

# POZNAN SUPERCOMPUTING AND NETWORKING CENTER



## **Metadata harvesting in regional digital libraries in PIONIER Network**

Cezary Mazurek, Maciej Stroiński,  
Marcin Werla, Jan Węglarz

## Polish Optical Internet PIONIER

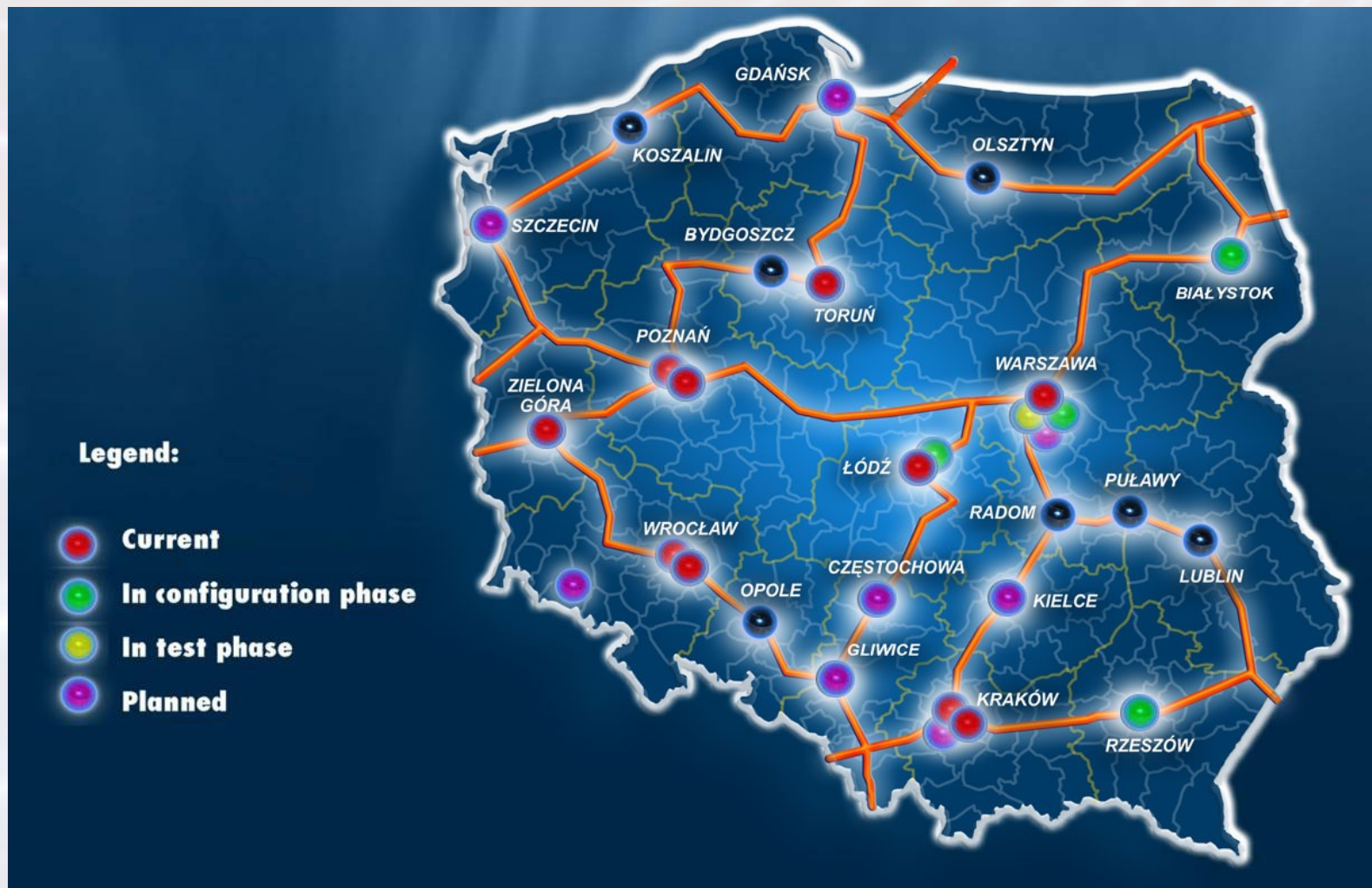




## Digital Libraries in PIONIER (2005)



## Digital Libraries in PIONIER (2006)





## Digital Libraries in PIONIER

- Support for multiple content types
- Hierarchic collections
- Resources versioning
- Advanced support for resources' metadata (DublinCore, MARC, RDF)
- Multiple searching mechanism
- Extended access management (LDAP, Single Sign-On)
- RSS-based notification
- Multitier, distributed and portable DL platform

# POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER



Kujawsko-Pomorska Digital Library - Postcards - Mozilla

File Edit View Go Bookmarks Tools Window Help



## Kujawsko-Pomorska Digital Library

readers on-line: 101

[MAIN PAGE](#) | [SEARCH](#) | [EXHIBITIONS](#) | [LOGIN](#) | [CONTACT](#) | [HELP](#)

### Information

- Project description
- Project Participants
- Technical information
- Frequently asked questions
- Contact

### Library

- Kujawsko-Pomorska Digital Library
  - Cultural heritage
  - Educational materials
- Regional materials
  - Kujawy-Pomorze
    - Art
    - Articles
    - Books
    - Ephemera
    - Maps
    - Photographs
    - Postcards
    - Printed music
    - Regional newspapers
    - Torun
      - graphics
      - school
    - Vilnensia

Digital Library of Wielkopolska - Main page - Mozilla

File Edit View Go Bookmarks Tools Window Help



## Digital Library of Wielkopolska

readers on-line: 115

[MAIN PAGE](#) | [SEARCH](#) | [EXHIBITIONS](#) | [LOGIN](#) | [CONTACT](#)

### Library

- Digital Library of Wielkopolska
  - Cultural heritage
  - Educational materials
  - Music notes
  - Regional materials

### Information

- Project description
- Project Participants
- Technical information
- Frequently asked questions
- Contact

### Extras

- Create account
- Statistics
- Planned publications
- Entire publication catalogue [ZIP] [HTML] [TXT]
- Add Firefox search plugin to your browser
- Add to favourites (Internet Explorer)



## Digital Library of Wielkopolska

[MAIN PAGE](#) | [SEARCH](#) | [EXHIBITIONS](#) | [LOGIN](#) | [CONTACT](#)

### Introduction

Digital Library of Wielkopolska gives you access to the oldest Polish writing relicts. On our page you can also read scientific scripts and monographs and see exhibitions with unique resources from various Wielkopolska libraries.

Number of publications in library: 15069  
Current number of users in library: 115  
Total number of users since Jun 10, 2004: 1570734

### Recently added

1. Amtsblatt des Landrats in Ostrowo 1940.12.05 R.1940 nr98
2. Amtsblatt des Landrats in Ostrowo 1940.12.02 R.1940 nr97
3. Uwag nad Panem Tadeuszem
4. Amtsblatt des Landrats in Ostrowo 1940.11.28 R.1940 nr96
5. Amtsblatt des Landrats in Ostrowo 1940.11.25 R.1940 nr95
6. Amtsblatt des Landrats in Ostrowo 1940.07.11 R.1940 nr94
7. Amtsblatt des Landrats in Ostrowo 1940.11.18 R.1940 nr93
8. Amtsblatt des Landrats in Ostrowo 1940.11.14 R.1940 nr92
9. Amtsblatt des Landrats in Ostrowo 1940.11.11 R.1940 nr91
10. 50 Jahre und leichter

Search scope:  
Everywhere  
Searched word(s):

- ☒ Use synonyms  
☐ Search remote libraries

Search

Advanced search

- Titles index
- Authors index
- Keywords index

25 last publications  
Planned publications  
News

What is RSS feed?

Digital Library of Zielona Gora - Nowa Sól / Neusalz; Gruß aus Neusalz; Pozdrowienia z Nowej Soli - Mozilla

File Edit View Go Bookmarks Tools Window Help



## Digital Library of Zielona Gora

now in library: 13

[MAIN PAGE](#) | [CATALOGUE](#) | [EXHIBITIONS](#) | [LOGIN](#) | [CONTACT](#) | [HELP](#)

### Library

- Digital Library of Zielona Gora
  - Science and Teaching
  - MA, MSc, PhD and Habilitation Thesis
  - Culture and Fine Arts
  - Regional Materials

### Publication

- Description
- Content
- Content in new window
- Entire publication as ZIP

### Publication description

Nowa Sól / Neusalz; Gruß aus Neusalz; Pozdrowienia z Nowej Soli

DublinCore schema version 1.1 attributes

**Title:** Nowa Sól / Neusalz; Gruß aus Neusalz; Pozdrowienia z Nowej Soli  
**Subject and Keywords:** pocztówka ; Neusalz ; Nowa Sól ; Lutherstraße ; Ulmühlenstraße ; Teilansicht  
**Description:** pocztówka czarno-biała ; Stempel pocztowy: Neusalz (Oder) 11.07.1936  
**Publisher:** Berlin: Paul Mecklenburg  
**Date:** 1936  
**Resource Type:** pocztówka  
**Format:** image/x.djvu  
**Language:** ger  
**Coverage:** Kolekcja zawiera kilkadziesiąt pozycji. Prezentujemy jej fragment  
**Rights Management:** Biblioteka Uniwersyteku Zielonogórskiego  
**Publication's collections:** Regional Materials  
Postcards

In our library since: May 16, 2006  
Last modified: May 16, 2006  
Number of publication content hits: 0  
All available editions: <http://zbc.uz.zgora.pl/publication/3015>  
Miniature:



Publication structure:

- Przedwojenne pocztówki lubuskie - Nowa Sól / Neusalz
  - Nowa Sól / Neusalz; Markt; Rynek
  - Nowa Sól / Neusalz; Promenade mit Jahnedenkmal; Promenade z pomnikiem
  - Nowa Sól / Neusalz; Kreiskrankenhaus; Szpital
  - Nowa Sól / Neusalz; Markt; Rynek
  - Nowa Sól / Neusalz; Schmuckplatz und Bahnhofstrasse
  - Nowa Sól / Neusalz; Markt; Rynek; Hansa-Café
  - Nowa Sól / Neusalz; Hafen; Port

Warsaw University of Technology Digital Library - Main page - Mozilla

File Edit View Go Bookmarks Tools Window Help



## Digital Library of Warsaw University of Technology





Kujawsko-Pomorska Digital Library - Toruń - Ratusz Staromiejski - Mozilla

File Edit View Go Bookmarks Tools Window Help

## Kujawsko-Pomorska Digital Library

readers on-line: 106


MAIN PAGE | SEARCH | EXHIBITIONS | LOGIN | CONTACT | HELP |  

### Information

**Title :**  
Toruń - Ratusz  
Staromiejski

### Publication

- Publication description
- **Content**
- Content in new window
- Entire publication as ZIP
- Search results



Thorn History Torunia! Rathaus  
Dr. Trenkler Co., Leipzig 20425

dLibra  
DIGITAL LIBRARY  
FRAMEWORK

http://kpbcr.umk.pl/Content/14152/Pocztowki\_00016\_0001.djvu DONE

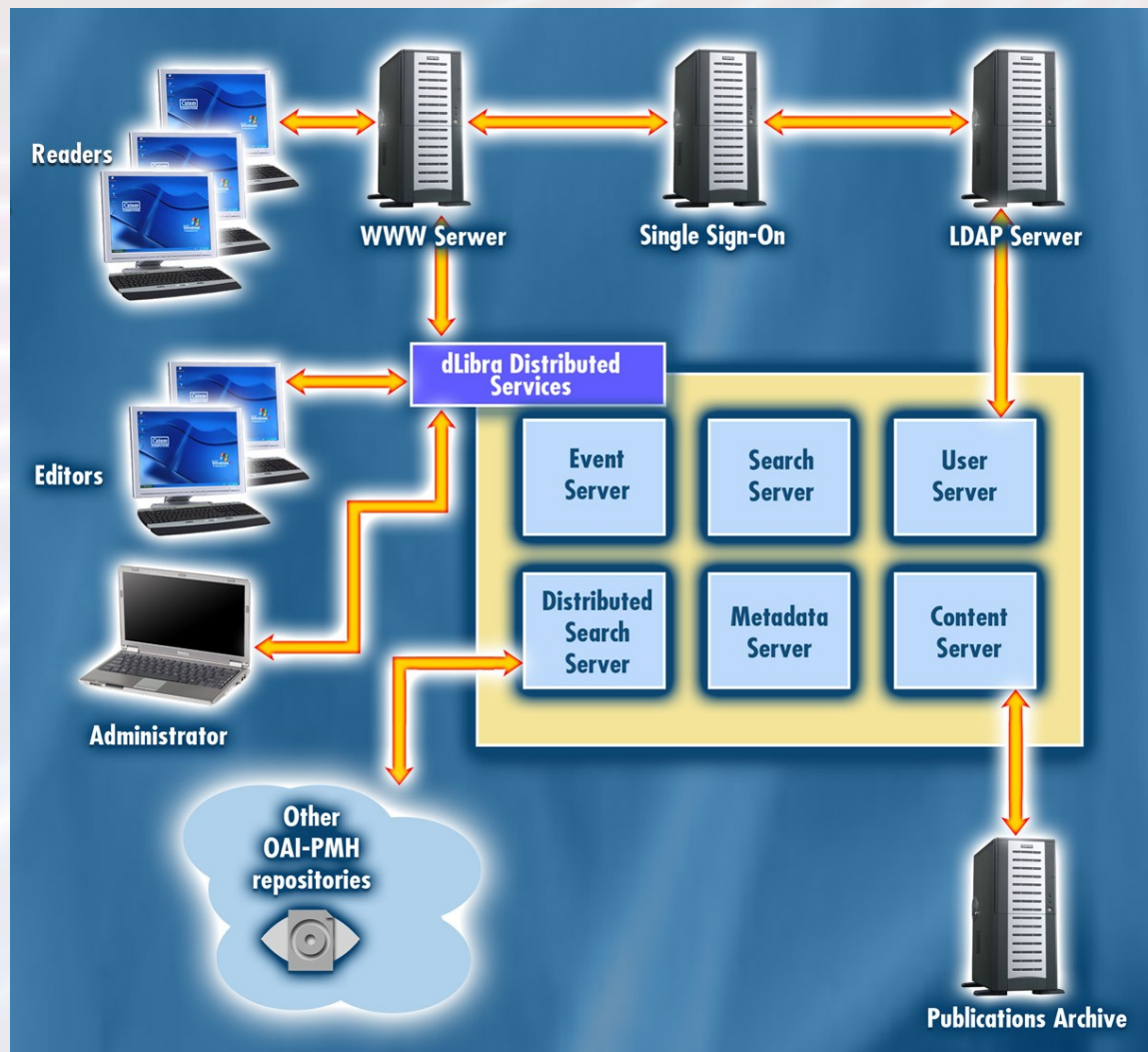
## dLibra Digital Library Framework



<http://dlibra.psnc.pl/>



## Architecture

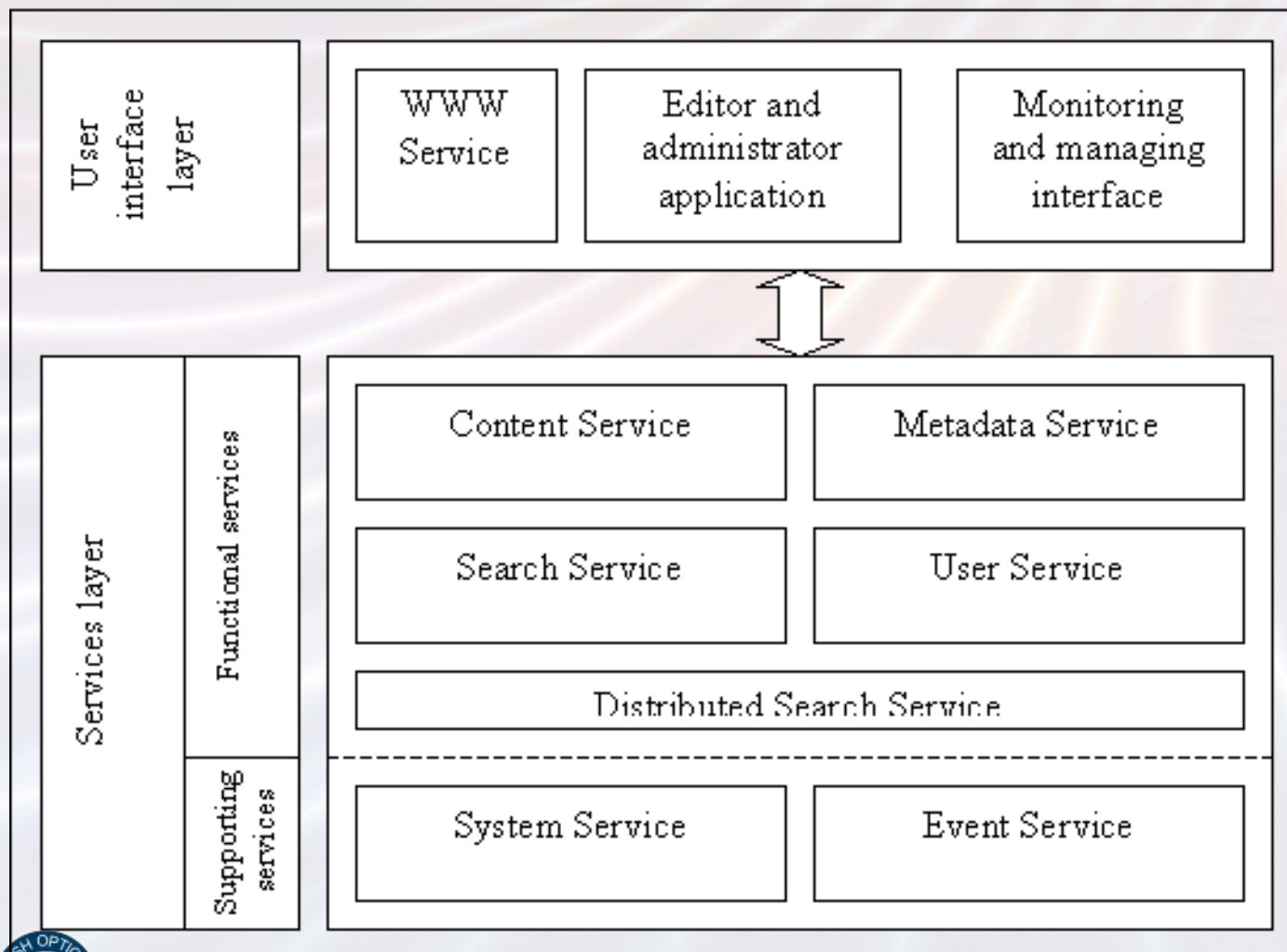


## Service Oriented Architecture

- Each service can be deployed on different host
- Services does not need to know their locations
- Services can send and receive events
- Services implementations can be easily replaced
- Services can access each other (if authorized)
- External services can access DL system (if authorized)
  - Via dLibra interfaces with Java RMI
  - Via OAI-PMH



## Service layers



## Services

- Functional Services
  - Metadata Service
  - Content Service
  - Search Service
  - Distributed Search Service
  - User Service
- Supporting Services
  - System Service
  - Event Service



## User interface

- WWW Service
  - Servlets
  - Portlets
- Editor/Administrator Application,
- Service Monitoring and Management Interfaces

## **dLibra release 2.2 reader features**

- **Distributed Search**
- **Latest Publications**
- **Advanced search**
- **RSS 2.0 feeds**
- **Search results as a RSS feed**
- **Publication description tabbed**
- ....



## OAI-PMH

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



## OAI-PMH Basic Terms

- Harvester
  - a system (client) which wants to access metadata stored in a remote system
- Repository
  - a system (server) which allows to access metadata stored in it.
- Element
  - a basic object described in a repository, identified by an identifier unique at least at the repository level
- Record
  - an element's metadata in particular format
- Set
  - a way of grouping elements in a repository
  - each repository accessible through the OAI-PMH protocol can define its own sets of elements. Sets can create a hierarchical structure – each set can have subsets. Each element can belong to any number of sets



## OAI-PMH Other Features

- Criteria for selective harvesting
  - Last date of resource metadata record modification
  - Element membership in Sets
- Deleted records
- List of friends
  - allows to create easy discoverable networks of the connected repositories

## OAI-PMH Requests

- Identify
  - gets information about the repository
- ListMetadataFormats
  - gets information about metadata formats supported by the repository. Each repository must at least support Dublin Core Metadata Element Schema.
- ListSets
  - gets information about sets defined in the repository
- GetRecord
  - gets a single record – describing a specified element in a specified metadata schema
- ListIdentifiers
  - gets a list of identifiers of elements available in the repository.
- ListRecords
  - gets a list of records available in the repository.

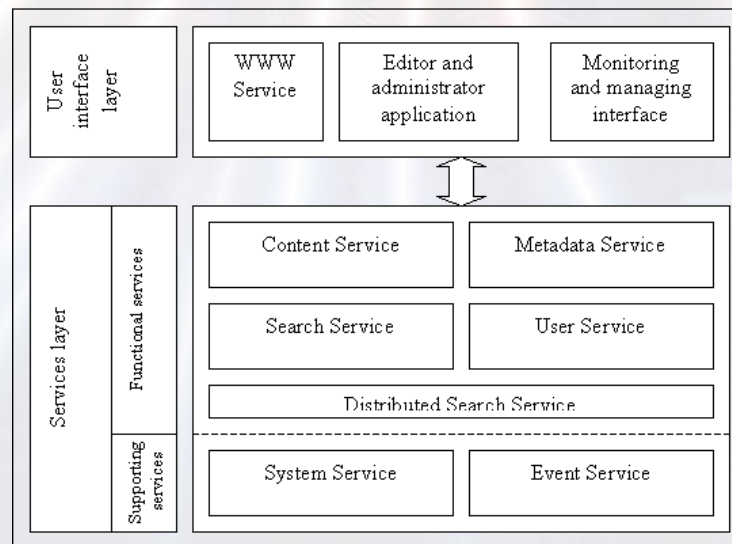


## **Metadata exchange between DLs**

- Periodic metadata synchronization
  - Selective harvesting
  - Deleted records
- Collections -> Sets
- Metadata -> DC MES
- OAI-Identifiers
- Each DL has all metadata from the platform

## Implementation

- dLibra became OAI-PMH repository as well as harvester
- Compatible with OAI-PMH 2.0
- Repository interface is embeded in WWW Service
  - Persistent support for deleted elements;
  - Sets
  - Flow control
  - Globally unique ids
- Harvester is provided as Distributed Search Service
  - List of repositories
  - Periodical harvesting





## **Content-based Services**


- Distributed searching
- Virtual collections
- Distributed exhibitions
- Comments and annotations

## **Distributed searching**

- As a result the user receives one search results list with both local and remote resources which metadata matched the user query
- Might be accessed by any other network service or information portal



Wielkopolska Biblioteka Cyfrowa - Wielkopolska Biblioteka Cyfrowa - Mozilla




obecnie czytających: 371
STRONA GŁÓWNA | WYSZUKIWANIE | WYSTAWY | LOGOWANIE | KONTAKT | POMOC |

### Informacje

- Opis projektu
- Uczestnicy Projektu
- Informacje techniczne
- Najczęściej zadawane pytania
- Kontakt

### Biblioteka

- Wielkopolska Biblioteka Cyfrowa
  - Dziedzictwo kulturowe
  - Materiały dydaktyczne
  - Materiały regionalne
  - Muzykalia

### Wyniki wyszukiwania ogólnego w kolekcji Wielkopolska Biblioteka Cyfrowa

Wyświetl wyniki jako RSS: [RSS 2.0](#)

Liczba znalezionych publikacji: 422









Lista publikacji spełniających warunków: Wszędzie: Toruń

Użyj synonimów: tak

Przeszukaj zdalne biblioteki: tak

[odnośnik do wyników wyszukiwania](#)

[powrót do zapytania](#)

-  Toruń - Ratusz Staromiejski [.....] [OAI](#)
-  Toruń - ulica Chełmińska [.....] [OAI](#)
-  Relacja o skrytych sztukach [.....]
-  Chronicon regum Poloniae, una cum origine gentis Regnique Polonici, provinciarum, districtuum [...] civitatum [...] descriptum per Erasmus Glicznerum [...] - Gliczner, Erazm (1535-1603) [.....]
-  Z ludem wielkopolskim przeciw zaborcom; wspomnienia - Jedlina-Jacobson, Wojciech [.....]
-  Książka adresowa miasta Torunia; według stanu z czerwca 1936 - Reinke, Marian; dyrektor Biura Zarządu Miejskiego w Toruniu; (oracował i wydał) [.....]
-  Toruń - ulica Kopernika 15 - Narębska-Dębska, Barbara (1921-2000) [.....] [OAI](#)
-  Klub Humoru w Toruniu : założony 11 maja 1919. [.....] [OAI](#)
-  Szkoła Podchorążych Artylerji : ku uczczeniu dziesięciolecia 1923-1933 [.....] [OAI](#)
-  Szkoła Podchorążych Artylerji ku uczczeniu dnia promocji roczników 1927-1929, 1928-1930. [.....] [OAI](#)

## Virtual collections

- Static collections
- Dynamic collections
  - a set of conditions that object metadata must satisfy to belong to the collection
  - Based on RSS feeds
- Unified collections
  - Unified approach for collection names
  - Building virtual collections through metadata exploration



RSSOwl - RSS / RDF / Atom Newsreader

File Edit View Go Newsfeed Favorites Tools Help

New Import Export Reload Search Mark Next Favorites History Tools

Address

Favorites

- AmphetaRate
- RSSOwl
- Samples
- Virtual Dynamic Collections
  - Chemistry (9)

Chemistry

Search results

| Headline                                                                                                                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Burewicz Andrzej, 2005, System Visual Basic dla aplikacji w algorytmice i wizualizacji (Published in digital library : 2006-03-06)                           |
| Górny, Mirosław, Kształtowanie się współczesnych form współpracy w bibliotekach naukowych (Published in digital library : 2006-03-06)                        |
| Jarczewski Arnold (red.), 2006, Materiały do ćwiczeń laboratoryjnych z chemii organicznej dla studentów biologii (Published in digital library : 2006-03-06) |
| Opis bibliograficzny dokumentów w formacie MARC 21 w systemie HORIZON. Część 2 Wydawnictwo ciągłe (Published in digital library : 2006-03-06)                |
| Podstawy chemii nieorganicznej - ćwiczenia laboratoryjne (Published in digital library : 2004-08-26)                                                         |
| Theatrum chemicum... T. 1 (Published in digital library : 2005-03-01)                                                                                        |
| Theatrum chemicum... T. 2 (Published in digital library : 2005-03-01)                                                                                        |
| Theatrum chemicum... T. 3 (Published in digital library : 2005-03-01)                                                                                        |
| Theatrum chemicum... T. 5 (Published in digital library : 2005-03-01)                                                                                        |
| Warchalewski Jerzy R., 1994, Stan aktualny i perspektywy rozwoju wybranych dziedzin przetwórstwa żywności. Polagra '94 : 2005-03-01                          |

Podstawy chemii nieorganicznej - ćwiczenia laboratoryjne (Published in digital library : 2004-08-26)

<http://www.wbc.poznan.pl/dlibra/doccontent2?from=rss&id=3551>

Podstawy chemii nieorganicznej - ćwiczenia laboratoryjne

Gotowe

Ready

sobota, 11 marzec 2006

## Concluding Numbers

- 9 digital libraries in PIONIER based on proprietary software
- Over 25.000 publications already digitalized and made available and searchable
- 70.000 of attribute values in dictionaries
- 320.000 references to dictionaries



**Thank you!**

Cezary Mazurek

mazurek@man.poznan.pl