają odpowiednik analogowy. W przypadku materiałów o szczególnej wartości i dużym zapotrzebowaniu, które nie mają substytutów analogowych, podejmowane są wszelkie możliwe działania pozwalające na odtworzenie ich treści.

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 technicznym zabiegom konserwatorskim 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.

W niniejszym programie przyjmuje się zasadę „oddzielenia” treści dokumentów cyfrowych od nośników fizycznych, na których są one zapisane, i umieszczenie ich w celu archiwizacji w repozytoriach. Do czasu ustalenia szczegółowych wytycznych dotyczących postępowania z dokumentami cyfrowymi są one przechowywane w repozytoriach (na dostępnych serwerach) poszczególnych instytucji i objęte zasadami ochrony i bezpieczeństwa przyjętymi dla systemów informatycznych.

W pierwszej, pilnej fazie działań archiwizacyjnych, poszczególne instytucje na swych serwerach umieszczają materiały najstarsze, o statusie dziedzictwa narodowego, niemające wersji analogowej. W kolejnej fazie działań archiwizacyjnych zapadają decyzje dotyczące materiałów użytecznych. Te, które z racji wartości ich treści powinny zostać poddane szczególnej ochronie, uzupełniają kolekcję archiwalną i trafiają do repozytorium, pozostałe natomiast pozostają na swych oryginalnych nośnikach i są poddawane stosownym zabiegom konserwatorskim.

MKiDN organizuje i zaprasza pracowników instytucji kultury i nauki do udziału w szkoleniach dotyczących metod archiwizacji, tak by przechowywane dokumenty, w zależności od potrzeby, zostały poddane operacjom odświeżenia nośnika, zmiany generacji nośnika, migracji bądź emulacji.

W programie zakłada się utrzymywanie użyteczności tak powstałych instytucjonalnych kolekcji cyfrowych do czasu, kiedy instytucja koordynująca działaniami archiwizacyjnymi oraz tworząca archiwa cyfrowe będzie gotowa przyjąć i długoterminowo archiwizować kompletne bądź fragmentaryczne kolekcje instytucjonalne.

W programie wyklucza się dublowanie zbiorów cyfrowych; do archiwum przyjęte zostaną te kolekcje instytucjonalne bądź ich fragmenty, które są w stanie uzupełnić narodową kolekcję cyfrową. Pozostałe zasoby cyfrowe

będą przechowywane w poszczególnych instytucjach i pozostaną tam do czasu opracowania i powołania procedur dotyczących tworzenia lokalnych archiwów cyfrowych.

Kolejnym, istotnym punktem proponowanego scenariusza jest współpraca instytucji koordynującej z wydawcami publikacji elektronicznych w zakresie długoterminowej archiwizacji.

W programie uwzględnia się fakt, że nie wszyscy wydawcy stosują się do zapisów ustawy o egzemplarzu obowiązkowym oraz że nie wszystkie typy publikacji są uwzględnione w treści ustawy o bibliotekach. Bierze się również pod uwagę, że publikacje elektroniczne, które stanowią dziedzictwo cyfrowe i powinny podlegać trwałej ochronie, są tworzone w różnych miejscach wydawniczych, np. instytutach badawczych i naukowych. Dlatego też instytucja koordynująca nawiązuje kontakty z wydawcami oraz miejscami wydawniczymi, a także innymi jednostkami organizacyjnymi i osobami fizycznymi niebędącymi wydawcami, ale prowadzącymi działalność polegającą na publikowaniu dokumentów⁷. Punktem wyjścia w tej kwestii jest tworzona przez BN bibliografia wydawnictw elektronicznych, która rejestruje m.in. polskich wydawców publikujących materiały cyfrowe. Dodatkowo tworzony jest rejestr miejsc wydawniczych, z którymi nawiązana zostanie współpraca.

Do wszystkich instytucji odnotowanych w bazie i rejestrze zostaje rozesłane powiadomienie o działaniach archiwizacyjnych w celu uzyskania akceptacji przedsięwzięcia oraz uzyskania ich pisemnej zgody dotyczącej współpracy.

W programie przyjęto założenie, że do czasu opracowania i wejścia w życie odpowiednich przepisów prawnych, udział instytucji wydawniczych w procesach archiwizacyjnych jest wynikiem ich dobrej woli i chęci przyłączenia się do przedsięwzięcia MKiDN oraz ścisłej współpracy z BN. Instytucje wydawnicze dobrowolnie zobowiązują się, na mocy porozumienia i umowy z instytucją archiwizującą, zgłosić i przekazać do jej systemu depozytowego każdy dokument opublikowany w wersji elektronicznej, w celu długoterminowej archiwizacji. Dodatkowo instytucje wydawnicze zobowiązują się stosować do wytycznych instytucji archiwizujących dotyczących standardów publikowania, tworzenia metadanych, zgłaszania oraz transferowania publikacji do archiwum.

Instytucja archiwizująca przyjmuje tym samym obowiązek archiwizacji zgłoszonych publikacji i zapewnienia ich trwałej użyteczności. Instytucja

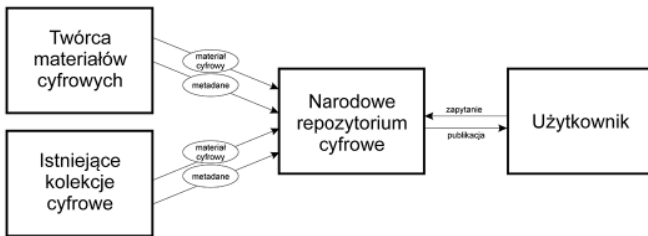
⁷ Pod pojęciem *miejsce wydawnicze* rozumie się w niniejszym programie instytucję bądź osobę, która nie ma statusu oficyny wydawniczej, ale publikuje dokumenty elektroniczne, które są wynikiem jej działalności. W przypadku proponowanego programu chodzi o współpracę np. z instytutami naukowymi, badawczymi, które publikują wyniki swych naukowych osiągnięć, a także z firmami organizującymi konferencje i publikującymi materiały konferencyjne, zarówno w wersji drukowanej, jak i cyfrowej. W programie miejsce wydawnicze będą również stanowić „twórcy”, czyli autorzy dokumentów elektronicznych, publikujący samodzielnie w Internecie, np. autorzy popularnych dzienników internetowych; w przypadku niektórych z nich, wartość dla nauki i edukacji jest niekwestionowana, dlatego warto je uwzględnić w programie i objąć trwałą ochroną.

archiwizująca zakłada również udostępnianie archiwizowanych publikacji, jednak z uwzględnieniem warunków dotyczących tego udostępniania, ustalonych i zapisanych w umowie z wydawcami i miejscami wydawniczymi.

Na podstawie doświadczeń krajów zaawansowanych w działaniach archiwizacyjnych wiadomo, że, w zależności od rodzaju dokumentu, instytucje wydawnicze mogą różnie ustalić termin karencji dla pierwszego udostępnienia publikacji przez archiwum. Jest to zabieg często stosowany w przypadku elektronicznych czasopism naukowych. Zanim wydawca czasopisma zezwoli na publiczny, bezpłatny dostęp do jego treści, jest ona dostępna przez określony czas, odpłatnie w serwisie wydawcy. Przez ten okres publikacja znajduje się albo w systemie depozytowym instytucji archiwizującej, ale tylko w celach archiwizacji, bez opcji udostępniania, albo tylko u wydawcy; wówczas instytucja archiwizująca otrzymuje zgłoszenie faktu opublikowania utworu oraz jego metadane. Określony zostaje także ostateczny termin odesłania przez wydawcę pełnego dokumentu do instytucji archiwizującej.

W programie przyjmuje się zatem, że wydawcy i miejsca wydawnicze decydują się w umowie na określony model współpracy w ramach działalności archiwizacyjnej.

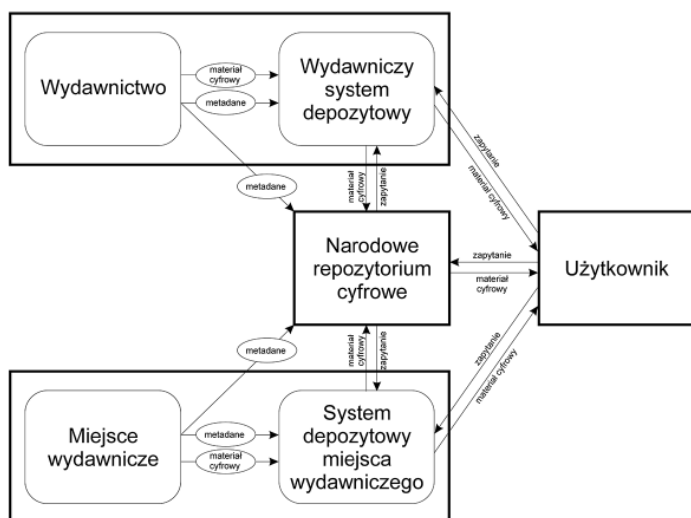
Pierwszy proponowany model współpracy zakłada, że publikacja elektroniczna zostaje zgłoszona oraz odesłana do instytucji archiwizującej w celu jej umieszczenia w systemie depozytowym. W metadanych użytkowych zostają określone warunki jej udostępniania. W modelu tym pełną odpowiedzialność za długoterminową ochronę opublikowanych dokumentów ponosi instytucja archiwizująca. Rola instytucji wydawniczych polega na stworzeniu i opublikowaniu dokumentów wraz z odpowiednimi metadanymi, a także ich zgłoszeniu i przekazaniu do instytucji archiwizującej, zgodnie z wytycznymi załączonymi do umowy (patrz schemat 2).



Schemat 2. Model współpracy instytucji wydawniczej z instytucją archiwizującą w zakresie długoterminowej archiwizacji publikacji elektronicznych w systemie depozytowym

Źródło: opracowanie własne.

W modelu drugim natomiast wydawcy samodzielnie realizują proces archiwizacji publikacji elektronicznych, czyli decydują się na tzw. model *self archiving*, jednak również w ścisłej współpracy z instytucją archiwizującą (patrz schemat 3).



Schemat 3. Model współpracy instytucji wydawniczej z instytucją archiwizującą w zakresie długoterminowej archiwizacji publikacji elektronicznych w systemie depozytowym instytucji wydawniczej, tzw. model *self archiving*

Źródło: opracowanie własne.

W drugim proponowanym podejściu wydawnictwa oraz wszelkie instytucje i organizacje tworzące i publikujące dokumenty elektroniczne ponoszą koszty związane z organizowaniem procesu długoterminowej archiwizacji oraz dostępu do archiwizowanych zasobów. Ponoszą również odpowiedzialność za utrzymanie użyteczności deponowanych zasobów cyfrowych w długim czasie.

W omawianym modelu do systemu depozytowego instytucji archiwizującej trafia jedynie zgłoszenie faktu opublikowania dokumentu oraz przekazanie jego metadanych. Udostępnianie dokumentu użytkownikom odbywa się bądź bezpośrednio z systemu depozytowego wydawcy, bądź za pośrednictwem instytucji archiwizującej.

W modelu *self archiving* istotna jest świadomość twórców dotycząca złożoności i odpowiedzialności związanej z procesem archiwizacji zasobów. Konieczne jest również ustalenie procedur, według których twórcy mieliby trwale chronić zdeponowane u siebie dokumenty cyfrowe.

W obu modelach bardzo istotne jest zapewnienie technicznej obsługi związanej z ochroną użyteczności i organizacją dostępu do archiwizowanych zasobów.

Na mocy przepisów o obowiązku wieczystej archiwizacji, instytucje pamięci zwykle dążą do ustawowego zobligowania wydawców do przekazywania wszystkich publikacji do depozytu, gdyż na tej podstawie mogą zgromadzić i przechować kompletny zasób. Jednak do czasu opracowania i obowiązywania odpowiednich przepisów prawnych, a także do czasu wdrożenia w instytucji archiwizującej profesjonalnego systemu depozytowego, model drugi może pełnić rozwiązanie przejściowe.

W programie przyjęto, że istotne znaczenie ma przygotowanie oferty szkoleń z zakresu tworzenia i archiwizacji dokumentów cyfrowych. Instytucja koordynująca 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 elektronicznych, z uwzględnieniem formatów dla różnych typów dokumentów,
- tworzenie metadanych dokumentów elektronicznych i formaty ich zapisu,
- systemy trwałego identyfikowania dokumentów elektronicznych,
- długoterminowa archiwizacja dokumentów elektronicznych – 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 koordynująca organizuje szkolenie i ustala miejsce jego przeprowadzenia. Szkolenia mogą odbywać się w siedzibie instytucji oferującej szkolenie bądź też zgłaszającej zapotrzebowanie.

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 corocznej 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 kierun-

kach rozwoju długoterminowej archiwizacji na podstawie działalności archiwizacyjnej instytucji pamięci z 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 dotyczących archiwizacji polskiego dziedzictwa cyfrowego.

Konferencja ma być okazją do nawiązania kontaktów z fachowcami, praktykami z zakresu archiwizacji, którzy mogą pełnić rolę instruktorów szkoleń dla kadr polskich instytucji pamięci odpowiedzialnych za zadania archiwizacji, ponadto mogą recenzować polskie plany i działania archiwizacyjne, a także doradzać instytucji koordynującej. Konferencja powinna również dostarczyć sposobność ustalenia, z którą instytucją można nawiązać współpracę odnośnie do przekazania jej pod ochronę kopii polskiej kolekcji archiwalnej.

15. MKiDN 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 materiałów cyfrowych oraz określa potrzebę i zakres ich uzupełnienia,
- konstituuje 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 aktów prawnych regulujących kwestie zgromadzenia w systemie depozytowym wszelkich dokumentów cyfrowych, także sieciowych; nowe ustawodawstwo ma też zobligować twórców do zgłaszania i odsyłania wszystkich dokumentów do depozytu oraz centra kompetencji do ich wieczystej archiwizacji,
 - pracuje nad nowelizacją ustawy o prawie autorskim i prawach pokrewnych⁸: celem noweli prawa ma być nadanie instytucji ar-

⁸ Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst ujednolicony: Dz.U. 1994 Nr 24 poz. 83).

- chiwizującej prawa do działań konserwatorskich na dokumentach 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 dokumentach cyfrowych z właścicielami praw (do czasu wejścia w życie noweli prawa regulującego tę kwestię),
- zapewnia obsługę prawną współpracy wydawców i miejsc wydawniczych z instytucją archiwizującą,
- wraz z zespołem ds. obsługi organizacyjnej oraz technicznej ustala procedury tworzenia metadanych użytkowych dla archiwizowanych obiektów.

Z uwagi na to, że w programie zakłada się oddzielenie treści dokumentów od ich fizycznego nośnika, czyli *de facto* wykonanie kopii dokumentu, zachodzi potrzeba prawnego uregulowania kwestii sporządzania kopii dokumentów w celu ich długoterminowej archiwizacji. Polskie ustawodawstwo zezwala wprowadzić instytucjom bibliotecznym i archiwalnym sporządzać kopie zgromadzonych materiałów w celach ich ochrony, jednak nie uwzględnia wiążących się z tym, podobnie jak i z innymi zabiegami konserwatorskimi, zmian formy i ewentualnie treści tych materiałów. Na takie zmiany ustawodawca nie wyraża zgody.

W programie proponuje się więc opracowanie projektu zmiany ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, w wyniku której instytucje archiwizujące mogłyby tworzyć kopie opublikowanych utworów oraz przeprowadzać na nich niezbędne prace konserwatorskie typowe dla procesu ich długoterminowej archiwizacji. Proponuje się również zapis dodatkowy, obligujący instytucje archiwizujące do szczegółowego dokumentowania ewentualnych zmian treści i formy publikacji, wraz z zachowaniem kopii utworu w pierwotnej formie. W projekcie powinno się również uwzględnić potrzebę dezaktywacji mechanizmów zabezpieczeń typu DRM (ang. *digital rights management*), w celu przeprowadzania prac konserwatorskich na cyfrowych obiektach archiwalnych, a także ich udostępniania.

Kolejny proponowany projekt zmiany prawa dotyczy przepisów regulujących działalność instytucji powołanych przez MKiDN do pełnienia funkcji centrów kompetencji. Projektowana zmiana powinna włączyć w zakres dotychczasowych obowiązków tych instytucji archiwizację, rozumianą jako zapewnienie długoterminowej/wieczystej użyteczności wszelkich typów materiałów, bez względu na ich formę i nośnik, a więc zarówno elektronicznych materiałów

opublikowanych na nośnikach fizycznych (określanych niekiedy jako nośniki materialne), jak i opublikowanych za pośrednictwem usług sieciowych.

Tym samym, proponuje się zapis w ustawie o egzemplarzu obowiązkowym, z którego wynika, że obowiązkowi przekazania do archiwum podlegają zarówno utwory opublikowane na nośnikach fizycznych, jak i w sieci.

Dodatkowo sugeruje się określić maksymalny czas (np. dwa tygodnie) na przekazanie utworu do instytucji archiwizującej od momentu jego opublikowania i rozpoczęcia rozpowszechniania.

Z uwagi na fakt, że projektowanie zmian aktów prawnych, następnie wprowadzanie poprawek, akceptacja przez właściwe organy państwa, wreszcie ich wejście w życie to procesy czasochłonne, w programie zakłada się, że do czasu ich obowiązywania przedstawione powyżej kwestie gromadzenia i działalności archiwizacyjnej odbywają się na podstawie dobrowolnych umów twórców i instytucji archiwizującej.

Do wszystkich wydawców, miejsc wydawniczych, jednostek organizacyjnych i osób fizycznych niebędących wydawcami, ale prowadzących działalność polegającą na publikowaniu utworów, zostaje skierowany pakiet informacji przybliżający cele i założenia projektu MKiDN i zachęcający do zgłaszania i nadsyłania opublikowanych materiałów 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,
- parametry dotyczące utworzenia podstawowego pakietu metadanych opisowych, technicznych oraz użytkowych,
- parametry dotyczące jakości oraz kompletności tzw. pakietów zgłoszeniowych; oprócz dokumentów i metadanych utworzonych według wytycznych zdefiniowanych przez instytucję archiwizującą, dopuszcza się określenie w indywidualnych umowach z wydawcami, w zależności od typu i specyfiki zgłaszanych materiałów cyfrowych, dodatkowych parametrów dla pakietów zgłoszeniowych,
- czas zgłoszenia dokumentu i przekazania pakietu zgłoszeniowego do instytucji archiwizującej od daty opublikowania,
- wytyczne dotyczące sposobu przekazania pakietu zgłoszeniowego do instytucji archiwizującej,
- założenia dotyczące karencji pierwszego udostępnienia archiwizowanego dokumentu,
- wytyczne dotyczące dezaktywacji mechanizmów kontroli użytkownika dokumentów, np. DRM,
- założenia dotyczące przeprowadzania niezbędnych prac konserwatorskich na dokumentach w celu zapewnienia ich długoterminowej

użyteczności, z uwzględnieniem prawdopodobieństwa odstępstw od pierwotnej formy utworu.

16. MKiDN 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 MKiDN.

Z uwagi na długoterminowy charakter przedsięwzięcia, instytucja koordynująca stara się uzyskać od właściwego ministra potwierdzenie stabilności finansowania działań archiwizacyjnych na długi okres.

Rozwiązaniem optymalnym byłoby ministerialne 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 koordynująca od początku działań opracowuje preliminarz dotyczący działalności archiwizacyjnej. Przy planowaniu poszczególnych zadań 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 odbywają 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 koordynująca prowadzi i, w razie potrzeby, udostępnia skrupulatną dokumentację wszelkich wpływów i ich źródeł oraz poniesionych kosztów.

17. MKiDN definiuje zadania szczegółowe zespołu ds. obsługi technicznej

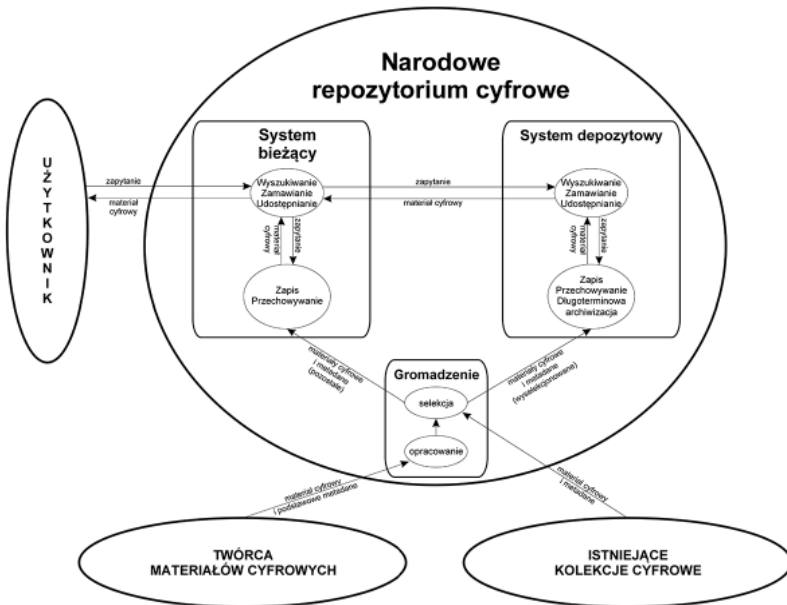
Zespół ds. obsługi technicznej:

- opracowuje koncepcję organizacji i funkcjonowania systemu depozytowego dla polskich zasobów cyfrowych; podstawowe założenie powinno dotyczyć tego, 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 mieć 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 materiałów zagrożonych z dotychczasowych środowisk do otoczenia bezpiecznego, czyli umieszczenie ich w 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.

W programie zakłada się, że do czasu wdrożenia profesjonalnego systemu depozytowego, w pełni odpowiadającego potrzebom długoterminowej archiwizacji polskiego zasobu cyfrowego, polski zasób cyfrowy jest gromadzony i archiwizowany na specjalnie w tym celu zakupionym/ch serwerze/ach, pełniącym/ch rolę tzw. archiwizera przejściowego (tymczasowego).

Planuje się jednak możliwie szybkie wdrożenie systemu depozytowego i umieszczenie w nim zgromadzonej kolekcji. W związku z tym opracowuje się koncepcję organizacji i funkcjonowania polskiego archiwum cyfrowego.



Schemat 4. Organizacja archiwum elektronicznego polskich zasobów cyfrowych

Źródło: opracowanie własne.

W zaproponowanej formie archiwum cyfrowego zasoby depozytowe są udostępniane na bieżąco. W schemacie 4 zakłada się, że docelowo, na mocy odpowiednich przepisów prawnych oraz umów zawartych z wydawcami i miejscami wydawniczymi, wydawcy odsyłają do niej w celach archiwalnych wszystkie dokumenty cyfrowe wraz z dodatkowymi elementami, w postaci pakietów zgłoszeniowych, zdefiniowanych przez instytucję archiwizującą (za OAIS⁹: *Submission Information Package*, SIP). W dziale gromadzenia pakiety zgłoszeniowe zostają poddane procesowi kontroli kompletności, opracowania oraz oceny i selekcji, których celem jest identyfikacja cech dokumentów, decydujących o ich włączeniu do kolekcji dóbr nauki i kultury, tym samym do systemu depozytowego i poddaniu procesowi długoterminowej archiwizacji. Wybrane dokumenty wraz z ich metadanymi, w postaci pakietów archiwalnych (za OAIS: *Archival Information Package*, AIP) zostają przekazane do systemu depozytowego, natomiast wszystkie pozostałe obiekty i ich metadane trafiają do systemu bieżącego i po zamówieniu są udostępniane użytkownikom w postaci pakietów udostępnianych (użytkowych) (za OAIS: *Dissemination Information Package*, DIP).

Istnieje potrzeba opracowania procedur dotyczących formatów zapisu i opisu, a także identyfikowania obiektów cyfrowych, umożliwiających utrzymanie ich długoterminowej autentyczności i integralności 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 publikacji cyfrowych przez okres przejściowy, do czasu opracowania formatów typowo archiwalnych.

Podobną propozycję instytucja archiwizująca wysuwa w związku z potrzebą opisu zasobów cyfrowych, czyli utworzenia ich metadanych. Rozważa się tworzenie metadanych zgodnie z formatem METS bądź 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 dokumentu cyfrowego 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

⁹ A. Januszko-Szakiel, *Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych*, „Przegląd Biblioteczny” 2005, nr 3(73), s. 341–358; *Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data Systems. CCSDS 650.0-B-1. Blue Book, Iss. 1. January 2002*, Consultative Committee for Space Data System, Washington D.C. [online], [dostęp: 10.02. 2012], dostępny w World Wide Web: <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

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 MKiDN 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 sensie stosownie 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.

Podsumowanie

Zaproponowany program zakłada ścisłą współpracę Ogólnopolskiej Grupy Roboczej, ukonstytuowanej głównie z bibliotekarzy, bibliotekoznawców, informatologów, archiwistów, muzealników, menedżerów kultury i oświaty, prawników, informatyków i ekonomistów, w celu opracowania strategii postępowania z cyfrowymi zasobami polskich instytucji nauki i kultury oraz utworzenia centralnego archiwum cyfrowego, zorganizowanego przy MKiDN.

Nie wyklucza to jednak powstawania kolekcji cząstkowych, prowadzonych np. przy centrach kompetencji. Wówczas istotnego znaczenia nabierze fakt opracowania zunifikowanej i powszechnie akceptowanej strategii tworzenia, funkcjonowania archiwów rozproszonych oraz opracowania mechanizmu agregowania metadanych zasobów, składowanych i trwale archiwizowanych w archiwach rozproszonych. W świecie z powodzeniem funkcjonują modele zarówno depozytu narodowego, na który składa się sieć archiwów rozproszonych, jak i depozytu centralnego, w którym w jednym miejscu są składowane i trwale chronione zasoby cyfrowe tworzące narodową kolekcję dziedzictwa nauki i kultury. W takim przypadku istotna jest dbałość o terytorialne oddalenie kilku kopii archiwalnego zasobu, najlepiej u zagranicznych partnerów.

Z analizy wypowiedzi prezentowanych w ostatnich miesiącach przez przedstawicieli czołowych instytucji nauki i kultury Polski (MKiDN, BN, NAC) wynika, iż są one wyraźnie skłonne 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.

Bibliografia

1. Statut BN [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.bn.org.pl/o-bn/statut-bn>.
2. Duńczyk-Szulc A., *Polskie i europejskie programy digitalizacyjne na przykładzie WPR Kultura+ i EuropeanaPhotography*, wystąpienie na konferencji: Dziedzictwo w sieci – różne aspekty digitalizacji, Kraków 29–30 listopada 2012.
3. Januszko-Szakiel A., *Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych*, „Przegląd Biblioteczny” 2005, nr 3(73), s. 341–358.
4. *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020* [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.kongreskultury.pl/library/File/Raport-Digitalizacja/Program%20digitalizacji%202009-2020.pdf>.
5. Statut Narodowego Instytutu Audiowizualnego [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://bip.mkidn.gov.pl/media/download_gallery/index6ba4.pdf.
6. Statut Narodowego Instytutu Muzealnictwa i Ochrony Zbiorów [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://bip.mkidn.gov.pl/media/download_gallery/20121231p-88_Z_NIMOZ.pdf.
7. O NAC. Zadania [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.nac.gov.pl/pl/zadania>.
8. Statut Narodowego Instytutu Dziedzictwa [online], [dostęp: 12.09.2012], dostępny w World Wide Web: <http://www.nid.pl/idm,13,statut.html>.
9. Zalecenie Komisji z 27.10.2011 w sprawie digitalizacji i udostępniania w Internecie dorobku kulturowego oraz w sprawie ochrony zasobów cyfrowych [online], [dostęp: 12.09.2012], dostępny w World Wide Web: http://ec.europa.eu/information_society/activities/digital_libraries/doc/recommendation/recom28nov_all_versions/pl.pdf.

