Grzegorz Gmiterek
(Warszawa)

Długoterminowa archiwizacja zasobów cyfrowych

Niezależnie od czasu, jaki upłynął od stworzenia obiektu cyfrowego,
użytkownik musi posiadać możliwość odtworzenia zawartej w nim informacji.
Adam Dudczak
Poznańskie Centrum Superkomputerowo-Sieciowe[1]

Ochrona zasobów cyfrowych to dzisiaj jedno z najważniejszych zadań stojących przed instytucjami kultury i nauki. Problemy z trwałością zasobów są postrzegane w kategoriach zagrożeń, które dotyczą nie tylko utraty lub zniszczenia samego nośnika zapisu, ale przede wszystkim jego trwałości czy starzenia się formatów umożliwiających użytkownikowi zapoznanie się z materiałami cyfrowymi. Zmianom podlegają zarówno stacje komputerowe, wykorzystywane w ramach tworzenia i odczytu cyfrowych treści, jak i oprogramowanie służące do ich prezentacji czy środowisko infrastrukturalne udostępniające obiekty (na przykład w ciągu ostatnich 20 lat mieliśmy do czynienia z pięcioma wersjami hipertekstowego języka znaczników HTML służącego do tworzenia stron internetowych). Taki stan rzeczy wymusza ciągłą migrację do nowych formatów, jak również emulację starszych aplikacji (migracja danych do nowszych formatów, ewentualnie emulacja starszych technologii na współczesnych platformach technologicznych[2]) tak, aby zarówno dzisiejszy, jak i przyszły użytkownik mógł bez problemu zapoznać się z elektronicznymi publikacjami bez obawy o ich użyteczność, kompletność czy integralność i autentyczności. Do niedawna wydawało się, że

[1] Zob. A. Dudczak, *Długoterminowe przechowywanie zasobów cyfrowych*. [on-line] [dostęp 3 października 2012 r.] Dostępny w World Wide Web: http://mbc.malopolska.pl/dlibra/doccontent?id=13392&dirids=1

[2] Por. *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*. [In:] kongreskultury.pl [on-line] [dostęp 25 września 2012 r.]. Dostępny w World Wide Web: http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf s. 47.

sposobem na archiwizację danych jest ich przechowywanie na takich nośnikach, jak dyski optyczne, CD-ROMy. Przy czym nietrwałość tych rozwiązań, niewystarczający (zbyt krótki) czas ich użytkowania został wielokrotnie potwierdzony i poparty doświadczeniami osób zarówno tworzących, jak i próbujących się zapoznać z danymi na nich zapisanymi[3]. Zwraca się także uwagę na niedostateczną refleksję osób odpowiedzialnych za ochronę danych przechowywanych na „tradycyjnych" nośnikach cyfrowych w różnych instytucjach. Nie ulega wątpliwości, że brak zrozumienia istoty ewolucji dokonującej się w świecie technologii i zapisu informacji może być powodem bezpowrotnej utraty możliwości odczytu elektronicznych zasobów. Niestety, w wielu przypadkach dopiero ta utrata prowadzi do wniosku, że cyfrowe zasoby (często ważne czy cenne dla danej instytucji dokumenty) powinny być chronione i przechowywane w bezpieczny dla nich sposób. Podobnie jak ma to miejsce z koniecznością ochrony i archiwizacji tradycyjnych dokumentów. Nie ma przy tym znaczenia, czy dany dokument powstał jako „born digital" (w wersji elektronicznej), czy jest cyfrową wersją dokumentu analogowego.

Sam termin „długoterminowa archiwizacja zasobów cyfrowych", analogicznie do „ochrony informacji cyfrowych", rozumiany jest w tym miejscu jako swoisty „zbiór wszystkich procesów służących zapewnieniu ciągłości dostępu do materiałów należących do dziedzictwa cyfrowego w okresie, w którym ciągłość taka jest pożądana"[4]. Zgodzić się też należy z A. Januszko-Szakiel, że sama dostępność materiałów to nic innego jak zachowanie środków zapewniających dostęp do autentycznej treści cyfrowego dokumentu, ale także jego wykorzystanie zgodnie z celem, w jakim został stworzony[5]. Z drugiej strony, mamy do czynienia z pewnego rodzaju paradoksem, który trafnie scharakteryzował Su-Sing Chen, profesor z Wydziału Informatyki i Inżynierii Informacji Uniwersytetu na Florydzie. Stwierdził on mianowicie, że „tradycyjnie, ochrona rzeczy oznaczała zachowanie ich w niezmienionej formie; tymczasem rzeczywistość cyfrowa zmieniła fundamentalnie naszą koncepcję wymagań konserwacyjnych. Jeżeli zachowamy

[3] M. Brzeźniak mówi o 5–10 latach w przypadku dysków twardych, płyt DVD i CD oraz 20 latach – pamięci flash, nośników taśmowych – ok. 30 lat. Zob. M. Brzeźniak, *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych*. „EBIB" 2010, nr 6, [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.ebib.info/2010/115/a.php?brzezniak

Zob. także A. Radwański, *Techniczne i organizacyjne aspekty archiwizacji dużych kolekcji cyfrowych*, [w:] *Od digitalizacji zaawansowanej do dojrzałej*. (Materiały CPI), Warszawa 2009, s. 49–62; S. Żabicki, *Przechowywanie danych – wyzwanie dla cyfrowych bibliotek*, „EBIB" 2010, nr 6, [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.ebib.info/2010/115/a.php?zabicki

[4] *Ochrona dziedzictwa cyfrowego: zalecenia*, Oprac. National Library of Australia, Warszawa 2003, s. 44.

[5] A. Januszko-Szakiel, *Długoterminowa archiwizacja*…, s. 92.

informację cyfrową bez modyfikacji, dostęp do niej stanie się niezmiernie trudny, jeżeli nie niemożliwy”[6].

Długoterminowa ochrona zasobów cyfrowych to dziś jeden z najistotniejszych problemów każdego rozwijającego się społeczeństwa informacyjnego. Jak podaje Piotr Beńke, architekt systemów informatycznych firmy IBM, rozmiar współczesnego świata cyfrowego to 1.2 zetabajty (1.2 x 1012 gigabajów), a z każdym rokiem świat ten jest większy o 50% (przy czym 25% to zasoby unikalne, 75% – kopie)[7]. Tempo przyrostu danych jest więc imponujące. Biblioteki jako instytucje archiwizujące i udostępniające tradycyjne i cyfrowe dokumenty są zobligowane do konsekwentnych działań na rzecz ochrony tych zasobów. Co więcej, należą one dzisiaj, obok uczelni, centrów badawczych, archiwów, do największych producentów elektronicznych danych.

Jednak samo archiwizowanie zasobów nie jest łatwym procesem[8]. Ze względu na duże ryzyko utraty danych często wymaga profesjonalnych oraz specjalistycznych rozwiązań[9]. Ale powodów takiego stanu rzeczy jest znacznie więcej. W pierwszej kolejności należałoby wymienić okresową konieczność migracji danych, rozbudowę systemu zarządzającego zarchiwizowanymi zasobami, prawidłową organizację metadanych[10], konieczność zapewnienia dostępu w perspektywie długoterminowej (rozumianej jako nieograniczonej w czasie czy możliwie najbardziej odległej w przyszłości[11]), bezpieczeństwo, poufność, kompletność i niezmienność danych, utrzymanie ich substancji (zapis treści na nośniku elektronicznym, ewentualna zmiana generacji nośnika), autentyczność (wszelkie informacje zawarte w chronionym dokumencie pochodzą od jego autora i od dnia ich publikacji nie uległy zmianie, tzn. przedstawiają dokładnie to,

---

[6] S. Chen, *The paradox of digital preservation*, “Computer”, 2001 March, pp. 2–6, tłum. Yaniv Levi (źródło: materiały wewnętrzne Ex Libris).

[7] P. Beńke, *Archiwum cyfrowe jako usługa modelu Cloud Computing*. [In:] ibm.com [dostęp 26 września 2012 r.] Dostępny w World Wide Web: http://www-05.ibm.com/pl/cyfrowa_kultura/download/Piotr_BeSke_IBM.pdf

[8] Sam termin „archiwizacja” za T. Kalotą i M. Szalą rozumiem jako „ogół metod i praktyk służących zabezpieczaniu i długoterminowemu przechowywaniu wytwarzanych zasobów cyfrowych. Na archiwizację składają się: infrastruktura techniczna, oprogramowanie, sprzęt, interfejsy użytkowników, mechanizmy wymiany danych, mechanizmy zabezpieczenia i migracji danych, obsługa techniczna”. Zob. T. Kalota, M. Szala, *Organizacja i logistyka digitalizacji*, [w:] *Biblioteki cyfrowe*, red. M. Janiak, M. Krakowska i M. Próchnicka, Warszawa 2012, s. 438.

[9] Zob. D. Wolski, *7 grzechów archiwizacji danych*, „PC World” 2012, nr 9, s. 98–101. Zob. także A. Radwański, *op. cit*.

[10] Metadane rozumiem jako „ustrukturyzowane dane o opisywanym obiekcie, wspomagające wyznaczone dla niego funkcje; przy czym obiektem jest wówczas każda jednostka, dla której można wyznaczyć i zapisać dane kontekstualne”. Za: M. Nahotko, *Opis dokumentów*, [w:] *Biblioteki cyfrowe*…, s. 363.

[11] A. Januszko-Szakiel, *Długoterminowa archiwizacja zasobów cyfrowych w świadomości pracowników polskich bibliotek*, [w:] *Polskie Biblioteki Cyfrowe 2009*, red. C. Mazurek, M. Stroiński i J. Węglarz, Poznań 2010, s. 92

co było zamierzeniem ich twórców)[12] czy integralność danych[13]. Bardzo ważne jest tutaj stosowanie odpowiedniego systemu kopii zapasowych (ewentualnie systemu hierarchicznego przechowywania[14]). Pojawiają się także inne problemy, z których najważniejszym wydają się rosnące koszty przechowywania danych cyfrowych[15].

Planując poszczególne procedury i etapy archiwizacji zasobów cyfrowych, niezbędne jest także dokonanie wyboru, jakiego standardu opisu dokumentów użyjemy[16]. Wybranie i wykorzystanie odpowiedniego to niezbędny wstęp do ustalania zasad postępowania z archiwizowanym dokumentem. Z punktu widzenia procesu długoterminowej archiwizacji zasobów cyfrowych w pierwszej kolejności oczywiście wymienimy metadane opisowe, tworzone w celu szczegółowej identyfikacji obiektu i późniejszego odnalezienia go przez użytkownika (np. standard Dublin Core i zestaw 15 elementów wykorzystywanych do opisu różnego rodzaju zasobów). Ważne będą również dane odnoszące się do proweniencji (tzw. metadane proweniencji), definiujące źródło lub pochodzenie treści danego obiektu, zawierające informacje, jakie zmiany zostały zastosowane w stosunku do dokumentu od jego powstania (np. informacje dotyczące migracji ze starszego formatu na nowszy). Wymienimy także techniczne metadane, na które składa się kompletny i wyczerpujący zestaw informacji odnoszący się do formatu, w jakim został zapisany obiekt, skali kompresji, sprzętu i oprogramowania niezbędnego do jego odtworzenia itp. Nie bez znaczenia będą też informacje dotyczące prawnych aspektów wykorzystania czy udostępniania dokumentów (prawa własności intelektualnej, np. informuje, kto posiada prawa do danego materiału), ale także informacje o odnośnikach czy relacjach z innymi obiektami. Również

[12] *Ibid*.

[13] P. Beńke, *op. cit*.

[14] System hierarchicznego przechowywania (Hierarchical Storage Management) to system przechowywania dużej ilości danych w celu ich archiwizowania, składowania, wykonywania kopii zapasowych i wsparcia procesu odtwarzania danych kopii. Wiąże się z automatycznym przenoszeniem danych archiwalnych na inne nośniki informacji. Warto zaznaczyć, że nad precyzyjną logiką stosowania tych automatów konieczny jest ścisły operatorski nadzór. Na podstawie J. Gnybek: *Information Lifecycle Management w Oracle Database 11g*. [In:] ploug.org.pl [on-line] [dostęp 26 września 2012 r.]. Dostępny w World Wide Web: http://www.ploug.org.pl/plougtki.php?action=read&p=45&a=6

[15] Zob. J. Palm, *Cyfrowa czarna dziura – Koszty długoterminowego przechowywania* [online] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.nina.gov.pl/docs/default-document-library/JonasPalm%20Cyfrowa%20czarna%20dziura%20-%20koszty%20przechowywania.pdf Zob. także D. A. Warner, *Libraries, archives, and digital preservation: a critical overview*, [w:] *Information Technology in Librarianship. New Critical Approaches*, Westport, Connecticut, London 2009, s. 274–276.

[16] Jest wiele standardów i schematów zapisu metadanych (mówi się nawet o stu pięciu standardach). Zob. *Seeing Standards: A Visualization of the Metadata Universe* [on-line] [dostęp 4 października 2012 r.]. Dostępny w World Wide Web: http://www.dlib.indiana.edu/~jenlrile/metadatamap/

autentyczność i integralność obiektu będzie związana z jego opisem. Jak zaznacza M. Nahotko, „autentyczność (czyli kontrola treści) oznacza odpowiedź na pytania o znaczenie źródła, jego proweniencję i w związku z tym – opiera się na metodach identyfikacji i weryfikacji źródła". Z drugiej strony, badacz wymienia autoryzację (kontrolę dostępu), która jest związana z metodami identyfikacji użytkownika oraz jego uprawnień w zakresie źródła informacji[17]. Jeszcze inne typy metadanych będą dotyczyły bezpieczeństwa (ochrony zasobów) czy magazynowania obiektów cyfrowych (np. informacje o „odświeżeniu" obiektu, poddaniu go okresowej kontroli)[18]. Mamy wreszcie do czynienia z metadanymi strukturalnymi, które są bezpośrednio związane z przechowywaniem obiektu i jego prezentacją. Trafne wydają się więc słowa autorów pracy *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*, którzy wskazują, że jedne typy metadanych „opisują oryginalny obiekt digitalizowany, inne charakteryzują proces jego digitalizacji, inne opisują strukturę złożonego obiektu cyfrowego, inne służą do bieżącego zarządzania obiektem cyfrowym, jeszcze inne do zarządzania obiektem cyfrowym podczas jego długotrwałego przechowywania"[19]. W tym przypadku różnorodność opisu jest więc kluczowym elementem dla prawidłowej organizacji, przechowywania, udostępniania, ale także ochrony dokumentów.

W przypadku długoterminowej archiwizacji zasobów cyfrowych bardzo ważne jest określenie zestawu i struktury metadanych w celu uniezależnienia jej od aktualnie stosowanej technologii. Dzisiejsze technologie, wykorzystywane w celu tworzenia elektronicznych dokumentów, w przyszłości zostaną zastąpione innymi. Prowadzić to może do problemów związanych zarówno z prawidłowym odczytem dokumentów, jak i ich metadanych. Z drugiej strony, głównym celem doboru odpowiedniego standardu metadanych powinno być takie opisanie dokumentu, aby nie było potrzeby poddawania go dodatkowemu opisowi podczas procesu archiwizowania (pomijając oczywiście czynności konieczne, takie jak na przykład porządkowanie obiektów czy zmianę formatu ich zapisu).

Mówiąc o długoterminowej archiwizacji zasobów cyfrowych prowadzonej przez instytucje biblioteczne i archiwa, warto zaznaczyć, że problem ten od dłuższego czasu jest obecny i dyskutowany w kręgach naukowo-badawczych, ale także w wielu placówkach gromadzących i udostępniających informacyjne zasoby. Coraz częściej programy służące ochronie cyfrowych danych traktowane są jako oczywistość w zabezpieczaniu wartościowych dokumentów dostępnych w formie elektronicznej. Przykładem takich działań mogą być dla nas projekty realizowane

[17] M. Nahotko, *Metadane. Sposób na uporządkowanie Internetu*, Kraków 2004, s. 43.

[18] Zob. N. Sherwinter, R. Wright, *Strategy for use of preservation metadata within a digital library*. [on-line] [dostęp 4 października 2012 r.]. Dostępny w World Wide Web: http://tinyurl.com/a4gspce

[19] *Standardy w procesie digitalizacji obiektów dziedzictwa kulturowego*, red. G. Płoszajski, Warszawa 2008, s. 9.

przez biblioteki narodowe państw zachodnich, przedsięwzięcia służące tworzeniu programów długoterminowej archiwizacji i właśnie ochrony danych cyfrowych. Za wzór może posłużyć narodowa strategia długoterminowej archiwizacji przygotowana w 2002 r. przez amerykańską Bibliotekę Kongresu oraz Radę Zbiorów Bibliotecznych i Informacyjnych. Jak czytamy w dokumencie głównym dotyczącym tego przedsięwzięcia, „misją Narodowej Infrastruktury Informacji Cyfrowej i Programu Ochrony jest gromadzenie, archiwizowanie i zabezpieczanie rosnących ilości danych, zwłaszcza materiałów, które mają tylko formę elektroniczną (tzw. born digital) dla obecnych i przyszłych pokoleń”[20]. Ważna z punktu widzenia omawianego tematu jest również „Karta ochrony dziedzictwa cyfrowego” UNESCO, która nakłada na poszczególne państwa obowiązek przeciwdziałania utracie cyfrowego dziedzictwa. Wspomnieć też należy o dokumencie opracowanym przez Bibliotekę Narodową Australii na zlecenie UNESCO, zawierającym wskazówki i zalecenia techniczne dotyczące ochrony oraz zapewnienia ciągłej dostępności powiększającego się dziedzictwa cyfrowej rzeczywistości[21]. Zresztą wymieniona biblioteka to instytucja, w której najwcześniej, bo już w latach 1993–1995, podjęto dyskusję nad problemem archiwizacji cyfrowych dokumentów tworzonych i udostępnianych przez książnice[22]. W Bibliotece Narodowej w Australii stworzono także tematyczny serwis internetowy. Preserving Access to Digital Information (PADI), w którym udostępnione są szczegółowe informacje na temat zabezpieczenia zasobów cyfrowych oraz szeroko pojętej tematyki archiwizacji internetu. Znajdziemy tutaj informacje o wielu światowych projektach ochrony elektronicznych dokumentów[23]. Na łamach tego serwisu przedstawiono też doświadczenia wiodących instytucji wdrażających krajowe programy ochrony zasobów sieciowych. We wszystkich przypadkach to właśnie biblioteki (główne, centralne) są współodpowiedzialne za proces archiwizacji i udostępniania zasobów[24].

---

[20] *Building a National Strategy for Digital Preservation: Issue in Digital Media Archiving*. [online] [dostęp 29 września 2012 r.]. Dostępny w World Wide Web: http://www.clir.org/pubs/reports/pub106/pub106.pdf

[21] Zob. *Ochrona dziedzictwa cyfrowego*…

[22] Wynikiem tych dyskusji prowadzonych wśród przedstawicieli bibliotek i archiwów było powołanie grupy roboczej PADI (Preserving Access to Digital Information), która miała na celu prowadzenie prac nad projektowaniem i rozwojem narodowej strategii długoterminowej ochrony zasobów cyfrowych w Australii. Więcej na ten temat można przeczytać w publikacji S. Paper, *PADI (Preserving Access to Digital Information) and Safekeeping* [on-line] [dostęp 30 września 2012 r.] Dostępny w World Wide Web: http://192.102.239.53/openpublish/index.php/nlasp/article/viewFile/1241/1526

[23] PADI – Projects and Case Studies. [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://pandora.nla.gov.au/pan/10691/20110824-1153/www.nla.gov.au/padi/format/case.html

[24] PADI – Web Archiving. [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://pandora.nla.gov.au/pan/10691/20110824-1153/www.nla.gov.au/padi/topics/92.html#cz

Charakteryzując programy długoterminowej archiwizacji zasobów cyfrowych, nie sposób nie wspomnieć o europejskim konsorcjum Digital Preservation Europe. Jego działania mają na celu „wspierać innowacje i przedstawiać prognozy o niebezpieczeństwach grożących procesowi przechowywania danych w formie cyfrowej, z którymi ma do czynienia współczesne społeczeństwo”[25]. W skład konsorcjum wchodzą uczelnie wyższe z różnych państw Europy (Wielkiej Brytanii, Austrii, Niemiec, Włoch, Litwy, Danii, Holandii, Czech). W ramach dofinansowania ze środków unijnych projekt zakłada przygotowanie systemu szkoleń (konferencje, kursy, seminaria) służącego doskonaleniu kwalifikacji w dziedzinie digitalizacji i archiwizacji cyfrowych dokumentów, publikacji materiałów na ten temat oraz analizie potrzeb już istniejących elektronicznych archiwów, stopnia ich przygotowania do długoterminowej ochrony zasobów[26].

Zresztą długoterminowa ochrona informacji to problem, który jest postrzegany przez europejskie instytucje administracyjne jako poważne wyzwanie dla rozwijających się społeczeństw. Jak dostrzegają A. Jakubiec i M. Pazdur, „Unia Europejska w ramach specjalnie powołanej agendy cyfrowej wspiera różne projekty, które umożliwiają nieograniczone w czasie zachowywanie, udostępnianie i wykorzystywanie zapisanych cyfrowo danych”. Badaczki wymieniają projekt CASPAR (Międzynarodowy Zintegrowany Projekt Badawczy) poświęcony zapisowi, udostępnianiu oraz wyszukiwaniu informacji dotyczących kultury, nauki i sztuki. Projekt był wdrażany w okresie od 1 kwietnia 2006 r. do 30 września 2009 r., a uczestniczyły w nim zarówno instytucje naukowe i akademickie, jak i firmy komercyjne z pięciu państw europejskich (Czech, Francji, Grecji, Wielkiej Brytanii i Włoch) oraz z Izraela[27].

Warto także choćby wspomnieć o innych europejskich projektach służących długoterminowej archiwizacji zasobów sieciowych. Joanna Potęga w publikacji *Cyfrowa Biblioteka Narodowa* wymienia w pierwszej kolejności realizowany w Szwecji od 1999 r. projekt Kulturarw3 (www.kb.se/kw3), fiński NEDLIB – Networked European Deposit Library (http://nedlib.kb.nl/index.html) czy program ochrony czeskich zasobów sieciowych WebArchiv (http://en.webarchive.cz), realizowany w czeskiej bibliotece narodowej, co ważne – wspierany przez Ministerstwo Kultury Czech. Warto zaznaczyć, że w przypadku fińskiego przedsięwzięcia stało się ono podstawą do uruchomienia międzynarodowego projektu o tej samej

[25] Zob. strona główna projektu [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.digitalpreservationeurope.eu/about/

[26] Więcej na temat projektów archiwizacji zasobów sieciowych, również przy udziale bibliotek, można przeczytać w artykułach dostępnych w Biuletynie „EBIB” poświęconemu archiwizacji internetu („EBIB” 2012, nr 1 [on-line] [dostęp 3 października 2012 r.] http://www.ebib.info/component/content/article/3/895-biuletyn-ebib-nr-12012128).

[27] Zob. A. Jakubiec, M. Pazdur, *CASPAR – długoterminowa archiwizacja obiektów cyfrowych*. [on-line] [dostęp 11 października 2012 r.]. Dostępny w World Wide Web: http://suw.biblos.pk.edu.pl/downloadResource&mId=219578

nazwie, jednak współtworzonego przez biblioteki narodowe Holandii, Finlandii, Francji, Niemiec, Norwegii, Portugalii, Szwajcarii oraz Włoch[28].

Również pracownicy polskich instytucji bibliotecznych dostrzegają wagę i złożoność tego problemu, o czym mogą świadczyć pojawiające się publikacje na ten temat na łamach branżowych i naukowych wydawnictw. Od niedawna również w Polsce mówi się o problemie długoterminowej archiwizacji zasobów cyfrowych, dotychczas tworzonych i udostępnianych przez biblioteki. Kompleksowego opracowania tego problemu dokonała w swoich publikacjach Aneta Januszko-Szakiel. Niestety, jak stwierdza badaczka, „problematyka długoterminowej ochrony cyfrowych zasobów bibliotecznych została podjęta w polskich instytucjach bibliotecznych z około piętnastoletnim opóźnieniem w stosunku do instytucji pamięci krajów zaawansowanych w archiwistyce cyfrowej"[29]. Efektem przeprowadzonych przez A. Januszko-Szakiel badań jest autorska, ambitna, ale niestety póki co mająca tylko hipotetyczny charakter propozycja Programu Długoterminowej Archiwizacji Polskich Zasobów Cyfrowych, któremu przewodniczyć miałaby Biblioteka Narodowa, jako centralna książnica Polski, ustawowo powołana do wieczystej archiwizacji polskiego dziedzictwa nauki i kultury. Zaproponowany Program składa się trzech części dotyczących następujących zagadnień: akceptacji nowych zadań i działań wstępnych w BN; organizacji pracy w zakresie długoterminowej archiwizacji polskich zasobów elektronicznych; planowania szczegółowych zadań Ogólnopolskiej Grupy Roboczej ds. długoterminowej archiwizacji polskich zasobów cyfrowych[30]. Ciekawy w odniesieniu do problemu długoterminowej archiwizacji wydaje się pomysł powołania do życia swoistego bibliotecznego systemu depozytowego, w którym deponowano by zarówno dokumenty będące wynikiem dotychczasowej digitalizacji (np. dokumenty, do których wygasły już prawa autorskie czy zbiory specjalne), jak i wydawane współcześnie (niezależnie od ich pierwotnej analogowej czy elektronicznej formy). W tym ostatnim przypadku wydawcy byliby zobligowani do przekazania cyfrowego egzemplarza obowiązkowego wydawanego przez nich dokumentu[31].

---

[28] J. Potęga, *Cyfrowa Biblioteka Narodowa*, [w:] *Biblioteki cyfrowe. Projekty, realizacje, technologie*, red. J. Woźniak-Kasperek i J. Franke, Warszawa 2007, s. 37.

[29] A. Januszko-Szakiel, *Nowy wymiar zabezpieczenia elektronicznych zasobów*, „EBIB" 2010, nr 6 [on-line] [dostęp 30 września 2012 r.]. Dostępny w World Wide Web: http://www.ebib.info/2010/115/a.php?januszko

[30] Zob. A. Januszko-Szakiel, *Długoterminowa archiwizacja zasobów cyfrowych – program dla polskich bibliotek*, „Przegląd Biblioteczny" 2011, nr 2, s. 211–230. Zob także *Długoterminowa archiwizacja zasobów cyfrowych dla polskich bibliotek* (program) [In:] historiaimedia.org [on-line] [dostęp 30 września 2012 r.]. Dostępny w World Wide Web: http://historiaimedia.org/2011/10/22/dlugoterminowa-archiwizacja-zasobow-cyfrowych-dla-polskich-bibliotek-program/

[31] Więcej na ten temat w artykule A. Januszko-Szakiel, *Długoterminowa archiwizacja*…, s. 218–223.

Warto także wspomnieć o projekcie Biblioteki Narodowej, który odnosi się do stworzenia „Narodowego Repozytorium Dokumentów Elektronicznych". Głównym celem tego projektu jest zabezpieczenie dotychczas zdigitalizowanych przez biblioteki materiałów, danych dostępnych na nośnikach fizycznych i online (zwłaszcza książek i czasopism), ale także wyselekcjonowanych zasobów sieciowych[32]. Zresztą mówiąc o działalności bibliotek narodowych, warto zaznaczyć, że mają one w przypadku długoterminowej ochrony zasobów cyfrowych szczególne zadania. Odnoszą się oczywiście do zarządzania i koordynacji poszczególnych procesów archiwizacji, ale w tym przypadku dotyczą zwłaszcza zapewnienia dostępu do różnorodnych kolekcji dla przyszłych pokoleń. Nie ma przy tym znaczenia, czy są to cyfrowe kolekcje dokumentów tekstowych, graficznych, audiowizualnych, audialnych czy na przykład 3D. Głównym celem jest zachowanie cyfrowego dziedzictwa kulturowego, a biblioteki narodowe mają największe możliwości sprostać temu problemowi[33].

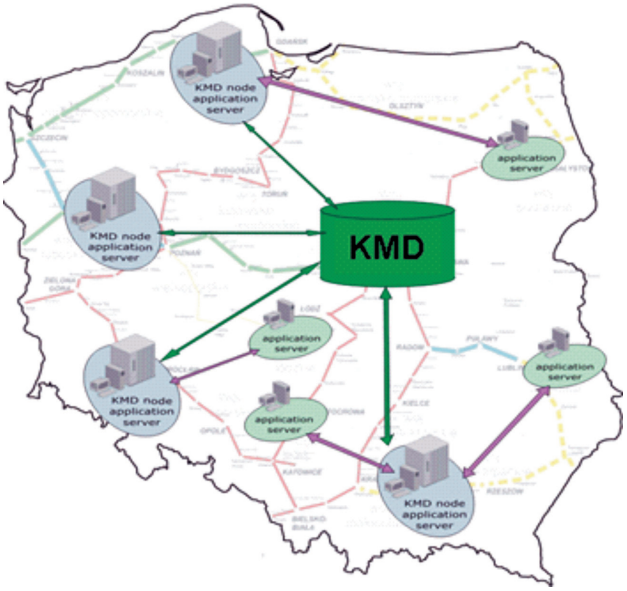
Innym ważnym przedsięwzięciem koncentrującym się na problemie archiwizacji zasobów cyfrowych i ich ochrony jest Krajowy Magazyn Danych, którego koordynatorem jest Poznańskie Centrum Superkomputerowo-Sieciowe. Jak czytamy na stronie głównej projektu: „Projekt Krajowy Magazyn Danych (KMD) ma na celu zbudowanie rozproszonego systemu przechowywania danych o zasięgu krajowym, który dostarcza wydajnych metod dostępu do danych oraz operacji na danych oraz zapewnia wysoki poziom wiarygodności, niezawodności oraz bezpieczeństwa danych. […] Usługi przechowywania mają być oferowane w obrębie sieci Polskiego Internetu Optycznego PIONIER, która łączy instytucje naukowe i edukacyjne w Polsce oraz w sieciach miejskich MAN. Głównym zastosowaniem systemu jest usługa tworzenia kopii zapasowych i archiwizacji (ang. Backup/Archive), dla instytucji akademickich, edukacyjnych, samorządów oraz innych instytucji publicznych"[34]. Przedsięwzięcie było realizowane w latach 2007–2009 i zakończyło się wytworzeniem prototypu oprogramowania do rozproszonego i wiarygodnego przechowywania danych z automatyczną replikacją geograficzną. W tym przypadku mamy więc do czynienia z projektem archiwi-

[32] Zob. J. Potęga, *Narodowe repozytorium dokumentów elektronicznych*, [w:] sbc.org.pl [on-line] [dostęp 25 września 2012 r.]. Dostępny w World Wide Web: http://www.sbc.org.pl/Content/10978/15_Potega_Narodowe_repozytorium.pdf Zob. także *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*, [w:] kongreskultury.pl [on-line] [dostęp 25 września 2012 r.]. Dostępny w World Wide Web: http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf s. 41–49.

[33] M. Breeding, *The Special Challenges for Nationals Libraries*, [on-line] [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://www.librarytechnology.org/ltg-displaytext.pl?RC=15788

[34] Więcej na ten temat można dowiedzieć się ze strony projektu. Zob. http://kmd.pcss.pl/ [on-line] [dostęp 30 września 2012 r.]. Zob. także M. Brzeźniak, *op. cit*.

zacji danych rozproszonych geograficznie, ale połączonych jednym systemem informatycznym[35].



Rys. 1. Rozmieszczenie Węzłów oraz usług systemu KMD. Źródło: http://kmd.pcss.pl/index.html

Z rozmowy z Maciejem Brzeźniakiem, jedną z osób odpowiedzialną za wdrożenie KMD, wynika, że prototyp ten został wdrożony (realizacja w latach 2008–2012) jako tzw. Usługa Powszechnej Archiwizacji w ramach projektu PLATON (Platforma Obsługi Nauki PLATON – Etap I: Kontener usług wspólnych)[36]. Usługę realizuje 10 ośrodków (5 centrów Komputerów Dużej Mocy w Gdańsku,

[35] Szczegółowe informacje na temat krajowej, naukowej, szerokopasmowej sieci optycznej Pionier znajdziemy w serwisie *Polish Optical Internet Pionier Magazine* [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.pionier.net.pl oraz artykule Katarzyny Siudzińskiej *Polski Internet Optyczny Pionier* [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1356/polski_internet_optyczny_pionier.html

[36] Więcej informacji na ten temat dostępnych jest na stronie internetowej powiązanej z tym projektem. [on-line] [dostęp 13 października 2012 r.] https://www.storage.pionier.net.pl/ Zob. także stronę WWW Projektu PLATON http://www.platon.pionier.net.pl/online/ [on-line] [dostęp 13 października 2012 r.]. Zob. także M. Brzeźniak, *Usługa Powszechnej Archiwizacji i jej zastosowanie w bibliotekach naukowych do zabezpieczenia i archiwizacji danych*, „EBIB" 2010, nr 6. [on-line] [dostęp 13 października 2012 r.]. Dostępny w World Wide Web: http://www.ebib.info/2010/115/a.php?brzezniak; M. Brzeźniak, *Usługi powszechnej archiwizacji*. [on-line] [dostęp 13 października 2012 r.]. Dostępny w World Wide Web: http://www.pionier.net.pl/magazine/pl/artykuly/1374/PLATON_Uslugi_powszechej_archiwizacji.html

Krakowie, Warszawie, Poznaniu i Wrocławiu, 5 ośrodków infrastruktury sieci miejskiej MAN i uczelni). Z usługi tej korzystają biblioteki cyfrowe, uczelnie, ośrodki badawcze itp.[37].

Koncepcja i architektura KMD jest nadal kontynuowana w ramach projektu KMD2 (System bezpiecznego przechowywania i współdzielenia danych oraz składowania kopii zapasowych i archiwalnych w Krajowym Magazynie Danych)[38]. Realizacja tego przedsięwzięcia jest przewidziana na lata 2012–2013. Jak zaznacza M. Brzeźniak, „w tym przypadku architektura rozwijana jest w kierunku mechanizmów zapewniających wysokie bezpieczeństwo składowania danych w systemie KMD (poufność, kontrola integralności danych), a także umożliwiających bezpieczne współdzielenie danych. W dziedzinie publikacji danych projekt KMD2 opracowuje mechanizm wydajnego udostępniania do sieci dużej objętości treści wideo".

Ta sama instytucja, czyli Poznańskie Centrum Superkomputerowo-Sieciowe, wdrożyła w ostatnim czasie projekt RCIN (Repozytorium Cyfrowe Instytutów Naukowych), którego jednym z głównych celów jest „zabezpieczenie bieżącego dorobku naukowego Instytutów wchodzących w jego skład, a także ułatwienie wszystkim zainteresowanym dostępu do publikacji zarezerwowanych dotychczas wyłącznie dla wąskich, wyspecjalizowanych grup badaczy"[39]. Projekt jest finansowany ze środków unijnych w ramach Programu Operacyjnego Innowacyjna Gospodarka (działanie: Inwestycje związane z rozwojem infrastruktury informatycznej nauki). W skład przedsięwzięcia wchodzą dokumenty zdigitalizowane głównie w bibliotekach instytutów Polskiej Akademii Nauk. W ramach tego projektu PCSS wdrożyło także zestaw usług wspierających długoterminową archiwizację zasobów sieciowych (zwłaszcza tekstowych, graficznych i audiowizualnych). Zestaw ten nosi nazwę dArceo. Należy dodać, że PCSS jest też producentem oprogramowania do budowy bibliotek cyfrowych dLibra (rozwijanego od 1996 r.), będącego podstawą działania blisko stu instytucjonalnych i regionalnych repozytoriów gromadzących dzisiaj ponad milion dokumentów dostępnych

[37] Na potrzeby tych instytucji udostępniono przestrzeń 12,5 PB pamięci taśmowej oraz 2 TB pamięci dyskowej z przeznaczeniem na przechowywanie i buforowanie danych. Cała infrastruktura, na której projekt się opiera, to ponad 70 serwerów, 5 bibliotek taśmowych oraz 10 macierzy dyskowych rozmieszczonych w 10 lokalizacjach. Każda instytucja akademicka i naukowa chcąca wykorzystywać tę usługę może przystąpić do usługi. Więcej informacji na ten temat znajdziemy na stronie https://www.storage.pionier.net.pl/usetheservice/index.html [on-line] [dostęp 13 październik 2012 r.] oraz serwisie wiki przedsięwzięcia https://www.storage.pionier.net.pl/wiki/index.php/Howto [on-line] [dostęp 23 października 2012 r.].

[38] Zob. M. Brzeźniak, *NDS2 – Secure storage, sharing and publishing of data in the NDS*. [on-line] [dostęp 13 października 2012 r.]. Dostępny w World Wide Web: http://www.terena.org/activities/tf-storage/ws13/slides/270912-nds2.pdf

[39] Więcej na temat repozytorium można dowiedzieć się ze strony projektu. [on-line] [dostęp 10 października 2012 r.]. Dostępny w World Wide Web: http://rcin.org.pl/dlibra/text?id=aboutRCIN

on-line[40]. dArceo to prototyp usług długoterminowego archiwizowania dużych wolumenów danych, będący odpowiedzią na problemy dotyczące zabezpieczenia dziedzictwa kulturowego, jakim niewątpliwie są zasoby cyfrowych bibliotek. Przedsięwzięcie wchodzi w skład projektu SYNAT (System Nauki i Techniki), będącego częścią Program Strategicznego Narodowego Centrum Badań i Rozwoju. Obecnie dArceo jest testowane w ramach wymienionego wcześniej Repozytorium Cyfrowego Instytutów Naukowych[41].

Mówiąc o archiwizacji dziedzictwa kultury, trudno nie wspomnieć Narodowego Archiwum Cyfrowego (NAC) uruchomionego w marcu 2008 r., następcy Archiwum Dokumentacji Mechanicznej. Głównymi zadaniami tej instytucji są: „gromadzenie, przechowywanie i udostępnianie materiałów cyfrowych, naukowe opracowywanie i udostępnianie fotografii, nagrań dźwiękowych oraz filmów, digitalizacja tradycyjnych materiałów archiwalnych oraz udostępnianie informacji o zbiorach archiwalnych i materiałów archiwalnych, także on-line"[42]. Warto zaznaczyć, że na dzień dzisiejszy w Archiwum zdeponowano prawie 190 tys. pozycji (fotografie i filmy) oraz zabezpieczono strony internetowe pond 40 instytucji (głównie archiwa i muzea)[43].

W dokumencie *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*, przygotowanym na zlecenie Ministerstwa Kultury i Dziedzictwa Narodowego, czytamy: „niearchiwizowane systematycznie i bezpiecznie obiekty cyfrowe, dokumenty elektroniczne dostępne on-line oraz strony internetowe są bezpowrotnie stracone dla obecnych i przyszłych pokoleń", a „skoordynowana budowa i długookresowa ochrona zasobów cyfrowych jest istotnym czynnikiem warunkującym rozwój ekonomiczny Polski oraz ważnym elementem tworzenia społeczeństwa informacyjnego"[44]. Ochrona zasobów cyfrowych, obok masowej digitalizacji tradycyjnych dokumentów, jest więc dzisiaj również dla instytucji centralnych kluczo-

[40] Zob. *Federacja Bibliotek Cyfrowych* [on-line] [dostęp 10 października 2012 r.]. Dostępny w World Wide Web: http://fbc.pionier.net.pl/owoc

[41] Więcej na ten temat można przeczytać w artykule *dArceo services: advancing long-term preservation* autorstwa Adama Dudczaka, Michała Kozaka, Cezarego Mazurka, Tomasza Parkoły, Szymona Pietrzaka i Marcina Werli. Zob. również prezentację multimedialną na ten sam temat [on-line] [dostęp 10 października 2012 r.]. Dostępny w World Wide Web: https://www.conftool.net/or2012/index.php?page=browseSessions&form_session=7

[42] Zob. *Narodowe Archiwum Cyfrowe*. Strona główna projektu. [on-line] [dostęp 30 września 2012 r.]. Dostępny w World Wide Web: http://www.nac.gov.pl/

[43] Dane na dzień 3 października 2012 r.

[44] *Program digitalizacji dóbr kultury oraz gromadzenia, przechowywania i udostępniania obiektów cyfrowych w Polsce 2009–2020*, Warszawa 2009. [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://www.kongreskultury.pl/library/File/RaportDigitalizacja/Program%20digitalizacji%202009-2020.pdf

wym elementem funkcjonowania polskiej kultury, dostępu do informacji i treści należących do dziedzictwa narodowego.

Na podstawie charakterystyki powyższych projektów służących długoterminowej archiwizacji nasuwa się wniosek, że również w Polsce poszukuje się uniwersalnego modelu ochrony bibliotecznych zasobów cyfrowych, który obejmowałby swoim zasięgiem zarówno dokumenty dostępne on-line (strony internetowe, zawartość repozytoriów, biblioteczne dokumenty życia społecznego, zawartość katalogów itp.), dokumenty zapisane na nośnikach cyfrowych (np. CD, DVD), jak i dane, które są jeszcze często przechowywane na medium starszej generacji (np. dyskietki).

Z drugiej strony, konieczna wydaje się edukacja pracowników polskich bibliotek i studentów studiów bibliologicznych w odniesieniu do narzędzi oraz sposobów ochrony zasobów cyfrowych. Jeszcze do niedawna niektórzy z nich mylili proces digitalizacji i tworzenie cyfrowych repozytoriów (bibliotek) z przedsięwzięciami mającymi na celu długoterminową archiwizację elektronicznych dokumentów[45]. Oczywiście mówiąc o digitalizacji, mamy też na myśli przechowywanie i archiwizowanie plików macierzystych (w tzw. formatach archiwalnych) cyfrowych publikacji z zachowaniem możliwie najwyższych parametrów i jak najwierniejszego odwzorowania oryginału[46]. Jednak długoterminowa ochrona zasobów sieciowych to bardziej skomplikowany proces, który bez zastosowania specjalistycznych rozwiązań bądź ogólnokrajowego systemu powszechnej archiwizacji jest praktycznie niemożliwy. Te specjalistyczne rozwiązania ułatwiać miałyby przy tym nie tylko samo składowanie i przechowywanie danych, ale także łatwy i bezpośredni do nich dostęp.

Mówiąc o długoterminowej archiwizacji cyfrowych zasobów bibliotecznych, mamy na myśli zarówno centralne projekty ich ochrony, jak i lokalne archiwa organizowane na podstawie ogólnie przyjętych zasad, ale też indywidualnych rozwiązań poszczególnych bibliotek. Również w dwóch ostatnich przypadkach konieczne jest zastosowanie specjalistycznego oprogramowania do składowania, przechowywania i udostępniania zarchiwizowanych dokumentów. Pomocny tutaj może okazać się system Rosetta firmy Ex Libris, którego wartość i użyteczność od stycznia 2009 r. (udostępnienie komercyjnej wersji 1.0) doceniły instytucje informacji ulokowane w różnych miejscach świata (np. archiwa i biblioteki w Nowej Zelandii, Niemczech, USA, Singapurze czy Anglii). Dodajmy, że nazwa systemu

[45] J. Januszko-Szakiel, *Długoterminowa archiwizacja zasobów cyfrowych w świadomości pracowników polskich bibliotek*. (prezentacja) [on-line] [dostęp 3 października 2012 r.]. Dostępny w World Wide Web: http://lib.psnc.pl/dlibra/docmetadata?id=222&from=publication&showContent=true

[46] W. M. Kolasa, *Formaty dokumentów w bibliotekach cyfrowych*, [w:] *Biblioteki cyfrowe*…, s. 405–407.

nawiązuje do odkrytego w 1799 r. zabytku staroegipskiego i greckiego piśmiennictwa, pochodzącego z 196 r. p.n.e. tzw. kamienia z Rosetty.

W czerwcu 2007 r. firma Ex Libris podpisała porozumienie o współpracy z Biblioteką Narodową Nowej Zelandii. Główny cel tej umowy to utworzenie cyfrowego archiwum oraz specjalistycznego narzędzia umożliwiającego efektywne zabezpieczenie elektronicznych dokumentów w ramach nowozelandzkiego programu Narodowego Archiwum Dziedzictwa Cyfrowego (National Digital Heritage Programme). Warto dodać, że program ten służyć miał rozwiązaniu problemu przechowywania, ochrony, ale także zapewnieniu – przyszłym pokoleniom – dostępu do stale powiększających się nowozelandzkich zasobów cyfrowych. Efektem tej współpracy było wprawdzie stworzenie systemu, udostępnianego jako produkt komercyjny, który jednakże oferowałby profesjonalne i sprawdzone sposoby archiwizacji danych. Ważnym etapem całego przedsięwzięcia było także poddanie go ocenie międzynarodowej grupy recenzentów składającej się z bibliotekarzy i pracowników nauki. Grupa ta stanowić miała ważne źródło informacji o zgodności wypracowywanych rozwiązań z ówczesnymi trendami i standardami w zakresie ochrony zasobów cyfrowych[47].

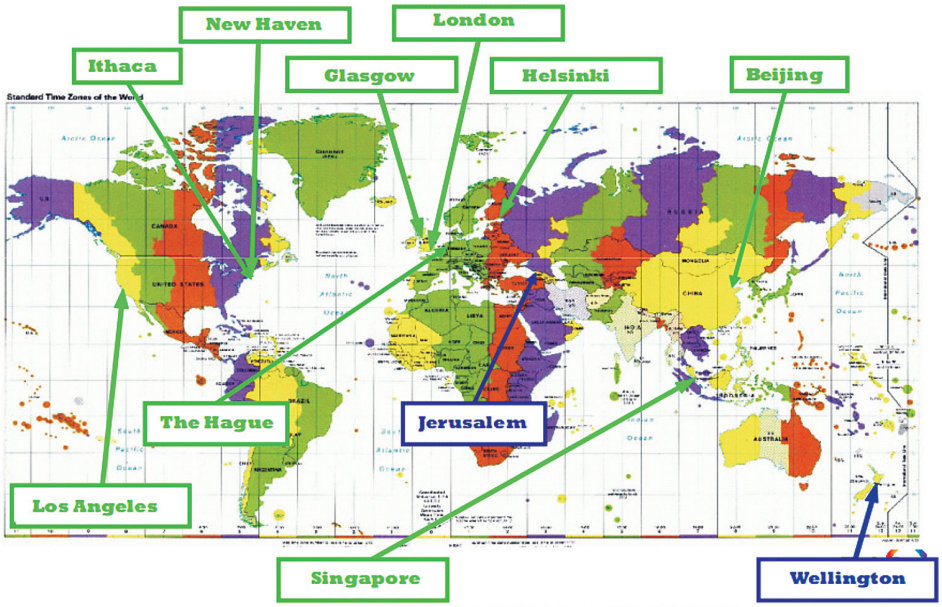
Prezentacji gotowego systemu, który pierwotnie nosił nazwę The Digital Preservation System, dokonano w listopadzie 2008 r. Warto zaznaczyć, że był on wówczas pierwszym tego rodzaju rozwiązaniem pozwalającym największym (naukowym, uniwersyteckim czy narodowym) bibliotekom organizować proces ochrony cyfrowych zasobów, jednocześnie zapewniając ich integralność i użyteczność w długoterminowej perspektywie[48].

Zanim jednak skupimy się na charakterystyce poszczególnych modułów oprogramowania, zwróćmy uwagę na fakt, że system jest zgodny z modelem organizacji i funkcjonowania archiwów elektronicznych OAIS (Open Archival Information System)[49]. Fakt ten wydaje się kluczowy z kilku względów. OAIS

[47] Więcej na temat tego projektu można przeczytać na stronie WWW firmy Ex Libris. Zob. *Ex Libris and the National Library of New Zealand Sign Agreement to Develop a Product to Preserve National Cultural Heritage* [on-line] [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://www.exlibrisgroup.com/default.asp?catid={CDDC8842-4FC9-4E7E-93B9-9B3E1E8E2D74}&details_type=1&itemid={FBBB3441-7544-4060-8453-BDC3FBA2FF49} Zob. także *Digital preservation AT the National Library of New Zealand*. [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://www.exlibrisgroup.com/files/CaseStudy/SunPreservationandNLNZ.pdf; S. Knight, *Early learnings from the National Library of New Zealand's. National Digital Heritage Archive project*. [on-line] [dostęp 18 października 2012 r.]. Dostępny w World Wide Web: www.emeraldinsight.com/0033-0337.htm

[48] *The Ex Libris Digital Preservation System Goes Live at the National Library of New Zealand*. [on-line] [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://www.exlibrisgroup.com/default.asp?catid={CDDC8842-4FC9-4E7E-93B9-9B3E1E8E2D74}&details_type=1&itemid={20151FEE-55D0-41BA-8626-86D034826982}

[49] Model ten został opracowany przez Consultative Committee for Space Data Systems (CCSDS) w odniesieniu do problemów związanych z archiwizacją i wymianą danych elektronicz-



Rys. 2. International peer review group – miejsca, z których pochodziły poszczególne osoby wchodzące w skład grupy. Źródło grafiki: http://tinyurl.com/9mds6at

charakteryzuje system otwartej informacji archiwalnej (na technicznym i organizacyjnym poziomie). Jest definiowany jako „organizacja składająca się z ludzi i stosownej infrastruktury, której starania skoncentrowane są na długoterminowym przechowywaniu, zabezpieczaniu i udostępnianiu obiektów elektronicznych wyznaczonej grupie użytkowników”[50]. Niezależnie od faktu, że model ten został opracowany dla konkretnego typu danych, jest on uznawany za uniwersalny sposób organizowania i funkcjonowania elektronicznych archiwów. Jest także wykorzystywany do gromadzenia, organizacji i udostępniania różnych typów dokumentów oraz stosowany w archiwizacji cyfrowych zasobów bibliotek, muzeów czy właśnie archiwów[51]. Charakteryzując przedsięwzięcia przygotowywane zgodnie z OAIS, A. Januszko-Szakiel zwraca także uwagę na ważne rozróżnienie pomiędzy danymi cyfrowymi (Data Object) a obiektami informacyjnymi (Information Object). Przy czym te drugie oprócz danych cyfrowych zawierają

nych dotyczących badań nad przestrzenią kosmiczną. W 2003 r. model OAIS został zaakceptowany przez ISO jako standard w zakresie archiwizacji dokumentów cyfrowych (*ISO 14721:2003 – Space data and information transfer systems – Open archival information system – Reference model*).

[50] A. Januszko-Szakiel, *Open Archival Information System – standard w zakresie archiwizacji publikacji elektronicznych*, „Przegląd Biblioteczny” 2005, z. 3, s. 342.

[51] *Ibid.*, s. 343.

też narzędzia umożliwiające przetworzenie surowych danych tak, aby użytkownik był w stanie się z nimi zapoznać[52].

Matthias Töwe, w trakcie spotkania użytkowników oprogramowania Ex Libris (IGeLU) w Zurichu we wrześniu 2012 r., jako jedną z największych zalet systemu Rosetta wymienia właśnie zgodność z OAIS[53]. Warto dodać, że system ten pomaga w przestrzeganiu obowiązujących zasad, które archiwum typu OAIS musi spełniać. Hubert Wajs w publikacji *Dokument elektroniczny jako obiekt archiwalny* wymienia te zasady. Przytoczmy w tym miejscu przynajmniej kilka z nich, te, które z punktu widzenia problemu długoterminowej archiwizacji bibliotecznych zasobów sieciowych wydają się najistotniejsze:

– Określenie zakresu, w jakim wykorzystywał informację pierwotny użytkownik (twórca);

– Zapewnienie, że informacja została przechowana w takiej formie, która (niezależnie od technologii) jest zrozumiała dla pierwotnych użytkowników; informacja w kontekście – zapewnienie czytelności także tego kontekstu, tak aby informacja była czytelna i zrozumiała dla przyszłych pokoleń; oraz określenie wynikających z powyższego metadanych;

– Opracowanie dokumentacji i procedur koniecznych do właściwego wypełnienia obowiązku przechowania, czyli mechanizmy przechowania[54].

Na podstawie powyższych słów warto pokusić się o stwierdzenie, że trwały dostęp do archiwizowanych zasobów odgrywa ważną, jeśli nie najważniejszą rolę. Z drugiej strony, trudno dzisiaj przewidzieć, jaką wartość naukową, kulturową czy informacyjną będą miały chronione zasoby w przyszłości. Jednakże udostępniając je w ramach określonego kontekstu, stwarzamy przyszłym czytelnikom (użytkownikom) szansę na prawidłowe odczytanie przechowywanych informacji. Już dzisiaj można zaryzykować tezę, że ten kontekst będzie miał decydujące znaczenie w odniesieniu do wartości poszczególnych danych.

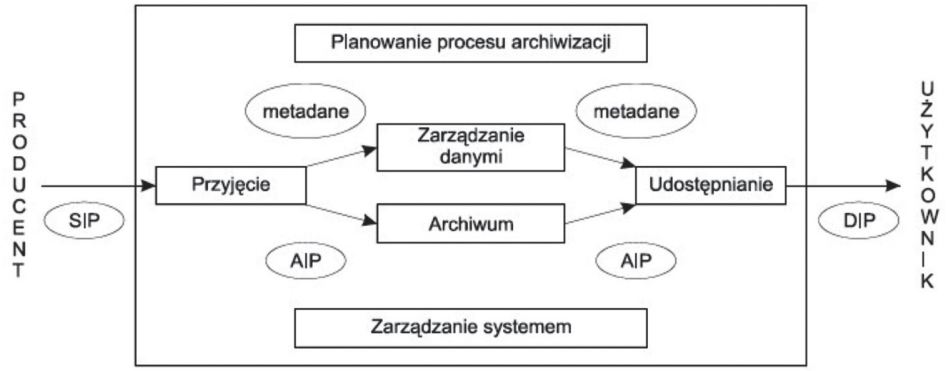
W archiwum typu OAIS przechowywane są obiekty informacyjne, które wchodzą w skład tzw. pakietów informacyjnych (Information Packages). Pakiet składa się z dwóch kluczowych części: kontenera informacyjnego (dane i narzędzia), a także informacji odnoszących się do samego procesu przechowywania (np. ochrony integralności i autentyczności). W ramach OAIS wyróżniamy Sumbission Information Package (SIP), Archival Information Package (AIP) i Dissemination Information Package (DIP). Rozróżnienie to dotyczy pakietów, które są przechowywane w archiwum (AIP), dostarczane do archiwum przez pro-

[52] *Ibid.*

[53] Zob. *Matthias Towe, Rosetta at ETH Zurich: routes into the digital archive*. [on-line] [dostęp 15 października 2012 r.]. Dostępny w World Wide Web: http://igelu.org/wp-content/uploads/2012/09/IGeLU_Toewe_20120912.pptx

[54] H. Wajs, *Dokument elektroniczny jako obiekt archiwalny*. [on-line] [dostęp 11 października 2012 r.]. Dostępny w World Wide Web: http://postdiploma.pl/do.pdf

ducenta (SIP) oraz wysyłane z archiwum i udostępnione użytkownikowi (DIP)[55]. Oczywiście na wymienione pakiety składają się zarówno dokumenty czy obiekty cyfrowe, jak i powiązane z nimi metadane. Poniżej zaprezentowano graficzny schemat funkcjonowania archiwum w modelu OAIS:



Rys. 3. Graficzny schemat funkcjonowania archiwum w modelu OAIS. Źródło obrazu: J. Januszko-Szakiel, *Open Archival Information*…, s. 347

System Rosetta spełnia warunki archiwum tworzonego zgodnie z OAIS. Jest również zgodny z modelem Trusted Digital Repository (TDR), dzięki czemu zapewnia infrastrukturę i technologię niezbędną do ochrony i udostępniania autentycznych zasobów cyfrowych[56]. Dodatkowo, dla lepszego przystosowania poszczególnych funkcjonalności systemu do oczekiwań jego odbiorców, stworzono strategiczny program mający na celu współpracę z obecnymi i przyszłymi użytkownikami Rosetty w ramach tzw. Ex Libris Rosetta Charter Program[57].

System dostarcza gotowych rozwiązań odnoszących się do problemów gromadzenia, organizacji, przechowywania, ochrony i upowszechniania różnych obiektów cyfrowych. Składa się trzech modułów głównych: archiwizacji i ochrony, przechowywania i depozytu oraz zarządzania poszczególnymi kolekcjami dokumentów. Obejmuje ładowanie danych, walidację, dostęp i trwałe repozytorium. Jest także rozwiązaniem typu end-to-end (E2ES), co oznacza, że Ex Libris jest

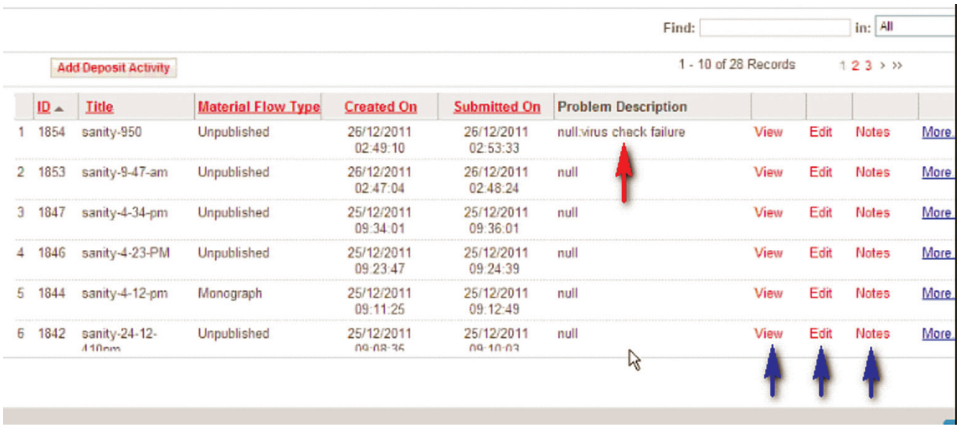
[55] Zob. A. Januszko-Szakiel, *Open Archival Information*…, s. 346.

[56] Trusted Digital Repository jest definiowany jako model zarządzania cyfrowym repozytorium, którego głównym celem jest zapewnienie niezawodnego, długoterminowego dostępu do zasobów cyfrowych tak dzisiaj, jak i w przyszłości. Zob. *Trusted Digital repositories. Attributes and Responsibilities. An RLG-OCLC Report*. Mountain View, CA. May 2002 [on-line] [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://www.oclc.org/resources/research/activities/trustedrep/repositories.pdf

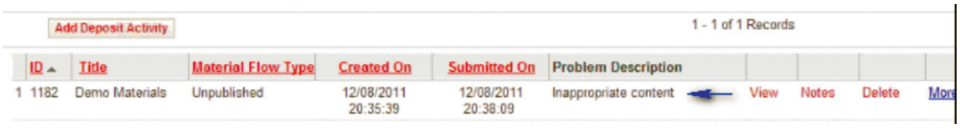
[57] *Ex Libris Group Releases New Digital Preservation System* [on-line] [dostęp 17 października 2012 r.]. Dostępny w World Wide Web: http://newsbreaks.infotoday.com/Digest/Ex-Libris-Group-Releases-New-Digital-Preservation-System-52203.asp

nie tylko dostawcą samego oprogramowania, ale współpracuje z daną instytucją, pomaga wdrażać poszczególne etapy projektu archiwizacji danych – od koncepcji do pełnej realizacji przedsięwzięcia. Również sam proces dodawania i archiwizacji poszczególnych obiektów został tak zaplanowany, aby w jak najskuteczniejszy i najefektywniejszy sposób pozwolić użytkownikom na organizację i zarządzanie zamieszczanymi materiałami. W pierwszym etapie obiekt jest umieszczany w specjalnym magazynie depozytowym (wsadowe wgrywanie materiałów), gdzie wchodzące w jego skład zasoby zostają poddane weryfikacji (sprawdzana jest poprawność i pochodzenie danych – kiedy, jak i przez kogo element został utworzony). Istnieje tutaj także możliwość określania hierarchiczności dla modułu depozytu.

Następnie zasoby są sprawdzane pod kątem poprawności i pełności zarówno opisowych, jak i technicznych metadanych. Celem jest kompleksowe przygotowanie obiektu do późniejszego długoterminowego przechowywania (jego przetwarzania, zarządzania, archiwizowania, ochrony i wykorzystania poszczególnych danych). Ważnym elementem jest także określenie prawnych podstaw udostępniania dokumentów (np. w ramach ich późniejszej prezentacji użytkownikowi).

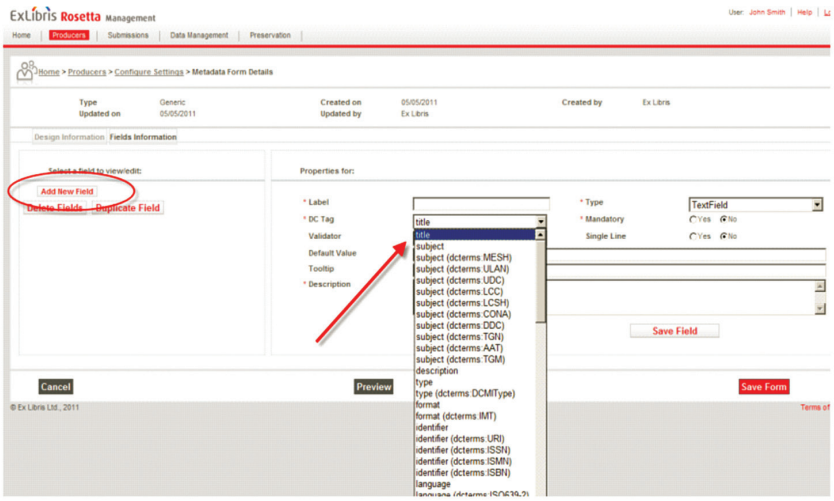


Rys. 4. System Rosetta. Moduł depozytu. Informacje o zweryfikowanych obiektach. Źródło obrazu: materiały wewnętrzne Ex Libris



Rys. 5. System Rosetta. Moduł depozytu. Informacja o zweryfikowanym obiekcie. Źródło obrazu: materiały wewnętrzne Ex Libris

Oczywiście istnieje także możliwość definiowania formularza metadanych (np. dodawania pól i podpól). Ta opcja pozwala nam na dość szczegółowe określenie, jakie informacje będą wchodziły w skład opisu obiektu.



Rys. 6. Definiowanie formularza metadanych. Źródło obrazu: E. M. Corrado, *Rosetta in Academic Library*[58]

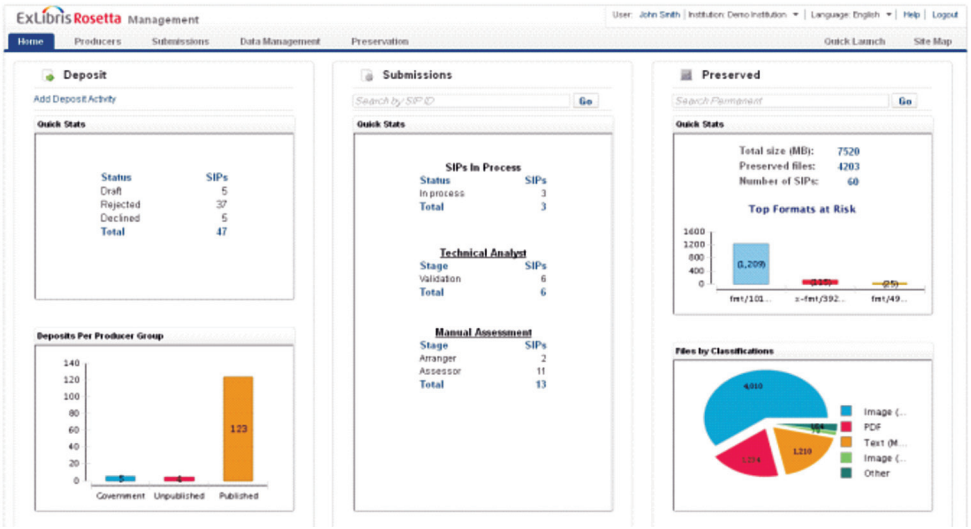
Wreszcie zasoby są zapisywane w możliwie najtrwalszym formacie. Ich długoterminowa archiwizacja i ochrona oznacza, że poszczególne obiekty są stale poddawane ocenie w celu zapewnienia ich ciągłej użyteczności. Mamy więc do czynienia z oprogramowaniem oferującym pełne bezpieczeństwo danych, sprawdzanie ich poprawności i integralności[59].

Na poniższej grafice zaprezentowany został widok panelu administracyjnego wchodzącego w skład modułu zarządzania archiwizowanymi zasobami. Mamy tutaj do dyspozycji zarówno dane dotyczące formatów, w jakich przechowywane są poszczególne dokumenty, wielkości czy ilości zarchiwizowanych zasobów, statusu pakietów informacyjnych dostarczanych od tzw. producentów do archiwum (SIP), możliwość edycji metadanych, dostęp do raportów, jak i opcję wyszukiwania i podglądu przechowywanych obiektów. Jak podkreśla Nil Shenwinter, podczas zorganizowanego przez firmę Ex Libris webinarium interfejs modułu za-

[58] E. M. Corrado, *Rosetta in Academic Library*. [on-line] [dostęp 18 października 2012 r.] Dostępny w World Wide Web: http://igelu.org/wp-content/uploads/2011/09/corrado_rosetta_igelu2011.pdf

[59] *Ex Libris Group Announces the General Release of its Digital Preservation System*. [on-line] [dostęp 17 październiak 2012 r.]. Dostępny w World Wide Web: http://www.exlibrisgroup.com/default.asp?catid={CDDC8842-4FC9-4E7E-93B9-9B3E1E8E2D74}&details_type=1&itemid={9B1F2C8A-3B03-459F-A2B4-4425A4D79689}

rządzania został tak przygotowany, aby dostarczał wszystkich informacji o realizowanym projekcie w jednym miejscu[60].



Rys. 7. Rosetta. Panel administracyjny. Moduł zarządzania. Źródło obrazu: materiały wewnętrzne Ex Libris

Moduł zarządzania pozwala nam także na organizację dokumentów wchodzących w skład poszczególnych kolekcji (ich aktualizację czy usunięcie), a także tworzenie podkolekcji. Ten element, umożliwiając składowanie, przechowywanie, odnajdywanie i udostępnianie danych, nosi cechy systemu zarządzania danymi cyfrowymi (tzw. Digital Asset Management).

Jeżeli chodzi o moduł archiwizacji cyfrowych zasobów, Rosetta oferuje kompleksowe rozwiązanie ochrony i dostępu zarówno do dokumentów zdigitalizowanych, jak i tzw. *born digital*. Co ważne, biblioteka formatów systemu obejmuje ponad 800 standardów, w których mogą być zapisane poszczególne typy dokumentów, jak również informacje o tych formatach. Dzięki temu modułowi Rosetta zapewnia długoterminową ochronę i dostęp do zarchiwizowanych obiektów informacyjnych. Pomaga też w strategicznym planowaniu działań odnoszących się do migracji danych do nowszych formatów. Warto dodać, że w skład biblioteki formatów wchodzą także narzędzia umożliwiające dostęp do poszczególnych dokumentów (np. Gimp, Irfan View).

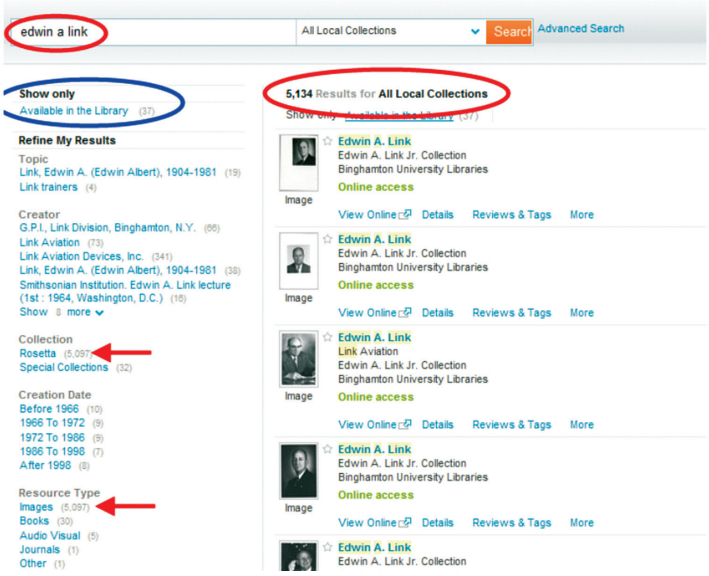
Wśród najważniejszych cech Rosetty niewątpliwie należałoby wymienić jego skalowalność. Termin ten oznacza, że budowa systemu ma formę rozproszo-

[60] Ciekawostką jest fakt, że w odniesieniu do tego panelu w dokumentacji Rosetty pojawia się angielski termin *dashboard*, czyli deska rozdzielcza.

nej architektury (możliwość funkcjonowania w złożonej konfiguracji serwerów). Magazynowanie i zarządzanie dotyczy cyfrowych kolekcji dowolnej wielkości. System jest także elastyczny i może być dostosowany do działalności dowolnej instytucji oraz podstaw prawnych, w ramach których ona funkcjonuje. Ważna jest też możliwość integracji z innymi informatycznymi narzędziami, w tym produktami firmy Ex Libris, wykorzystywanymi do wyszukiwania (np. Primo) i udostępniania dokumentów. Możliwe jest również publikowanie katalogu tego systemu za pośrednictwem wyszukiwarki Google. Rosetta może stać się swoistym „modułem" kompleksowego systemu zarządzania, przechowywania i udostępniania zarchiwizowanych cyfrowych zasobów[61].

System oferuje rozwiązanie kluczowych problemów dotyczących długoterminowej archiwizacji elektronicznych danych.

- Dostępność (*availability*) w tym przypadku oznacza wykorzystanie w ramach systemu modułu depozytu oraz opcji umożliwiających użytkownikom bezpośredni dostęp do archiwizowanych zasobów. Możemy się też posłużyć wyszukiwaniem zaawansowanym odnoszącym się do niemal każdego elementu wchodzącego w skład opisu dokumentów. Termin „dostępność" dotyczy także stałej kontroli obiektów.

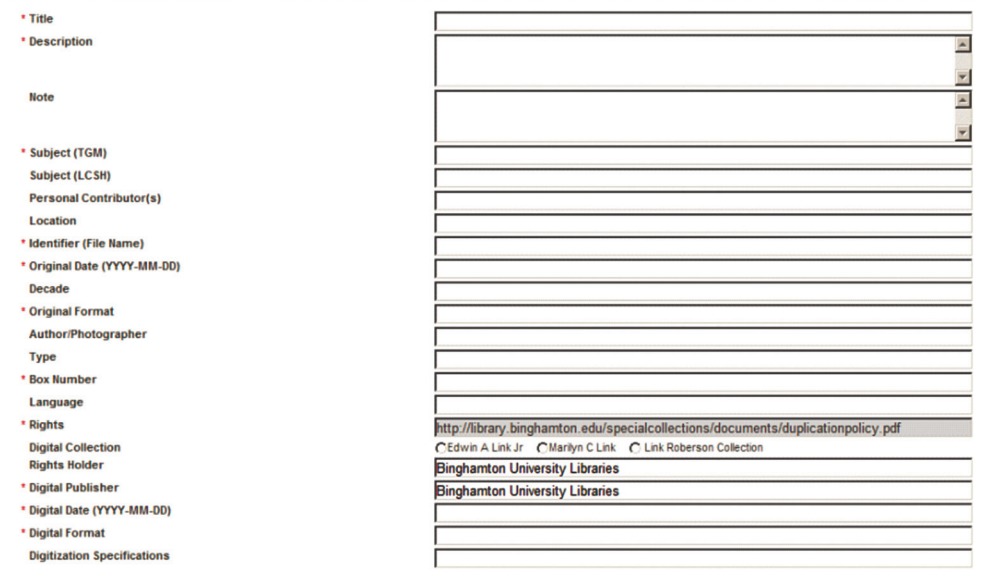


Rys. 8. Dostęp do archiwizowanych zasobów za pośrednictwem Primo (Tzw. OPAC-a 2.0)[62]
Źródło obrazu: materiały wewnętrzne Ex Libris

[61] Zob. charakterystykę Rosetty na stronie WWW firmy Aleph Polska. [on-line] [dostęp 18 października 2012 r.]. Dostępny w World Wide Web: http://www.aleph.pl/produkty/rosetta/

[62] E. M. Corrado, *op. cit*. Więcej na temat OPAC-ów 2.0 można przeczytać w mojej książce *Biblioteka w środowisku społecznościowego Internetu. Biblioteka 2.0*, Warszawa 2012, s. 104–115.

• Tożsamość (*identity*) dotyczy szczegółowości opisu danego obiektu, który pozwoli w przyszłości znaleźć poszukiwany dokument. Ostatnio udostępniona wersja Rosetty (3.0) charakteryzuje się nowymi możliwościami w zakresie wyszukiwania (filtrowanie, sortowanie, wyszukiwanie fasetowe, podświetlenie trafień itp.) czy rozbudowanym modułem zarządzania poszczególnymi kolekcjami.



Rys. 9. Formularz metadanych w bibliotece uniwersyteckiej w Binghamton. Źródło obrazu: E. M. Corrado, *Rosetta in Academic Library*[63]

• Możliwość prawidłowego odczytania i zrozumienia zasobu (*understandability*) jest związana z zachowaniem dodatkowych informacji, pozwalających na zapoznanie się z dokumentem w sposób zrozumiały dla jego przyszłego odbiorcy. Chodzi tutaj także o właściwą strukturę i metadane, jak również o wsparcie dla odmiennych materiałów i formatów w ramach jednego modelu danych – bazujący na standardzie PREMIS (Preservation Metadata: Implementation Strategies)[64] i METS, opis w Dublin Core itp.[65].

[63] E. M. Corrado, *Rosetta in Academic Library*. [on-line] [dostęp 18 października 2012 r.] Dostępny w World Wide Web: http://igelu.org/wp-content/uploads/2011/09/corrado_rosetta_igelu2011.pdf

[64] Zob. P. Caplan, *Understanding PREMIS. Washington 2009*. [on-line] [dostęp 18 października 2012 r.]. Dostępny w World Wide Web: http://www.loc.gov/standards/premis/understanding-premis.pdf

[65] Zob. Y. Levi, *A Preservation System. Chalanges and practical experience*. [on-line] [dostęp 18 października 2012 r.]. Dostęp w World Wide Web: http://tinyurl.com/9npelwp

• Trwałość, stałość (*fixity*), czyli zabezpieczenie obiektów cyfrowych przed nieautoryzowanymi zmianami (celowymi bądź niezamierzonymi). Ważnym aspektem jest też możliwość przechowywania pełnych obiektów informacyjnych (dane, metadane, narzędzia).

• Identyfikacja i odczyt dokumentu zapisanego w danym formacie (*viability* i *renderablity*). Rosetta dostarcza rozwiązań umożliwiających definiowanie zasad, jakich narzędzi użyć dla każdego z formatów (ważna rola biblioteki formatów).

• Wreszcie charakteryzowana wcześniej autentyczność (*authenticity*), czyli zachowanie oryginalności danych (możliwa weryfikowalność źródła danych z zawartością obiektu informacyjnego)[66].

Ex Libris, jako firma od wielu lat dostarczająca oprogramowanie dla bibliotek i archiwów z niemal całego świata, dzięki wykorzystaniu najnowszych technologii, sprawdzonych przez międzynarodową grupę recenzentów, zaproponowała zaawansowany produkt zapewniający długoterminową ochronę, ale także bezpośredni dostęp do archiwizowanych danych. Wchodzące w skład Rosetty moduły oferują efektywną i funkcjonalną archiwizację, ochronę, przechowywanie oraz zarządzanie poszczególnymi kolekcjami dokumentów. Co więcej, paleta możliwości wykorzystania Rosetty jest o wiele szersza niż pojedynczych systemów umożliwiających tworzenie repozytoriów i cyfrowych bibliotek (np. Fedora, DSpace, Omeka), archiwizacji dokumentów (np. systemy iRODS, Lockss) czy ich ochrony (np. systemy Tessella czy Dura Space). Rosetta charakteryzuje się kompleksowym rozwiązaniem problemów dotyczących zarówno tworzenia cyfrowych kolekcji, ich archiwizacji, jak i długoterminowej ochrony[67].

A LONG-TERM ARCHIVING OF DIGITAL RESOURCES

A long term protection of digital resources is a fundamental problem for a developing informational society, which is characterised by an increasing pace of generating new informational resources. Also the libraries, as archiving institutions, making the traditional as well as the digital documents available, are obliged to consistent actions towards the archiving of the electronic resources, but also their meta-data. This article aims at analysing the phenomenon of a long-term archiving of libraries' digital resources, with using the professional Rosetta software by Ex Libris, more widely unknown to the Polish environment.

[66] Opracowano na podstawie A. Kaschte, *Digital Preservation System ExLibris Rosetta*. [online] [dostęp 17 października 2012r.]. Dostępny w World Wide Web: http://indico.cern.ch/getFile.py/access?contribId=39&resId=1&materialId=slides&confId=48321

[67] N. Sherwinter, *Rosetta Product Overview*. Wystąpienie w ramach webinarium Ex Libris w maju 2012 r. Materiały wewnętrzne Ex Libris.