

Agnieszka Lewandowska, Marcin Werla
{jagna,mwerla}@man.poznan.pl
Poznańskie Centrum Superkomputerowo-Sieciowe

Przyłączanie zasobów polskich bibliotek cyfrowych do Europeany

Biblioteki cyfrowe w Polsce – modele organizacyjne

- Główne modele organizacyjne
 - **Biblioteki regionalne**
 - Tworzone i utrzymywane przez instytucje kultury i nauki z danego regionu
 - Zawierają przede wszystkim obiekty związane z regionem, jego historią i kulturą, ale również akademickie materiały edukacyjne i narodowe dziedzictwo kulturowe
 - **Instytucjonalne biblioteki cyfrowe**
 - Tworzone i utrzymywane przez pojedyncze instytucje (np. uczelnie)
 - Zawierają obiekty związane z obecną działalnością tych instytucji i ich historią, ale również inne obiekty, w tym narodowe i regionalne dziedzictwo kulturowe

Biblioteki cyfrowe w Polsce – stan obecny

Liczba obiektów cyfrowych

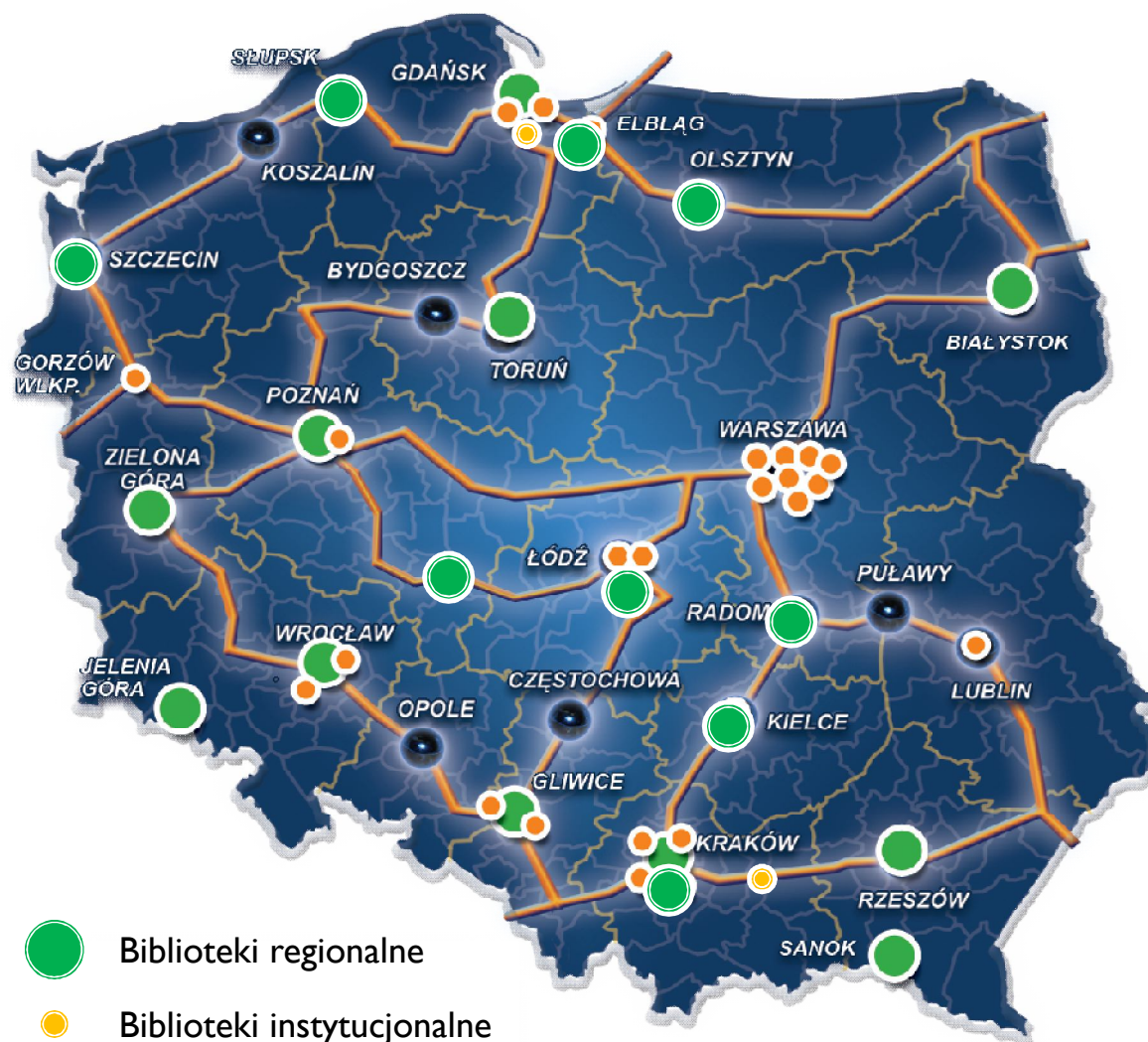
- ✓ 312 000 obiektów łącznie
- ✓ 9 000 nowych obiektów w miesiącu (średnio)

Liczba aktywnych bibliotek cyfrowych:

- ✓ 19 regionalnych
- ✓ 22 instytucjonalnych

+ **nieustannie powstające nowe biblioteki cyfrowe**

Tę sieć tworzy kilkaset współpracujących instytucji kultury i nauki



Federacja Bibliotek Cyfrowych

- Dostępny publicznie od czerwca 2007 roku serwis internetowy
- Gromadzi opisy (agreguje metadane) obiektów cyfrowych znajdujących się w polskich bibliotekach cyfrowych
- Technologiczną podstawą działania są otwarte standardy komunikacji i wymiany danych
- Dane są aktualizowane raz na dobę, w nocy
- Serwis jest utrzymywany i prowadzony przez PCSS

<http://fbc.pionier.net.pl/>

Federacja Bibliotek Cyfrowych

- Podstawowe funkcje
 - Przeszukiwanie opisów dostępnych obiektów
 - Plany digitalizacji
 - Przeszukiwanie
 - Raport
 - Interfejs programistyczny (API) wspierające zapobieganie wielokrotnej digitalizacji tych samych obiektów
 - Lokalizacja obiektów cyfrowych na podstawie ich identyfikatorów (OAI Id)
 - Baza danych polskich bibliotek cyfrowych
 - <http://fbc.pionier.net.pl/owoc/libs-map>
 - Statystyki i raporty
 - Dodatki ułatwiające promowanie FBC i przyłączonych bibliotek
 - **Udostępnianie zgromadzonych metadanych innym serwisom**

EuropeanaLocal

- EuropeanaLocal

- Projekt w ramach programu eContent*Plus*
- Czas trwania: 3 lata (czerwiec 2008 – maj 2011)
- Typ projektu: Sieć najlepszych praktyk
- <http://europeanalocal.eu>
- PCSS jest w ramach projektu regionalnym koordynatorem na Polskę oraz należy do rady zarządzającej projektu



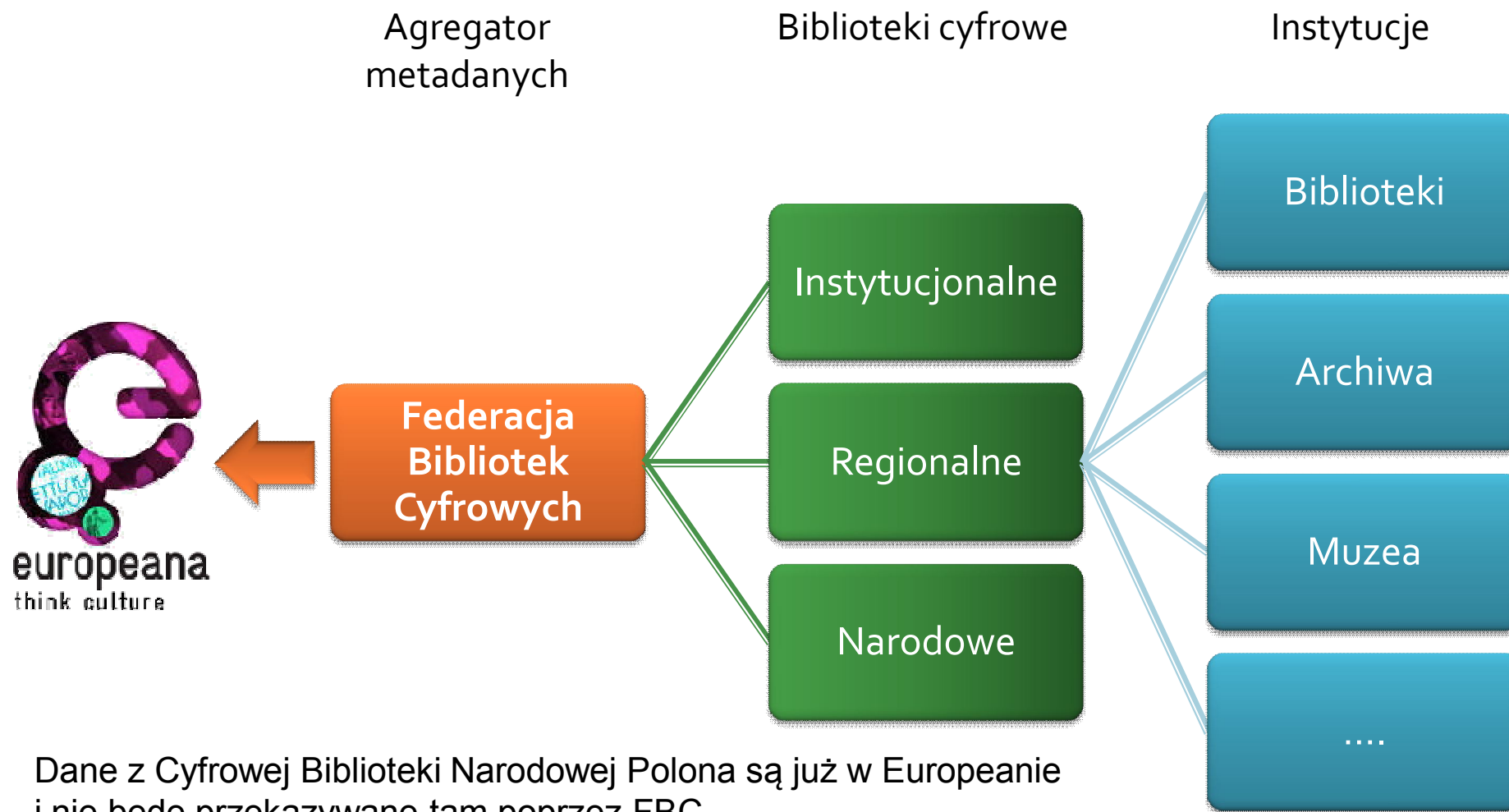
EuropeanaLocal

- Główne cele projektu EuropeanaLocal



- Zwiększenie możliwości zautomatyzowanego użycia treści cyfrowych z lokalnych i regionalnych bibliotek cyfrowych w zewnętrznych usługach sieciowych
- Doprowadzenie do stworzenia sieci regionalnych repozytoriów cyfrowych będących w stanie komunikować się z Europeaną

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany



Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Gromadzimy metadane obiektów, które powinny być widoczne również w Europeanie
 - Jest to proces w pełni zautomatyzowany
 - W większości wypadków podstawą jest protokół OAI-PMH
 - W szczególnych sytuacjach realizujemy połączenie w inny sposób (np. Polska Biblioteka Internetowa)
 - Obecnie pobierane jest tylko 15 atrybutów Dublin Core Simple
 - Prace nad nowym schematem metadanych zostały zapoczątkowane we wrześniu 2009
 - Nowy schemat ma być z założenia listą rekomendowanych elementów DC Terms i ewentualnie dodatkowych rozszerzeń
 - Np. miejsce wydania publikacji

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Próbujemy oczyścić pobierane metadane, znormalizować je i wzbogacić
 - Ze względu na dużą liczbę obiektów wszystkie te operacje muszą być zautomatyzowane
 - Na poziomie Federacji budujemy słowniki wszystkich występujących wartości
 - Osobno dla każdego elementu schematu metadanych
 - Osobno dla każdego języka
 - Różnice w opisach pochodzących z różnych instytucji utrudniają użytkownikom przeszukiwanie takiego połączonego zbioru obiektów

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Próbujemy oczyścić pobierane metadane, znormalizować je i wzbogacić
 - Normalizujemy wartości atrybutu dc:language do standardu RFC 4646
 - Np.: polski, polish, pol, Polnisch => pl
 - Generujemy wartość atrybutu europeana:type na podstawie dc:type lub dc:format
 - Dozwolone wartości: TEXT, IMAGE, AUDIO, VIDEO
 - Dodajemy nazwę i adres WWW biblioteki cyfrowej do pola dc:source
 - Przykład modyfikacji metadanych pod kątem interfejsu użytkownika

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Próbujemy oczyścić pobierane metadane, znormalizować je i wzbogacić
 - Problem opisów wielojęzycznych

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Fragment przykładowego opisu przed uporządkowaniem:
 - dc:publisher (pl) : Warszawa
 - dc:publisher (pl) : Gebethner i Wolff
 - dc:publisher (en) : Warszawa
 - dc:publisher (en) : Gebethner i Wolff
 - dc:language (pl) : pol
 - dc:language (en) : pol
 - dc:type (pl) : książka
 - dc:type (en) : book
 - dc:subject (pl) : Darwin, Charles (1809-1882) -- krytyka i interpretacja.
 - dc:subject (pl) : Słowacki, Juliusz (1809-1849) -- krytyka i interpretacja.
 - dc:subject (en) : Darwin, Charles, 1809-1882.
 - dc:subject (en) : Słowacki, Juliusz, 1809-1849 -- Criticism and interpretation.

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Fragment przykładowego opisu po uporządkowaniu:
 - dc:publisher (pl) : Warszawa
 - dc:publisher (pl) : Gebethner i Wolff
 - dc:publisher (en) : Warszawa
 - dc:publisher (en) : Gebethner i Wolff
 - dc:language (pl) : pol
 - dc:language (en) : pol
 - dc:type (pl) : książka
 - dc:type (en) : book
 - dc:subject (pl) : Darwin, Charles (1809-1882) -- krytyka i interpretacja.
 - dc:subject (pl) : Słowacki, Juliusz (1809-1849) -- krytyka i interpretacja.
 - dc:subject (en) : Darwin, Charles, 1809-1882.
 - dc:subject (en) : Słowacki, Juliusz, 1809-1849 -- Criticism and interpretation.

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Fragment przykładowego opisu po uporządkowaniu:
 - dc:publisher : Warszawa
 - dc:publisher : Gebethner i Wolff
 - dc:language : pol
 - dc:type (pl) : książka
 - dc:type (en) : book
 - dc:subject (pl) : Darwin, Charles (1809-1882) -- krytyka i interpretacja.
 - dc:subject (pl) : Słowacki, Juliusz (1809-1849) -- krytyka i interpretacja.
 - dc:subject (en) : Darwin, Charles, 1809-1882.
 - dc:subject (en) : Słowacki, Juliusz, 1809-1849 -- Criticism and interpretation.

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Pracujemy nad przekazaniem danych do Europeany:
 - 1 etap: Sprawdzenie metadanych w tzw. Europeana Content Checker
 - Jest to kopia serwisu Europeana
 - Można tam samodzielnie wgrać metadane obiektów cyfrowych i zobaczyć jak wyglądają one w interfejsie Europeany
 - 24 listopada 2009 umieściliśmy w serwisie Content Checker metadane 254 478 obiektów
 - TEXT: 239 551
 - IMAGE: 14 766
 - AUDIO: 126
 - VIDEO: 35
 - Etap zakończony powodzeniem

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Pracujemy nad przekazaniem danych do Europeany:
 - 2 etap: Sprawdzenie technicznych możliwości pobierania informacji z agregatora przy pomocy protokołu OAI-PMH
 - Test został przeprowadzony 27 listopada 2009
 - Wynik: „*Everything is fine for harvesting*”

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Pracujemy nad przekazaniem danych do Europeany:
 - 3 etap: Finalne ustalenia i pobranie metadanych przez Europeanę
 - W toku...
 - <http://www.europeana.eu/portal/brief-doc.html?start=1&view=table&query=COUNTRY:poland>

Federacja Bibliotek Cyfrowych jako agregator metadanych dla Europeany

- Pracujemy nad przekazaniem danych do Europeany:
 - Zalecenia ze strony Europeany:
 - Wprowadzenie miniatur obiektów cyfrowych na poziom agregacji i dalej do Europeany
 - Poprawa intuicyjności korzystania z publikacji w formacie DjVu dla „zwykłych” użytkowników

Podsumowanie

- Mamy w Polsce kilkadziesiąt bibliotek cyfrowych gromadzących i udostępniających obiekty cyfrowe pochodzące z kilkuset instytucji kultury i nauki
- Agregujemy metadane tych obiektów i dajemy pojedynczy punkt dostępu do nich w postaci Federacji Bibliotek Cyfrowych
- Współpraca z Europeaną (i nie tylko z Europeaną) wymaga czyszczenia, normalizowania i wzbogacania metadanych
 - Tylko część z pożądanых efektów da się osiągnąć w pełni automatycznie
- Metadane z polskich bibliotek cyfrowych powinny być widoczne w Europeanie w ciągu kilku najbliższych dni
 - A gdy to się stanie...

Agnieszka Lewandowska, Marcin Werla
{jagna,mwerla}@man.poznan.pl
Poznańskie Centrum Superkomputerowo-Sieciowe

Przyłączanie zasobów polskich bibliotek cyfrowych do Europeany

Dziękuję za uwagę!