

## Infrastruktura bibliotek cyfrowych w sieci PIONIER

CEZARY MAZUREK, MACIEJ STROIŃSKI, MARCIN WERLA, JAN WĘGLARZ

*Poznańskie Centrum Superkomputerowo-Sieciowe*  
{mazurek | stroins | mwerla | weglarz}@man.poznan.pl

### Streszczenie

W artykule przedstawiono infrastrukturę bibliotek cyfrowych funkcjonującą w ramach sieci PIONIER. Jest ona wynikiem prac rozwojowych, rozpoczętych, w ramach ogłoszonego w 2001 roku, programu PIONIER oraz prowadzonych od 1999 roku w Poznańskim Centrum Superkomputerowo-Sieciowym prac nad rozwojem oprogramowania do zarządzania zawartością bibliotek cyfrowych dLibra. Uruchomienie we wrześniu 2002 roku Wielkopolskiej Biblioteki Cyfrowej stało się istotnym krokiem w kierunku budowy hierarchicznej, rozproszonej infrastruktury, udostępniającej obecnie ponad 150 tysięcy obiektów cyfrowych w blisko 40 instalacjach bibliotek cyfrowych. Omawiana infrastruktura posiada także jeden wspólny punkt dostępu: usługę Federacji Bibliotek Cyfrowych, która zapewnia zestaw narzędzi ułatwiających wykorzystanie tych zasobów oraz ich popularyzację. Taki model realizacji usług infrastruktury bibliotek cyfrowych pozwala także na włączenie ich zasobów do europejskiej biblioteki cyfrowej Europeana.

**Słowa kluczowe:** biblioteki cyfrowe, usługi atomowe, infrastruktura informatyczna, federacje usług

### Wstęp

Biblioteki cyfrowe są jednym z istotnych składników infrastruktury informatycznej nauki, który w ostatnich latach w Polsce rozwijał się niezwykle dynamicznie. W roku 2001, w ramach programu „PIONIER. Polski Internet Optyczny – Zaawansowane Aplikacje, Usługi i Technologie dla Społeczeństwa Informacyjnego”, zdefiniowana została potrzeba budowy zaawansowanych usług sieciowych jako niezbędnego składnika infrastruktury informatycznej dla nauki. Jednym z elementów tej infrastruktury był system sieciowych usług bibliotek cyfrowych. W wyniku tych działań, w ramach istniejącej sieci PIONIER, funkcjonuje nowoczesna, dobrze zorganizowana i rozproszona infrastruktura bibliotek cyfrowych. Treści dostępne w środowisku bibliotek cyfrowych są przetwarzane przez narzędzia i usługi stanowiące składnik tej infrastruktury i określają funkcjonalność biblioteki cyfrowej, a także zakres odbiorców (tj. użytkowników). W tym kontekście możemy już mówić o globalnej infrastrukturze bibliotek cyfrowych jako o zasobie wartościowym jednocześnie pod względem naukowym i społecznym.

Budowa sieci rozproszonych bibliotek cyfrowych i repozytoriów w Polsce rozpoczęła się uruchomieniem we wrześniu 2002 roku pierwszej w Polsce biblioteki cyfrowej – Wielkopolskiej Biblioteki Cyfrowej (WBC). Ta, obecnie największa w Polsce, biblioteka cyfrowa (ponad 60 tys. obiektów cyfrowych) stała się wzorem dla kolejnych tego typu inicjatyw. Od momentu jej publicznego udostępnienia w Polsce uruchomiono kilkadziesiąt kolejnych bibliotek cyfrowych i repozytoriów, które aktualnie gromadzą ponad 150 tysięcy obiektów cyfrowych i są obszarem współpracy kilkuset instytucji.

Z ogłoszonej przez Komisję Europejską w 2005 roku inicjatywy *i2010 Digital Libraries* [1] oraz z planów budowy Europejskiej Biblioteki Cyfrowej wynikała konieczność wirtualnego połączenia

polskich, rozproszonych bibliotek cyfrowych w jedną spójną platformę bazującą na automatycznej wymianie informacji pomiędzy poszczególnymi bibliotekami. Podjęcie tych prac pozwoliło w efekcie na uzyskanie szeregu dodatkowych funkcji i usług infrastruktury bibliotek cyfrowych. Należą do nich m.in. mechanizmy wymiany metadanych pomiędzy rozproszonymi bibliotekami cyfrowymi i opartych na tym rozwiązaniu zaawansowanych funkcji, takich jak wyszukiwanie rozproszone czy dynamiczne wirtualne kolekcje, a także uruchomiony w czerwcu 2007 roku serwis Federacji Bibliotek Cyfrowych (FBC).

Federacja Bibliotek Cyfrowych, z wykorzystaniem otwartych standardów i protokołów, udostępnia nowe, zaawansowane funkcje, takie jak: przeszukiwanie zasobów rozproszonych bibliotek cyfrowych, tworzenie trwałych referencji do poszczególnych obiektów cyfrowych oraz dostęp do raportów zawierających plany digitalizacji poszczególnych bibliotek. Dzięki FBC możliwe jest również automatyczne wykrywanie duplikatów pomiędzy bibliotekami cyfrowymi oraz zapobieganie dublowaniu digitalizacji już na etapie jej planowania w rozproszonych ośrodkach. Najnowsze funkcje FBC związane są z sieciowym profilem użytkownika, pozwalającym czytelnikom polskich bibliotek cyfrowych korzystać z jednej tożsamości we wszystkich bibliotekach Federacji.

### **Infrastruktura Bibliotek Cyfrowych w Polsce**

Uruchomienie w 2002 roku Wielkopolskiej Biblioteki Cyfrowej zapoczątkowało proces tworzenia kolejnych bibliotek cyfrowych. W 2004 roku powstała Biblioteka Cyfrowa Politechniki Wrocławskiej, przekształcona później w Dolnośląską Bibliotekę Cyfrową. Poza WBC, największe obecnie biblioteki cyfrowe to: Kujawsko-Pomorska Biblioteka Cyfrowa (ok. 18 000 publikacji), Biblioteka Cyfrowa Uniwersytetu Wrocławskiego (ok. 12 000 publikacji) oraz Śląska Biblioteka Cyfrowa, Małopolska Biblioteka Cyfrowa i Cyfrowa Biblioteka Narodowa POLONA (każda po ok. 6 000 publikacji). Obecnie w sieci PIONIER funkcjonuje 38 bibliotek cyfrowych, których zasoby dostarczane są przez ok. 200 instytucji zaangażowanych w zarządzanie ich zawartością (ryc. 1). Niemal wszystkie te biblioteki cyfrowe funkcjonują z wykorzystaniem oprogramowania dLibra, rozwijanego w Poznańskim Centrum Superkomputerowo-Sieciowym pakietu narzędzi i usług do zarządzania zawartością bibliotek cyfrowych.

Ze względu na charakter organizacyjny biblioteki cyfrowe w Polsce możemy podzielić na trzy grupy:

- instytucjonalne biblioteki cyfrowe (obecnie 24 instalacje),
- regionalne biblioteki cyfrowe (obecnie 13 instalacji),
- konsorcjalne biblioteki cyfrowe (obecnie 1 instalacja).

Instytucjonalne biblioteki cyfrowe prowadzone są zazwyczaj przez jedną instytucję. Umieszczone w niej obiekty cyfrowe związane są najczęściej z aktualną działalnością i historią tej instytucji. Zarówno opieka merytoryczna nad biblioteką oraz techniczna nad systemem informatycznym spoczywa na barkach tej konkretnej instytucji. Przykładami mogą tu być biblioteki cyfrowe wielu uczelni wyższych (m.in. Uniwersytetu Wrocławskiego i Warszawskiego czy Politechniki Warszawskiej) oraz innych instytucji państwowych (Biblioteka Cyfrowa Centralnego Ośrodka Doskonalenia Nauczycieli, <http://bc.codn.edu.pl/>).

Regionalne biblioteki cyfrowe tworzone są najczęściej z myślą o digitalizacji dziedzictwa kulturowego i naukowego danego regionu oraz dziedzictwa narodowego. Współtworzone są przez różno-



Ryc. 1. Biblioteki cyfrowe w sieci PIONIER

rodne instytucje regionu, takie jak biblioteki uczelni publicznych i prywatnych, biblioteki publiczne, muzea, archiwa, organizacje kultury, a nawet osoby prywatne. Z tego względu charakter, a przez to także zasoby takich bibliotek cyfrowych, są bardzo różnorodne, choć w większości związane tematycznie z określonym regionem kraju. Działania techniczno-merytoryczne polegają na modelu, w którym regionalne, naukowe centrum sieciowe czy obliczeniowe opiekuje się systemem biblioteki cyfrowej od strony technicznej. Współtworzące bibliotekę cyfrową, regionalne instytucje kultury i nauki (w tym koordynująca wszystkie działania) odpowiadają za merytoryczny rozwój biblioteki.

Do grupy konsorcjalnych bibliotek cyfrowych zaliczyć można Wirtualny Księgozbiór Federacji Bibliotek Kościelnych FIDES (<http://digital.fides.org.pl/>), którego zawartość jest współtworzona przez wszystkie biblioteki wchodzące w skład Federacji (obecnie jest ich ponad 80).

Tak duża liczba bibliotek cyfrowych oraz treści cyfrowych dostępnych on-line stała się podstawą do rozpoczęcia prac nad utworzeniem jednolitej infrastruktury działającej w skali całego kraju. W dążeniach do ujednoczenia infrastruktury pomógł projekt badawczy nr 3 T11C 023 30 Ministerstwa Nauki i Szkolnictwa Wyższego *Mechanizmy usług atomowych dla rozproszonych bibliotek*

cyfrowych, zrealizowany przez PCSS. Dzięki pracom badawczo-rozwojowym uruchomiony w sieci Polski Internet Optyczny – PIONIER połączył biblioteki cyfrowe za pomocą protokołu OAI-PMH [2]. Protokół służący do wymiany metadanych umożliwił zrealizowanie mechanizmu wyszukiwania rozproszonego, pozwalającego na przeszukiwanie opisów zasobów cyfrowych wszystkich bibliotek opartych na oprogramowaniu dLibra z poziomu ich stron WWW [3]. Wykorzystanie protokołu OAI-PMH spowodowało również powstanie automatycznie nadawanych, unikalnych w skali światowej, identyfikatorów udostępnianych obiektów cyfrowych.

Kolejny element infrastruktury bibliotek cyfrowych w Polsce to wyszukiwarka Federacja Bibliotek Cyfrowych (FBC) udostępniona w czerwcu 2007 roku pod adresem <http://fbc.pionier.net.pl/>.

### **Federacja Bibliotek Cyfrowych w sieci PIONIER**

U podstaw funkcjonowania Federacji Bibliotek Cyfrowych (FBC) znajduje się zbiór atomowych usług dla rozproszonych bibliotek cyfrowych, których koncepcję opracowano w ramach wspomnianego powyżej projektu badawczego [4-6]. Główne cechy FBC to: łatwy dostęp do zasobów polskich bibliotek cyfrowych, zwiększenie widoczności tych zasobów w Internecie oraz udostępnianie użytkownikom Federacji nowych, zaawansowanych usług wykorzystujących te zasoby. Fundamentem dla realizacji powyższych funkcji są dwie wyspecjalizowane instancje usługi wyszukiwania. Pierwsza z nich odpowiada za gromadzenie i przeszukiwanie metadanych związanych z już istniejącymi obiektami cyfrowymi, druga natomiast pełni analogiczne zadania dla metadanych opisujących obiekty dopiero przeznaczone do digitalizacji. Taki podział umożliwi m.in. realizację automatycznego wsparcia dla koordynacji digitalizacji w Polsce. Obydwie te usługi okresowo (obecnie raz na dobę) pobierają z poszczególnych bibliotek cyfrowych metadane dodane lub zmienione od momentu ostatniej synchronizacji oraz informacje na temat rekordów metadanych powiązanych z usuniętymi obiektami. Pobieranie metadanych realizowane jest za pomocą protokołu OAI-PMH, a podstawowym schematem opisu obiektów cyfrowych jest Dublin Core [7]. Poza usługami wyszukiwania w FBC wykorzystywane są również usługi związane z sieciowym profilem użytkownika, pozwalające na korzystanie z jednego konta użytkownika w całej sieci bibliotek cyfrowych [8].

Na bazie usług atomowych funkcjonują komponenty poziomu aplikacji FBC. Komponenty te odpowiedzialne są za realizację interfejsu użytkownika FBC oraz interfejsów komunikacyjnych dla systemów zewnętrznych. Obecnie, poza wyszukiwarkami internetowymi typu Google, dane pobierane z FBC widoczne są między innymi w portalu slawistycznym Slavistik-Portal.de. Ponadto funkcjonalność przeszukiwania polskich zasobów cyfrowych za pomocą FBC dostępna jest na stronach wielu instytucji, takich jak: Biblioteka Kórnicka PAN, Poznańska Fundacja Bibliotek Naukowych czy Biblioteka Główna UMCS. Co więcej, metadane zebrane przez FBC wkrótce przekazywane będą również do Europejskiej Biblioteki Cyfrowej Europeana.

Od czerwca 2008 roku rozpoczęto prace związane z projektem EDLocal. Jest on trzyletnim projektem typu Best Practice Network, finansowanym w ramach programu eContentPlus, a jego podstawowym celem jest udostępnienie zasobów cyfrowych zgromadzonych w lokalnych i regionalnych bibliotekach cyfrowych poprzez Europejską Bibliotekę Cyfrową Europeana. Poznańskie Centrum Superkomputerowo-Sieciowe jest jednym z partnerów tego projektu i pełni w nim m.in. rolę regionalnego koordynatora na Polskę. W ramach tego projektu zasoby wszystkich bibliotek cyfrowych współpracujących w ramach Federacji Bibliotek Cyfrowych będą widoczne również w Europeanie.

## Podsumowanie

W niniejszym artykule przedstawiono infrastrukturę bibliotek cyfrowych w sieci PIONIER. Do najważniejszych jej cech należą: silnie rozproszona architektura, jeden punkt dostępu do zasobów i usług, otwarte standardy komunikacji, skoordynowane zarządzanie zasobami oraz wykorzystanie innych elementów infrastruktury naukowej (tzw. usług wspólnych). Infrastruktura ta obejmuje obecnie blisko 40 instalacji systemów bibliotek cyfrowych, w których zgromadzono ok. 150 tysięcy obiektów cyfrowych. Dostęp do wszystkich zasobów możliwy jest dzięki Federacji Bibliotek Cyfrowych. Uruchomiona w roku 2007 i dostępna pod adresem [fbc.pionier.net.pl](http://fbc.pionier.net.pl) usługa ma na celu ułatwienie wykorzystania zasobów polskich bibliotek cyfrowych i repozytoriów, zwiększenie widoczności tych zasobów w Internecie oraz udostępnienie ich użytkownikom nowych, zaawansowanych możliwości bazujących na tych zasobach. Dynamiczny rozwój Federacji, stały wzrost wykorzystania jej funkcji i zasobów przez użytkowników, włączenie FBC do oficjalnej dystrybucji przeglądarki Firefox i wreszcie wykorzystanie jej przy budowie Europejskiej Biblioteki Cyfrowej Europeana pokazują, że jej uruchomienie stanowiło właściwy krok w rozwoju polskiej infrastruktury bibliotek cyfrowych w sieci naukowej.

Zawartość bibliotek cyfrowych dostępna w ramach opisanej infrastruktury stanowi przedmiot badań wielu naukowców nauk humanistycznych na świecie, stąd też jest istotnym czynnikiem umożliwiającym zwiększenie dynamiki prowadzonych w tym zakresie badań. Istotne jest przy tym, iż polski model realizacji infrastruktury bibliotek cyfrowych pozwala na synergiczny rozwój zasobów cyfrowych oraz usług, które są niezbędne do pracy w ramach nowoczesnego warsztatu naukowego osadzonego w infrastrukturze informatycznej nauki. Ponadto, istotną właściwością tego składnika jest również społeczne znaczenie zgromadzonych zasobów informacyjnych, a w przyszłości powszechne wykorzystanie tych narzędzi i usług.

## Piśmiennictwo

- [1] i2010: Digital Libraries Initiative  
[http://ec.europa.eu/information\\_society/activities/digital\\_libraries/index\\_en.htm](http://ec.europa.eu/information_society/activities/digital_libraries/index_en.htm)
- [2] C. Lagoze et al., 2002. *The Open Archives Initiative Protocol for Metadata Harvesting*.  
<http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [3] Mazurek C., Stroiński M., Węglarz J., Werla M. *Metadata harvesting in regional digital libraries in PIONIER Network*. *Campus-Wide Information Systems* 2006; 23(4): 241-253.
- [4] Mazurek C., Parkoła T., Werla M. *Atomowe usługi w środowisku rozproszonych bibliotek cyfrowych*. II Krajowa Konferencja Naukowa „Technologie przetwarzania danych”, Poznań, 24-26 września 2007 r. Wydawnictwo Politechniki Poznańskiej, Poznań 2007: 446-457.
- [5] Kosiedowski M., Mazurek C., Stroiński M., Werla M., Wolski M. *Federating digital library services for advanced applications in science and education*. *Comput. Meth. Sci. Technol.* 2007; 14.
- [6] Dudczak A., Mazurek C., Werla M. *RESTful Atomic Services for Distributed Digital Libraries*. 2008. 1st International Conference on Information Technology, 18-21 maja 2008 r., Gdańsk 2008: 267-270.
- [7] Norma ISO 15836: 2003. Schemat metadanych Dublin Core w wersji 1.1.  
<http://www.iso.org/iso/search.htm?published=on&qt=15836:2003>
- [8] Dudczak A., Heliński M., Mazurek C., Mielnicki M., Werla M. *Extending the Shibboleth Identity Management Model with a Networked User Profile*. 1st International Conference on Information Technology, 18-21 maja 2008 r., Gdańsk 2008: 179-182.

