



# Biblioteki cyfrowe w środowisku sieciowym

Marcin Werla

Poznańskie Centrum Superkomputerowo-Sieciowe

---

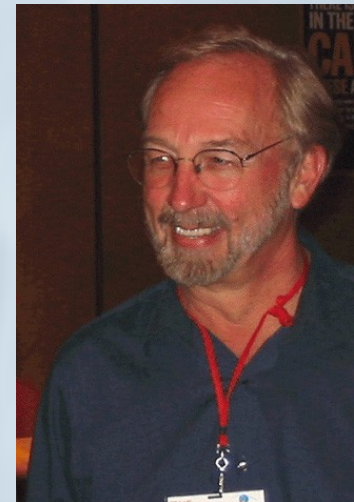
IV Warsztaty „Biblioteki cyfrowe”

Poznań, 2007







# The Network Is The Computer™

*Dopiero sieć to komputer*

-- John Gage



# Biblioteki cyfrowe dLibra

- dLibra
-  2002
  -  2004
  -  2005
  -  2006
  -  do 11.2007
  -  od 11.2007



# Protokół OAI-PMH

- **Open Archives Initiative**
  - **Protocol for Metadata Harvesting**
- <http://www.openarchives.org>



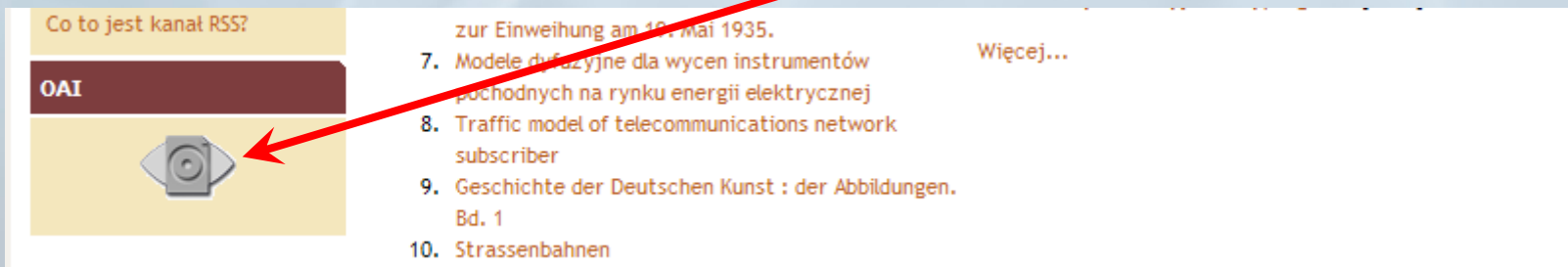
# Protokół OAI-PMH

- Otwarty
- Łatwy do rozszerzenia o nowe funkcje
- Oparty na popularnych standardach: HTTP, XML, DublinCore
- Łatwy do zaimplementowania w różnych technologiach
- Pozwala na wymianę informacji pomiędzy oprogramowaniem różnych producentów

# Protokół OAI-PMH

- Oprogramowanie dLibra wspiera protokół OAI-PMH w jego najnowszej wersji (2.0) ze wszystkimi elementami opcjonalnymi

<http://<biblioteka.cyfrowa>/oai-pmh-repository.xml>



Co to jest kanał RSS?

OAI

zur Einweihung am 19. Mai 1935.

7. Modele dyfuzyjne dla wycen instrumentów pochodnych na rynku energii elektrycznej

8. Traffic model of telecommunications network subscriber

9. Geschichte der Deutschen Kunst : der Abbildungen. Bd. 1

10. Strassenbahnen

Więcej...

# Polska sieć bibliotek cyfrowych

- Podstawa - komunikacja bibliotek cyfrowych przy pomocy protokołu OAI-PMH
- Odbywa się okresowo – raz na dobę (w nocy)
- Przesyłane są opisy nowych i zmienionych publikacji oraz informacje o usuniętych publikacjach

# Polska sieć bibliotek cyfrowych

- Dostępne usługi
  - Wyszukiwanie rozproszone
  - Wirtualne kolekcje rozproszonych zasobów
  - Nadawane automatycznie, globalnie unikalne identyfikatory obiektów cyfrowych



# Federacja Bibliotek Cyfrowych

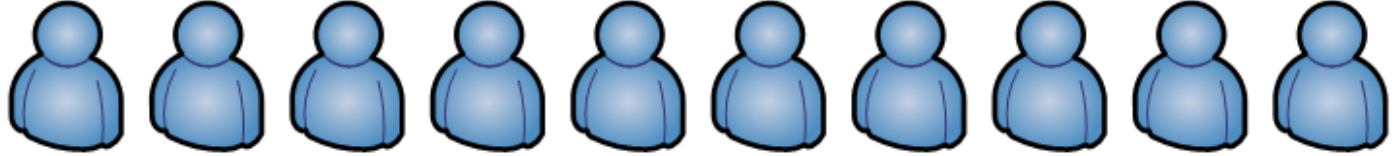
- Portal dostępowy dla usług opartych na bibliotekach cyfrowych dostępnych w sieci PIONIER
- Dostępny dla końcowych użytkowników, jak i dla innych usług
- Komunikacja z repozytoriami oparta na otwartym protokole OAI-PMH i schemacie metadanych Dublin Core
- Wykorzystanie otwartych standardów pozwala na dołączenie dowolnych repozytoriów spełniających te standardy
- W ramach prac nad FBC znaleziono i poprawiono błąd w specyfikacji protokołu OAI-PMH



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

# Federacja Bibliotek Cyfrowych

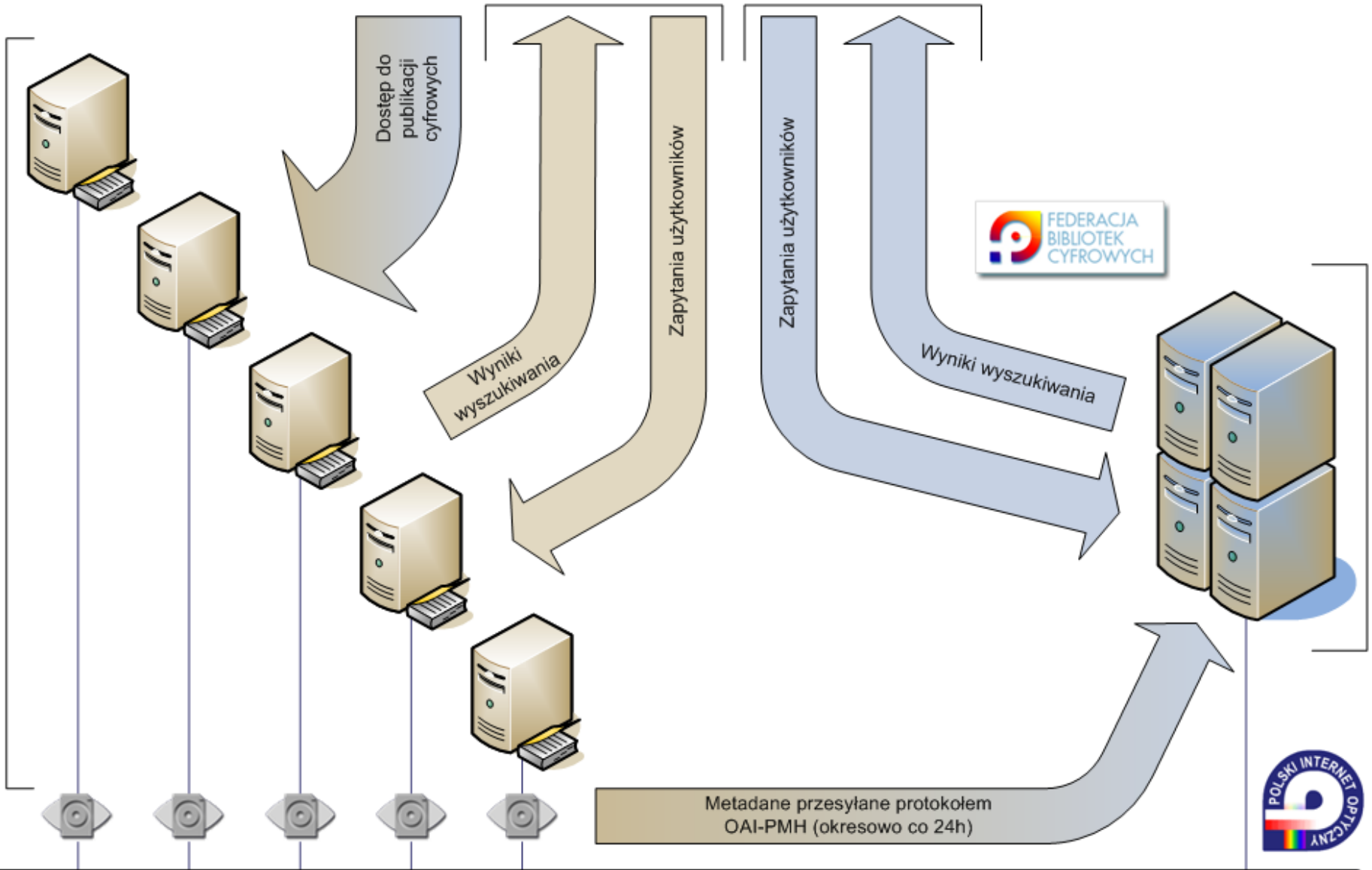
- Wyszukiwanie obiektów cyfrowych
  - Okresowe pobieranie i indeksowanie metadanych ze zdalnych repozytoriów
  - Jeden interfejs wyszukiwawczy dla wszystkich repozytoriów OAI-PMH w Polsce
  - Funkcjonalność dostępna również w postaci mini-wyszukiwarki i dodatków do przeglądarek WWW
  - Największe podobne projekty na świecie: OAIster.org (wszystkie repozytoria), ScientificCommons.org (repozytoria naukowe)  
Polskie zasoby są w nich również widoczne



Rozproszone biblioteki cyfrowe z interfejsami OAI-PMH

Wyszukiwanie lokalne

Wyszukiwanie rozproszone



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

Polski Internet Optyczny PIONIER

# Federacja Bibliotek Cyfrowych

- Rozpoznawanie unikalnych identyfikatorów obiektów cyfrowych
    - W pełni automatyczne nadawanie i rozpoznawanie identyfikatorów oraz aktualizacja metadanych i referencji
      - Identyfikator w danej bibliotece cyfrowej:  
**8711**
      - Unikalny identyfikator OAI:  
**oai:www.wbc.poznan.pl:8711**
      - Trwała referencja:  
<http://fbc.pionier.net.pl/id/oai:www.wbc.poznan.pl:8711>
    - Podobne projekty na świecie:
      - DOI (system komercyjny)
      - Persistent-Identifier.de (system dla repozytoriów niemieckich)
- Obydwa systemy wymagają „ręcznego” tworzenia identyfikatorów i aktualizacji danych

# Federacja Bibliotek Cyfrowych

- Wyszukiwanie obiektów przeznaczonych do digitalizacji
  - Okresowe pobieranie i indeksowanie metadanych obiektów przeznaczonych do digitalizacji
  - Wyniesienie cyklu życia obiektu cyfrowego z pojedynczego repozytorium na poziom platformy bibliotek cyfrowych
  - Przechowywanie planów digitalizacji to unikalna cecha oprogramowania dLibra, opracowana we współpracy ze środowiskiem bibliotekarzy
  - Udostępnianie planów digitalizacji przy pomocy protokołu OAI-PMH dostępne od wersji 3.0 systemu dLibra
  - Wykorzystanie otwartych protokołów umożliwia dodanie innych systemów w przyszłości
  - Podobne projekty na świecie: **nie znane**

# Federacja Bibliotek Cyfrowych

- Automatyczne wykrywanie duplikatów
  - Automatyczna analiza i porównywanie opisów wszystkich obiektów cyfrowych zgromadzonych w FBC
    - obecnie około 100 000 obiektów
  - Wynikiem jest raport pokazujący potencjalne duplikaty powstałe w różnych bibliotekach cyfrowych
    - obecnie wykryto około 70 różnych tytułów
  - Podobne projekty na świecie:
    - eksperymenty prowadzone na Old Dominion University w USA

# Wyzwania

- Automatyczna koordynacja digitalizacji
- Identyfikacja i trwałe referencje do obiektów cyfrowych
- Rozproszone uwierzytelnianie użytkowników
- Spójność metadanych w rozproszonych bibliotekach cyfrowych

# Wyzwania

- Automatyczna koordynacja digitalizacji
  - Dlaczego powstają duplikaty?
    - Brak wymiany informacji między bibliotekami
    - Spójność i kompletność posiadanych kolekcji
    - Zła jakość obiektów już zdigitalizowanych
    - ...
  - Czy duplikaty to coś złego?
  - Co możemy zrobić, aby zmniejszyć liczbę duplikatów do minimum?



# Wyzwania

- Identyfikacja i trwałe referencje do obiektów cyfrowych
  - Jak zapewnić unikalne i trwałe identyfikatory i referencje do obiektów cyfrowych?
  - Czy da się to zrobić „bezpłatnie”?
  - Jak przekonać użytkowników do korzystania z właściwych referencji?

# Wyzwania

- Rozproszone uwierzytelnianie użytkowników
  - Po co wykorzystywać istniejące bazy użytkowników w bibliotekach cyfrowych?
  - Jak to zrobić?
  - Jakie nowe możliwości daje taki system uwierzytelniania?

# Wyzwania

- Spójność metadanych w rozproszonych bibliotekach cyfrowych
  - Czy obecne opisy bibliograficznych są takie złe?
  - Jak możemy zapewnić spójność opisów?
    - NUKAT? ePoradnik? Konsorcjum?
  - Co nam da ujednoczenie opisów?

# POZNAŃSKIE CENTRUM SUPERKOMPUTEROWO-SIECIOWE



To wszystko (a nawet więcej ;-)  
na drugim dniu warsztatów.

Serdecznie zapraszamy!!!

---

Toruń, Biblioteka UMK  
6 grudnia 2007 r.