

Rozproszone biblioteki cyfrowe

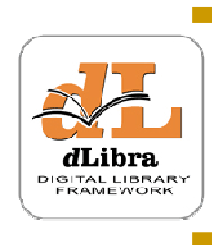


Marcin Werla

Poznańskie Centrum Superkomputerowo-Sieciowe

mwerla@man.poznan.pl

[Plan prezentacji



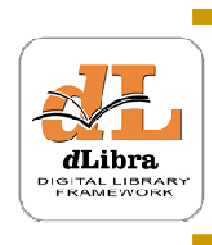
- Wprowadzenie
- Protokół OAI-PMH
- Narzędzia OAI-PMH
- Wyszukiwanie zasobów rozproszonych w systemie dLibra
- Podsumowanie

[Plan prezentacji

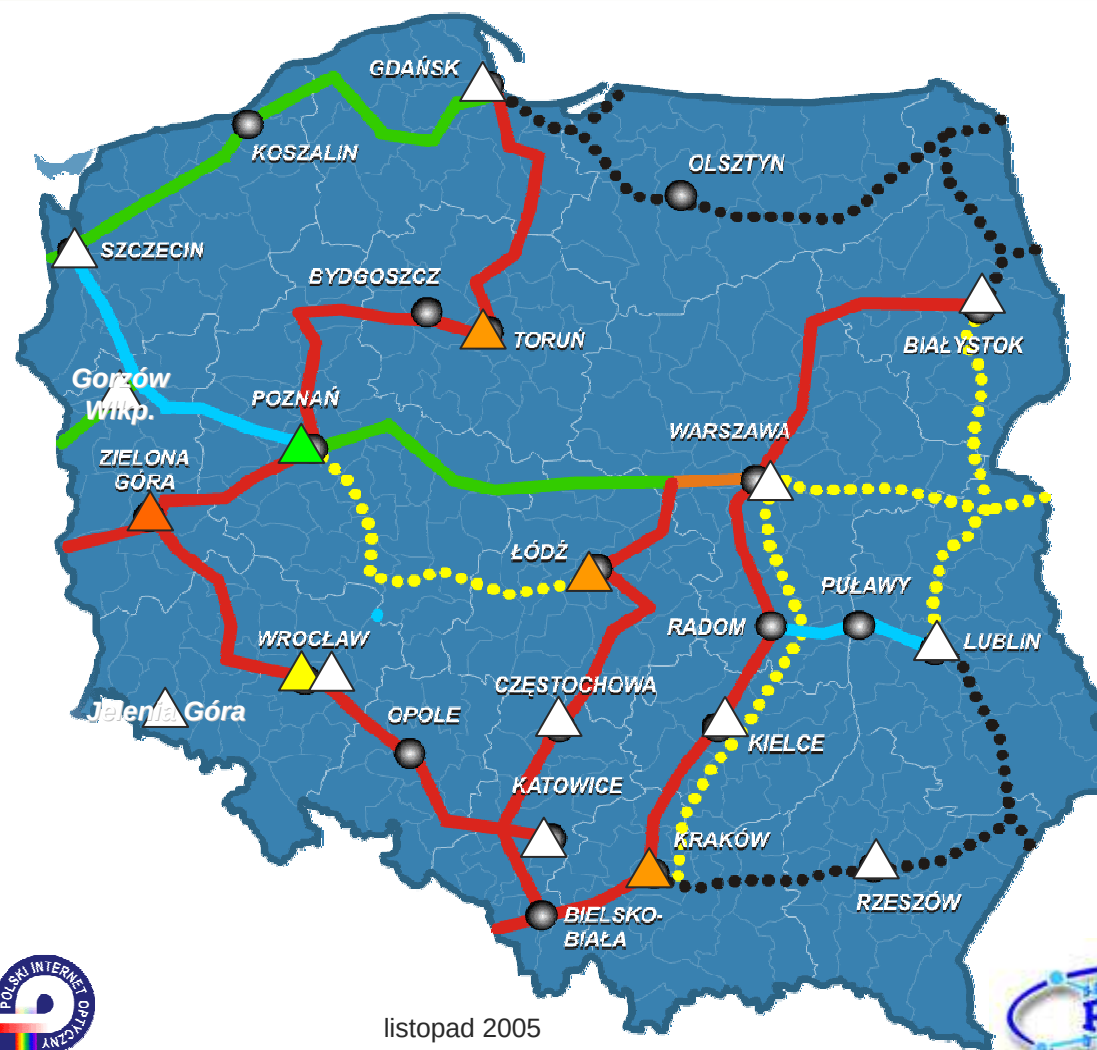


- **Wprowadzenie**
- Protokół OAI-PMH
- Narzędzia OAI-PMH
- Wyszukiwanie zasobów rozproszonych w systemie dLibra
- Podsumowanie

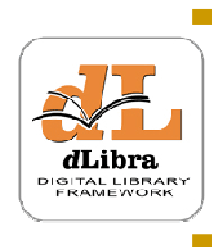
[Instalacje systemu dLibra]



- dLibra
- ▲ 2002 (WBC)
 - ▲ 2004
 - ▲ 2005
 - △ 2006?

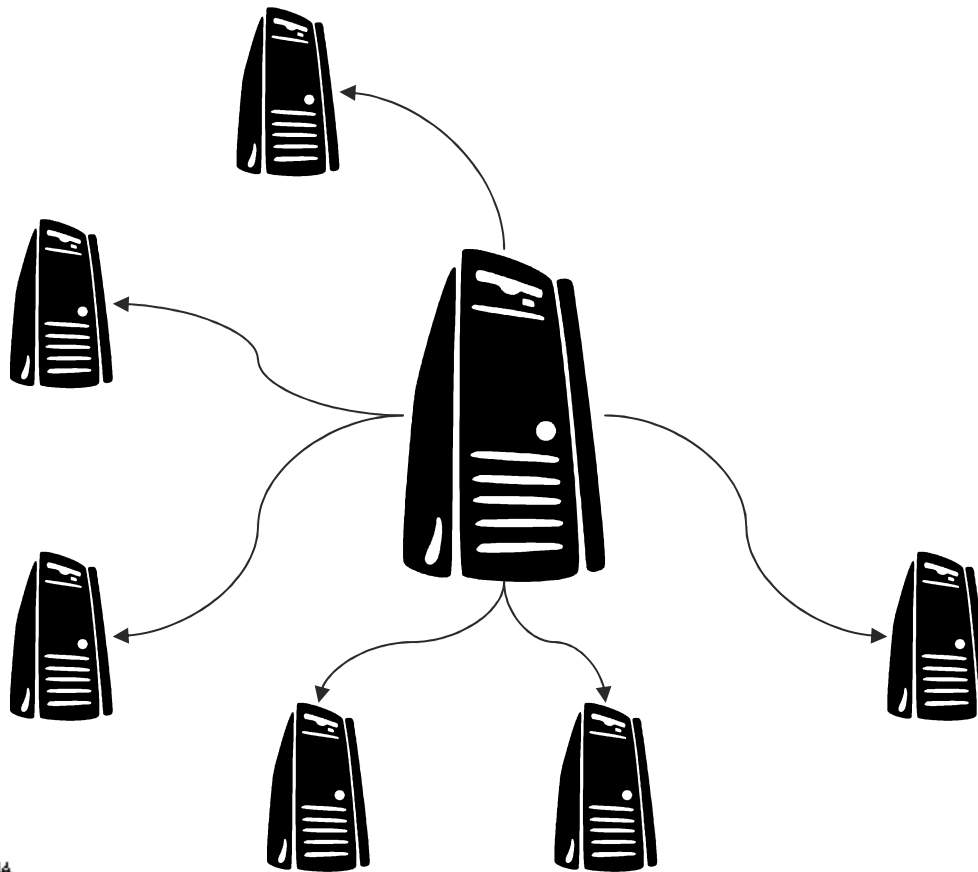


[Wprowadzenie

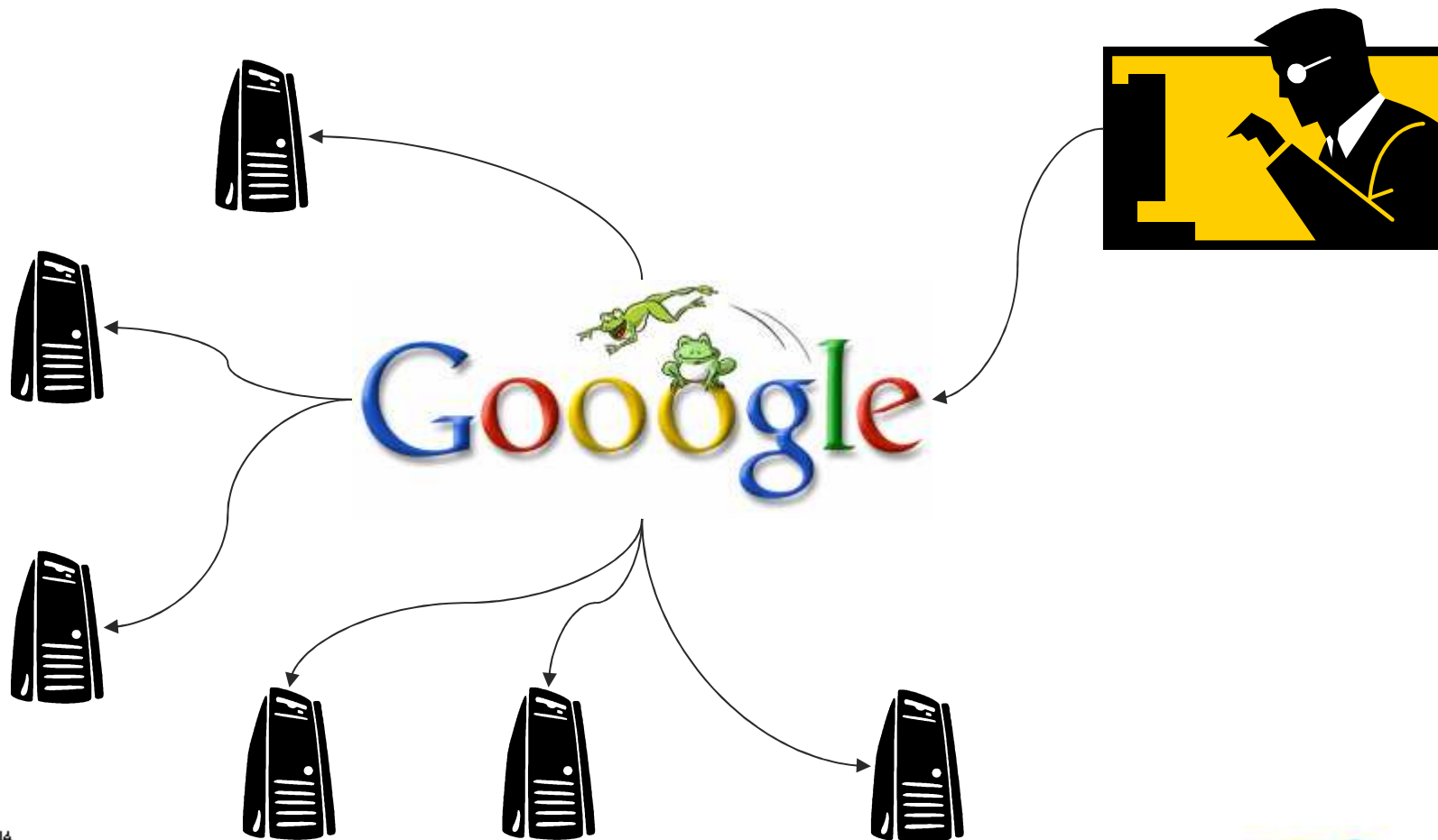


- Dwa podstawowe schematy architektury systemów rozproszonych
 - Węzeł nadrzędny i węzły podrzędne
 - „Każdy z każdym” (P2P)

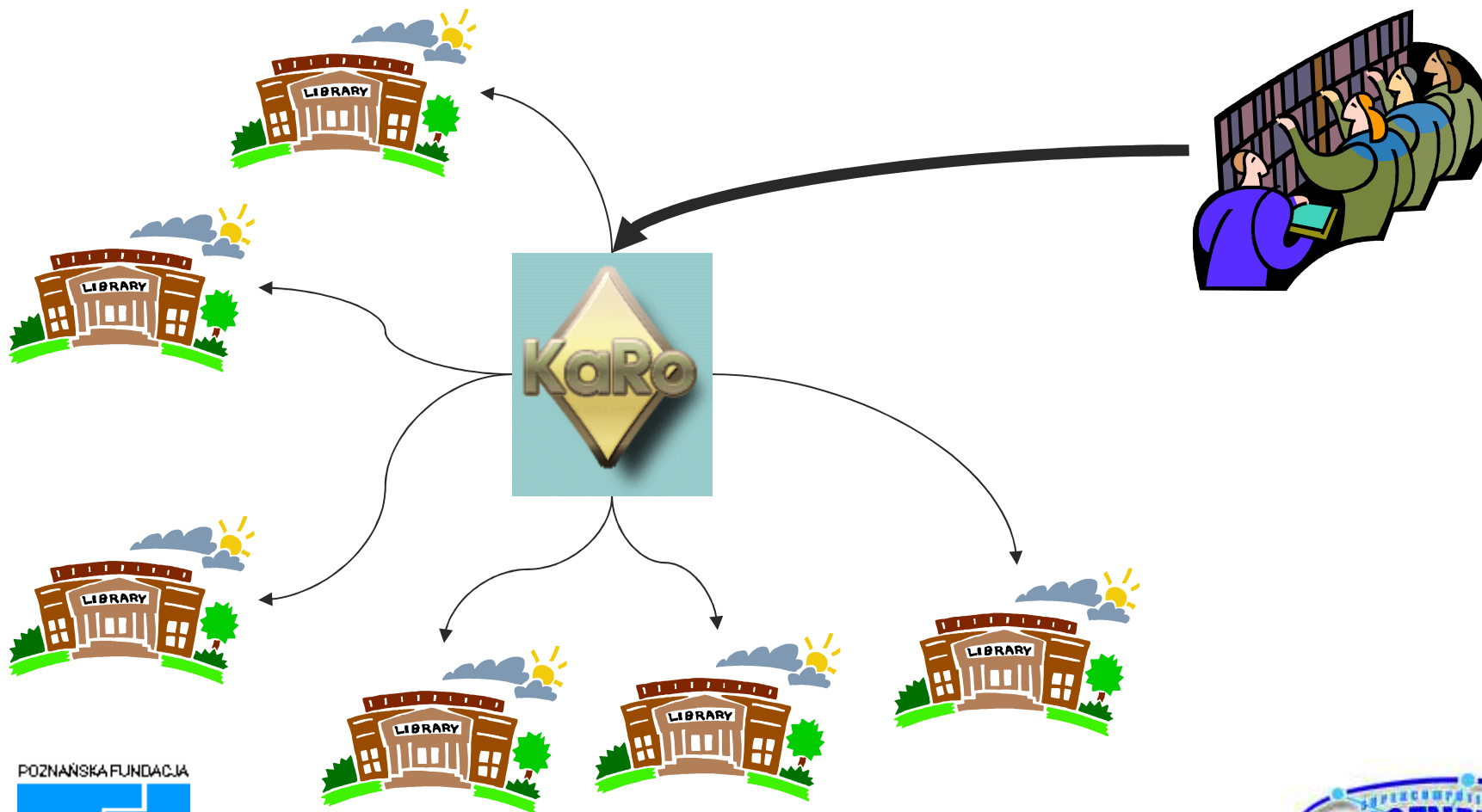
Węzeł nadrzędny i węzły podrzędne



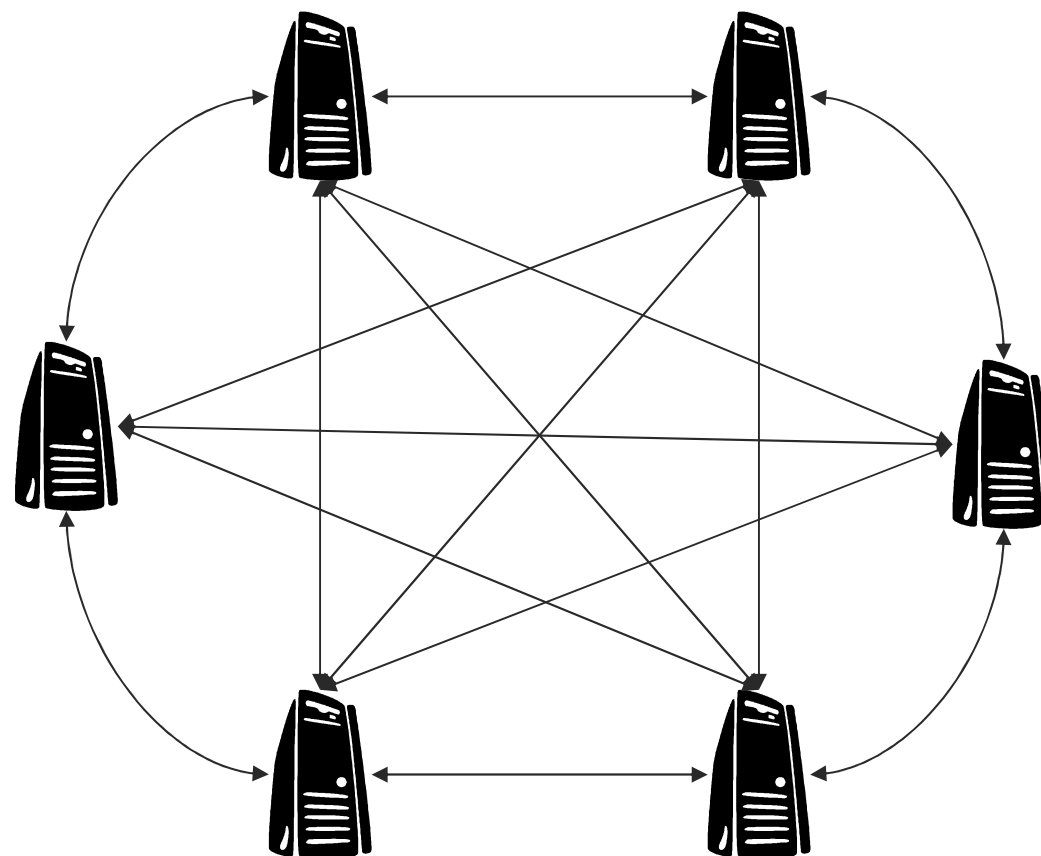
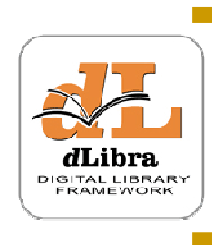
Węzeł nadrzędny i węzły podrzędne



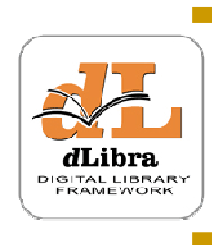
Węzeł nadrzędny i węzły podrzędne



[„Każdy z każdym” (P2P)

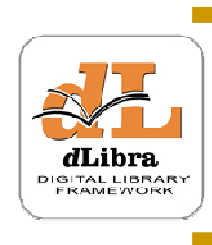


[Plan prezentacji



- Wprowadzenie
- **Protokół OAI-PMH**
- Narzędzia OAI-PMH
- Wyszukiwanie zasobów rozproszonych w systemie dLibra
- Podsumowanie

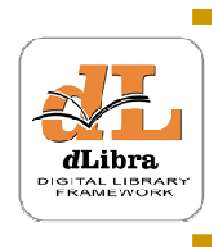
[Protokół OAI-PMH



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

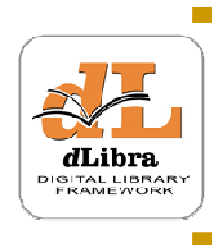


[Protokół OAI-PMH



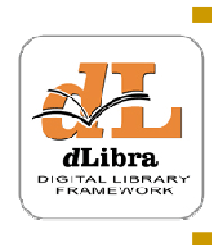
- 21-22 października 1999
Santa Fe, New Mexico, USA
- Podstawowe wnioski:
 - Potrzebny jest protokół do selektywnego pobierania metadanych
 - Potrzebne są kryteria selektywnego pobierania
 - Potrzebne są formaty, w których metadane będą przesyłane

[Protokół OAI-PMH]



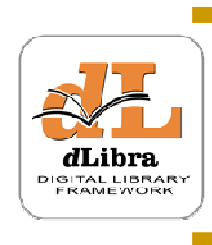
- Żądania: HTTP
- Odpowiedzi: XML
- Kryteria do selektywnego pobierania:
 - Data modyfikacji zasobu
 - Przynależność do zbioru
- Obecna wersja: 2.0
 - 14 czerwca 2002 r.

[Protokół OAI-PMH 2.0]



- Podstawowe pojęcia:
 - „Harvester” – system pobierający metadane (wysyłający żądania)
 - Repozytorium – system udostępniający metadane (wysyłający odpowiedzi)
 - Element – pojedynczy obiekt opisywany przez repozytorium, musi mieć unikalny identyfikator
 - Rekord – metadane zapisane w konkretnym schemacie
 - Zbiór – sposób grupowania elementów na potrzeby selektywnego pobierania metadanych

[Protokół OAI-PMH 2.0]



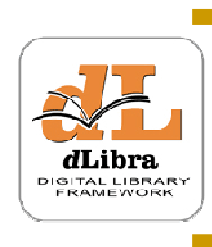
- Rodzaje żądań
 - Identify – pobiera informacje o repozytorium
 - ListMetadataFormats – pobiera informacje o dostępnych schematach metadanych
 - identifier – identyfikator rekordu
 - ListSets – pobiera informacje o zbiorach
 - resumptionToken – identyfikator żądania

[Protokół OAI-PMH 2.0]



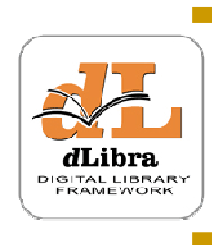
- GetRecord – pobiera pojedynczy rekord
 - identifier – identyfikator rekordu
 - metadataPrefix – schemat metadanych
- ListIdentifiers/ListRecords – pobiera identyfikatory/rekordy
 - from – początkowa data modyfikacji rekordów
 - until – końcowa data modyfikacji rekordów
 - metadataPrefix – schemat metadanych
 - set - zbiór
 - resumptionToken – identyfikator żądania

[Protokół OAI-PMH 2.0]



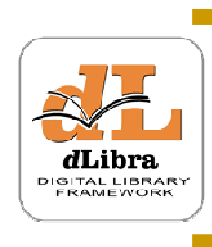
- Dodatkowe elementy protokołu
 - Usunięte rekordy
 - Kompresja odpowiedzi
- Każde repozytorium może wskazać listę swoich „przyjaciół”
- Repozytoria można rejestrować na stronach Open Archives Initiative

[Protokół OAI-PMH 2.0]



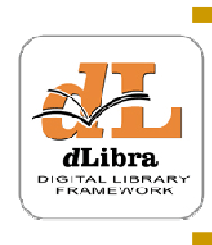
- Oprogramowanie dLibra wspiera protokół OAI-PMH w jego najnowszej wersji z wszystkimi elementami opcjonalnymi
 - <http://<biblioteka.cyfrowa>/dlibra/oai-pmh-repository.xml>

[Plan prezentacji



- Wprowadzenie
- Protokół OAI-PMH
- **Narzędzia OAI-PMH**
- Wyszukiwanie zasobów rozproszonych w systemie dLibra
- Podsumowanie

[OAI Repository Explorer]



- <http://re.cs.uct.ac.za/>
- Umożliwia
 - Przeglądanie repozytoriów OAI-PMH
 - Testowanie zgodności repozytoriów ze standardem opisującym protokół
- Udostępnia listę repozytoriów

[UIUC OAI Repository Registry]



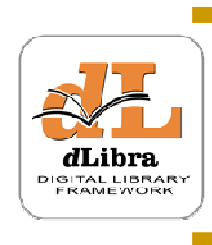
- UIUC = University of Illinois at Urbana-Champaign
- Największa lista repozytoriów OAI
 - 870 aktywnych repozytoriów
 - Google widzi tylko 836
- <http://gita.grainger.uiuc.edu/registry/searchform.asp>
- Repozytoria są okresowo odwiedzane i sprawdzane

[OAIster



- <http://oaister.umd.umich.edu/o/oaister/>
- University of Michigan Digital Library Production Service
- Indeksuje opisy zasobów z repozytoriów OAI i umożliwia ich przeszukiwanie
 - 6 058 086 rekordów z 568 instytucji

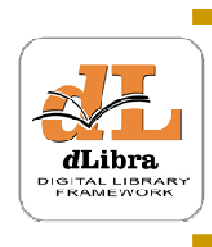
[OCKHAM



- <http://www.ockham.org/>
- Zestaw bezpłatnego oprogramowania realizującego usługi dla bibliotek cyfrowych
- OCKHAM Harvest-to-Query (H2Q) – dodatkowa warstwa umożliwiająca dostęp do repozytoriów OAI przy pomocy:
 - Protokołu Z39.50 – obecna wersja (0.5.3), tylko metadane w formacie DublinCore
 - Protokołu SRU/SRW i metadane w dowolnym formacie – planowane dla docelowej wersji 1.0

■ Repozytoria OAI w KARO?

[Plan prezentacji



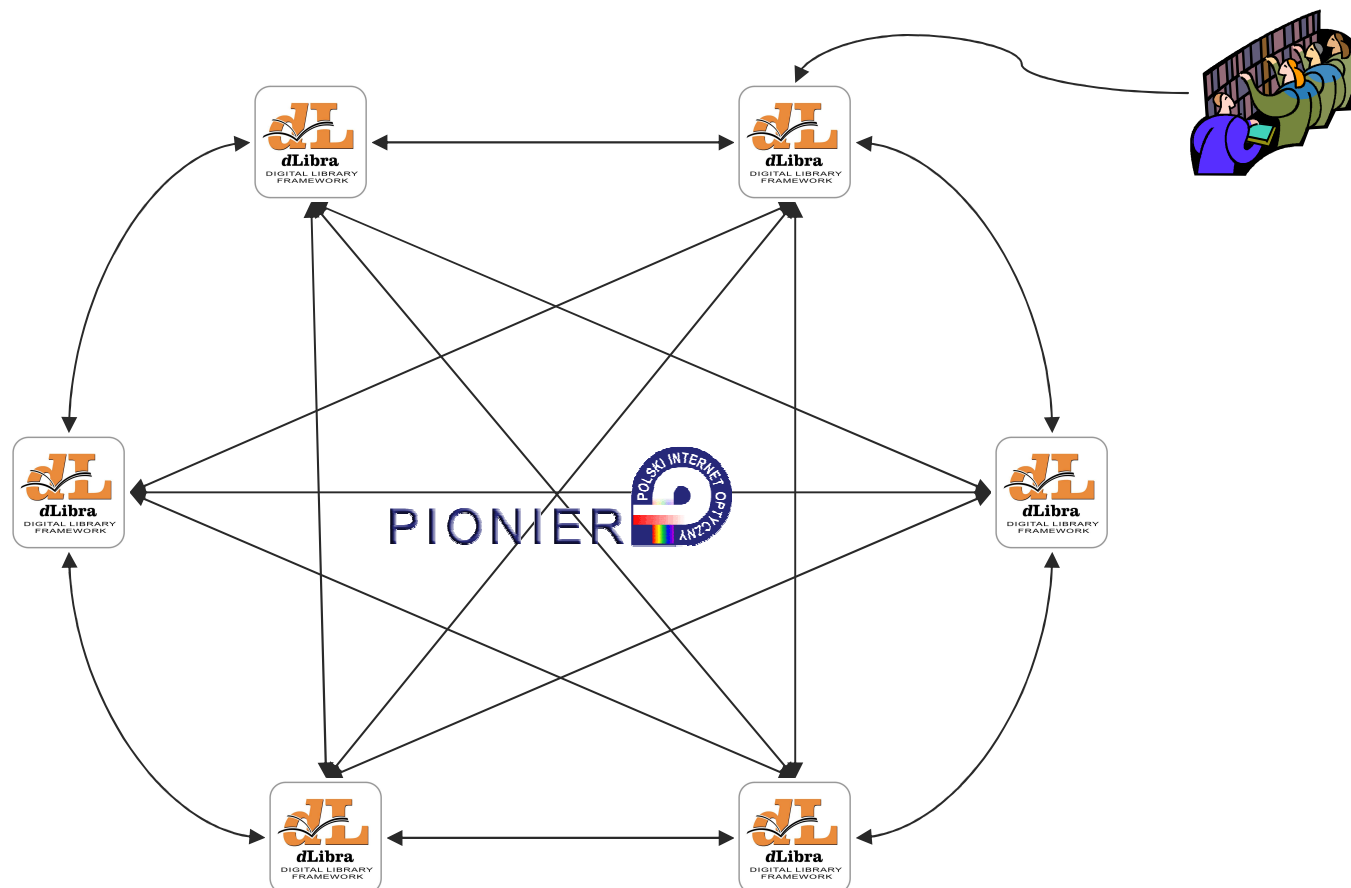
- Wprowadzenie
- Protokół OAI-PMH
- Narzędzia OAI-PMH
- **Wyszukiwanie zasobów rozproszonych w systemie dLibra**
- Podsumowanie

Wyszukiwanie zasobów rozproszonych w systemie dLibra



- Wyszukiwanie zasobów rozproszonych w systemie dLibra wykorzystuje podejście typu P2P
- Do wymiany metadanych wykorzystywany jest protokół OAI-PMH
 - Otwarty, popularny na świecie, nie ogranicza do określonych technologii

Wyszukiwanie zasobów rozproszonych w systemie dLibra



Wyszukiwanie zasobów rozproszonych w systemie dLibra



- Każda biblioteka decyduje, jakie inne biblioteki indeksować
- Biblioteki są indeksowane okresowo, w sposób przyrostowy
 - Niezbędne są informacje o usuniętych publikacjach
- Proces przeszukiwania odbywa się lokalnie

[Wyszukiwanie zasobów rozproszonych w systemie dLibra]



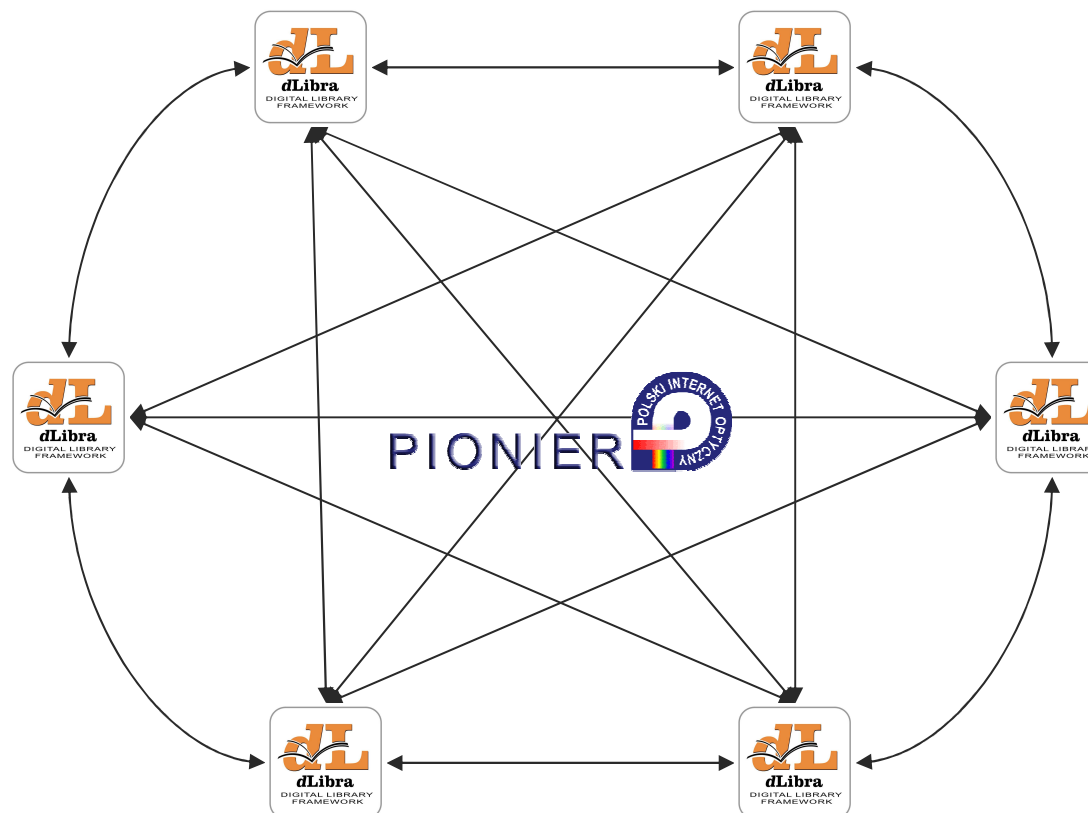
- Takie podejście:
 - pozwala zrównoważyć obciążenie poszczególnych bibliotek
 - Jest możliwe przy indeksowaniu metadanych ze względu na ich niewielki rozmiar

Wyszukiwanie zasobów rozproszonych w systemie dLibra

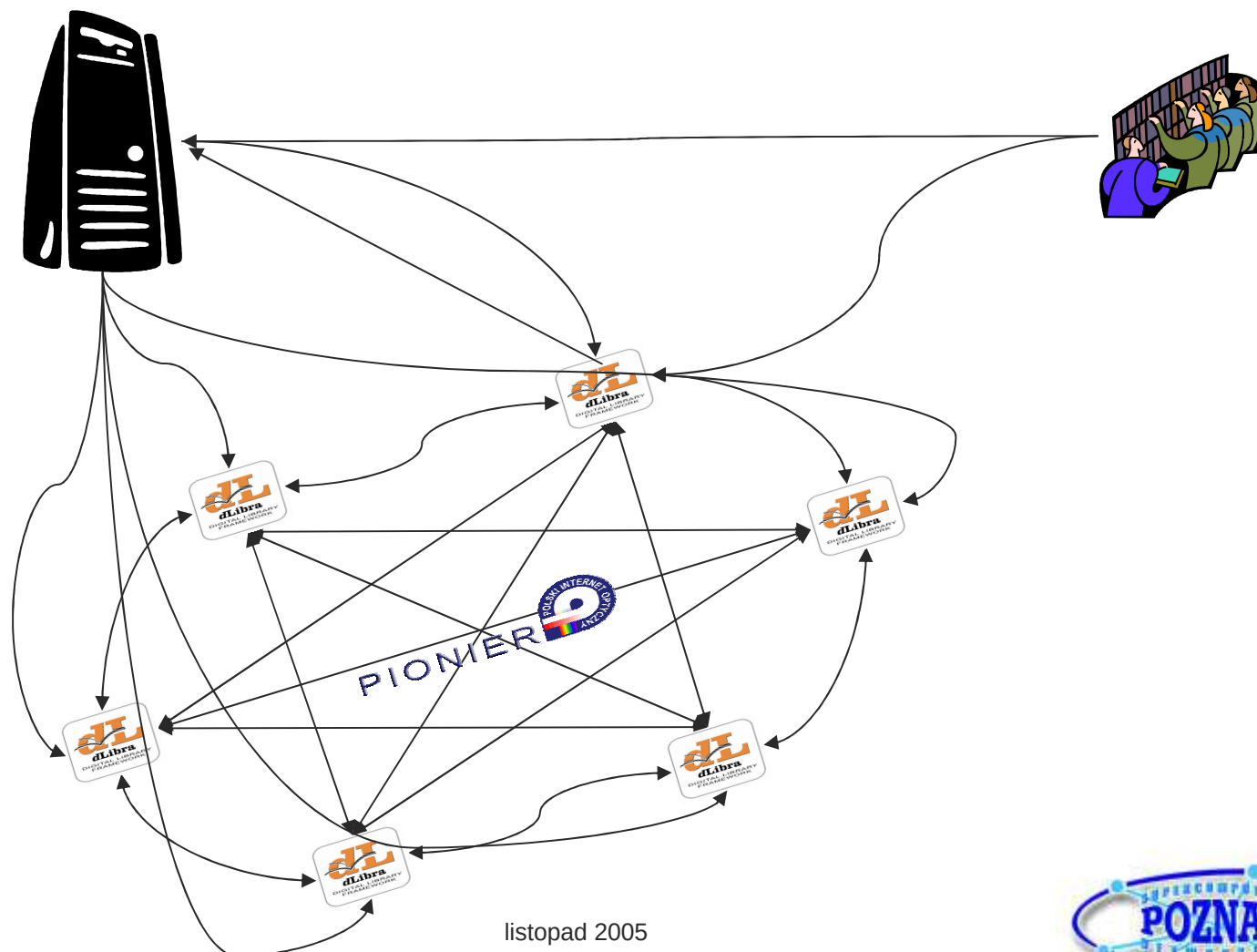
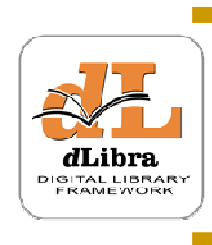


- Wyszukiwanie w treści zasobów rozproszonych
 - Można wykorzystać protokół OAI-PMH
 - Problem ilości przesyłanych danych
 - Można przysyłać sam tekst publikacji
 - Google vs DjVu
 - Można zmodyfikować architekturę systemu rozproszonych bibliotek

Wyszukiwanie zasobów rozproszonych w systemie dLibra



Wyszukiwanie zasobów rozproszonych w systemie dLibra



Wyszukiwanie zasobów rozproszonych w systemie dLibra



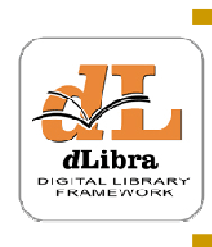
- Takie podejście spowoduje stworzenie dedykowanej wyszukiwarki bibliotecznej działającej na zasadzie Google w oparciu o protokół OAI-PMH
- Określenie ostatecznej architektury wymaga jeszcze wielu analiz i testów

[Plan prezentacji



- Wprowadzenie
- Protokół OAI-PMH
- Narzędzia OAI-PMH
- Wyszukiwanie zasobów rozproszonych w systemie dLibra
- **Podsumowanie**

[Podsumowanie]



- W Polsce mamy:
 - Coraz więcej bibliotek cyfrowych
 - Bardzo szybką i nowoczesną sieć PIONIER
- Ograniczenia technologiczne są stosunkowo łatwe do pokonania
- Pozostają kwestie organizacyjne i finansowe.....

Rozproszone biblioteki cyfrowe



Dziękuję za uwagę