

ECDL 2009 Tutorial

# Aggregation and reuse of digital objects' metadata from distributed digital libraries

Prepared by:

PSNC Digital Libraries Team

(<http://dl.psnc.pl/>)

Speaker:

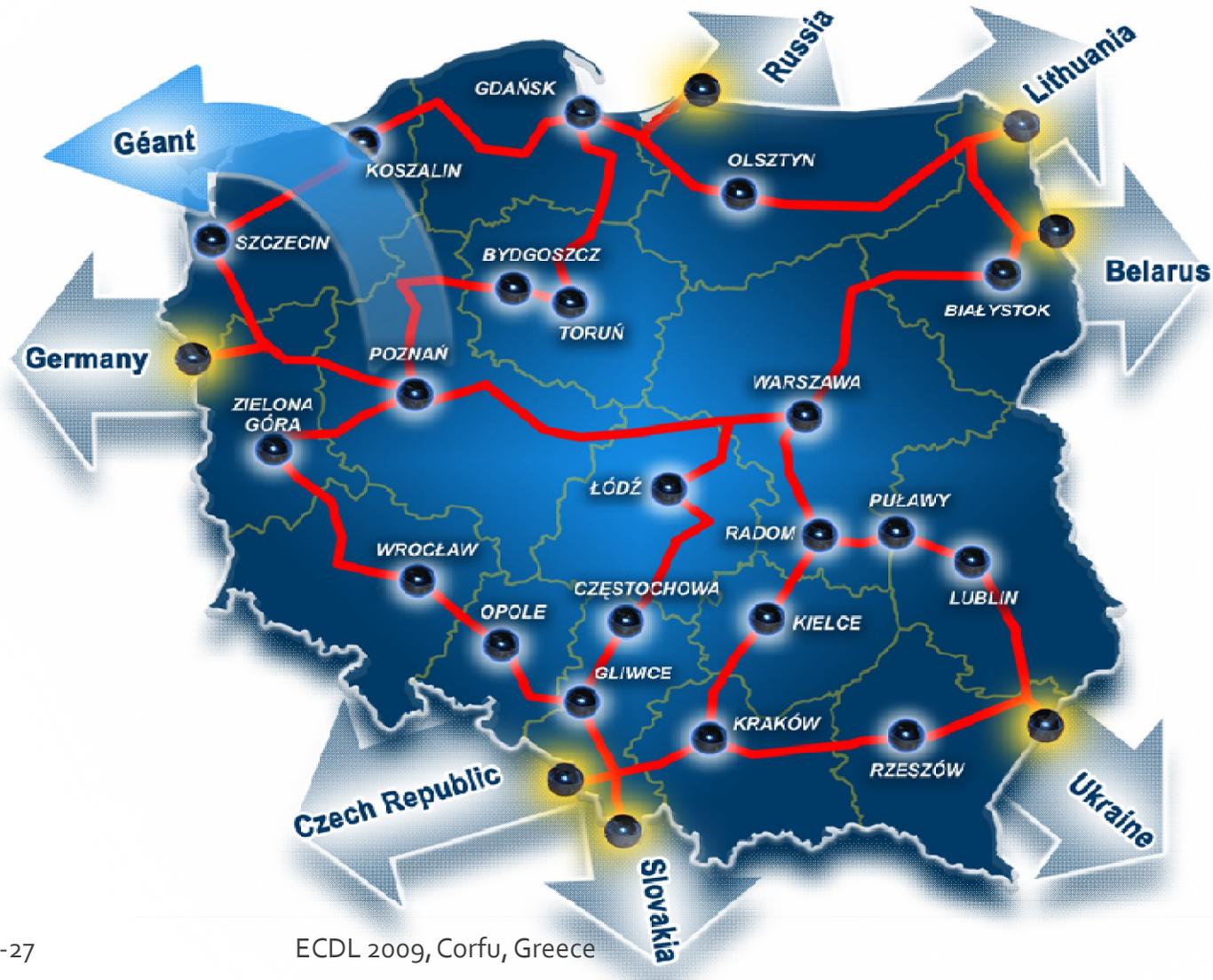
Marcin Werla

([mwerla@man.poznan.pl](mailto:mwerla@man.poznan.pl))

# Part 1: The development and the current state of the digital libraries infrastructure in Poland

ECDL 2009 Tutorial: Aggregation and reuse of digital objects' metadata from distributed digital libraries

# Polish Optical Internet PIONIER



2009-09-27

ECDL 2009, Corfu, Greece

# Paradigmes of the digital libraries infrastructure

- PIONIER Programme allowed the creation of a modern **distributed infrastructure** of digital libraries
- Digital content available in this infrastructure is processed by **tools and services** being part of this infrastructure
- The variety of content and tools makes the infrastructure valuable both in **scientific and social context**
- Tools and services identified on the level of a single digital library can be **federated** in different models on the level of the (global) digital libraries infrastructure

# Digital libraries in the PIONIER Network – organizational models

- Main organizational models
  - **Regional digital libraries**
    - Created and maintained by several institutions from particular region
    - Gather mostly resources related to the region, its history and culture but also academic educational materials and national cultural heritage
  - **Institutional digital libraries**
    - Created and maintained by single institutions (like universities)
    - Gather mostly resources related to present activities (like institutional repositories) and history of the institution
  - **Consortional digital libraries**
    - Created and maintained by consortia/federations of institutions with similar profile
    - Gather mostly resources related to present activities (like institutional repositories) and history of the consortium/federation institutions
- In many cases the technical base and support for digital libraries is provided by local computing or networking centers (like PSNC)

# European vision and Polish digital libraries

- The **i2010 Digital Libraries** initiative announced in 2005 by the European Commission and plans of the creation of European Digital Library (europeana.eu)
  - Virtual integration of Polish distributed digital libraries in one compound platform allowing the exchange of information between particular digital libraries
- Such approach allowed to resolve some of the problems mentioned in the i2010 initiative by
  - Facilitation of use of the resources from Polish digital libraries
  - Increase of its visibility in the Internet
  - Creation of automated mechanisms supporting the coordination of digitisation in distributed institutions

# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# Digital Libraries in Poland

Overall number of digital objects

✓ 265 thousands

Number of active digital libraries:

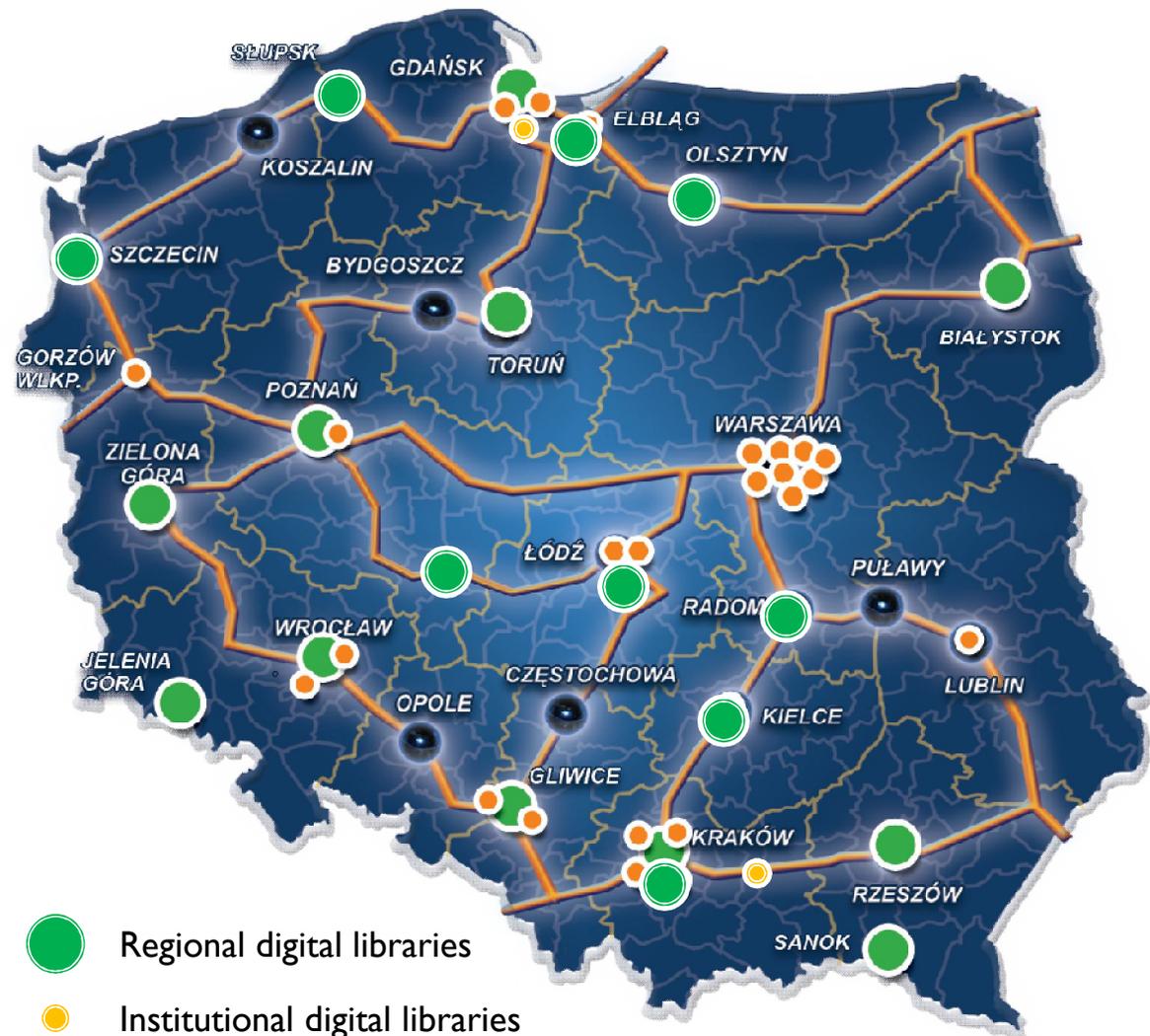
✓ 19 regional

✓ 21 institutional

+ **several other digital libraries** in the phase of planning, configuration or initial content uploading

Number of cooperating institutions:

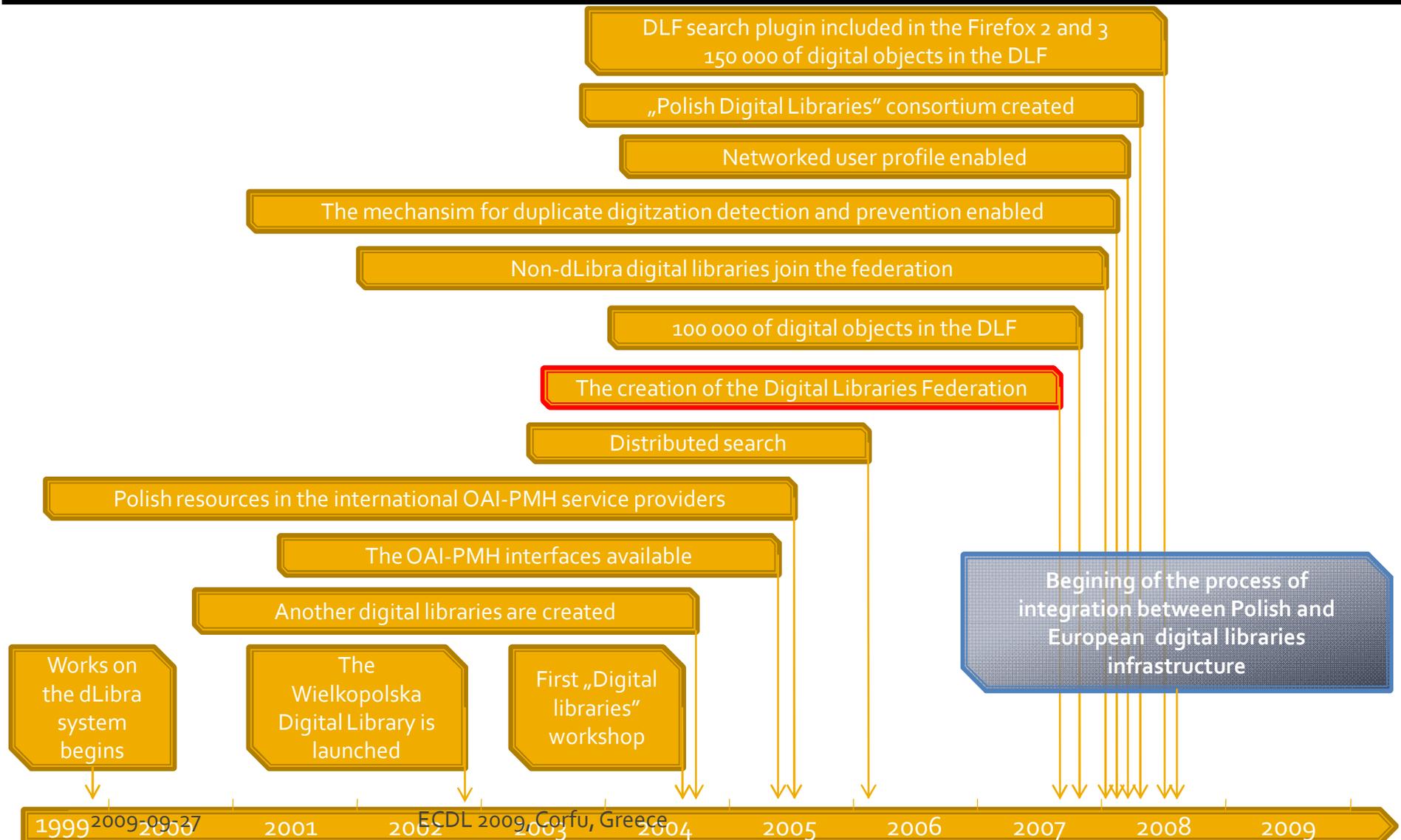
✓ Several hundreds of libraries, museums and archives



# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# The development of the digital libraries infrastructure in the PIONIER Network



# Digital Libraries Federation

- Main aims
  - **To facilitate the use of resources** from Polish digital libraries
  - **To increase the visibility** of these resources **in the Internet**
  - **To create new, advanced network services** both for end-users and digital libraries creators on the base of these resources

# Digital Libraries Federation

- Basic assumptions
  - **No need** nor requirement **to move** resources to the DLF
  - **No fees** for the use of the DLF and for being a part of it
  - **Open standards are the basis** for cooperation
    - Particular digital libraries can use different technological platforms

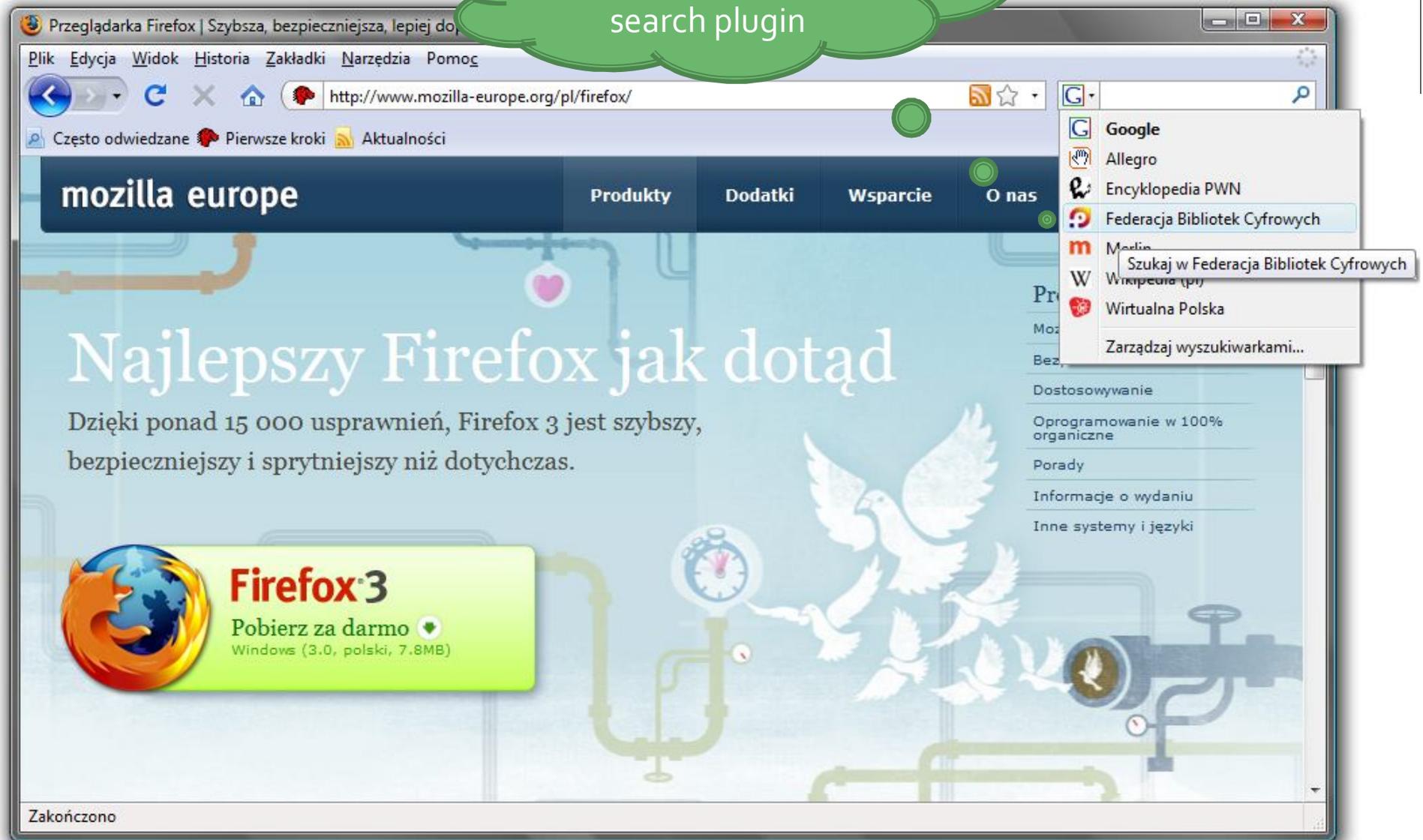
# Digital Libraries Federation

- Basic functions
  - Search in the available publications
    - Simple
    - Advanced
  - Digitization plans
    - Searchable
    - Report
    - API for the prevention of duplicated digitization
  - Location of digital objects on the basis of their OAI Identifiers
  - Database of Polish digital libraries
  - Statistics and reports
- Information in the DLF is updated on the daily (nightly) basis

# Promotion of Polish Digital Resources

- OpenSearch plug-in for web browsers available since the beginning of the DLF
  - Must be installed manually by the user
  - The user must know that there are digital libraries and the Federation
- Several months of activities towards the inclusion of DLF plug-in into official Polish Firefox releases
  - Negotiations with Polish Mozilla team, and finally with Mozilla Foundation
  - Technical changes related to high performance requirements
    - Autosuggest service response time less than 300 ms

Digital Libraries  
Federation  
search plugin

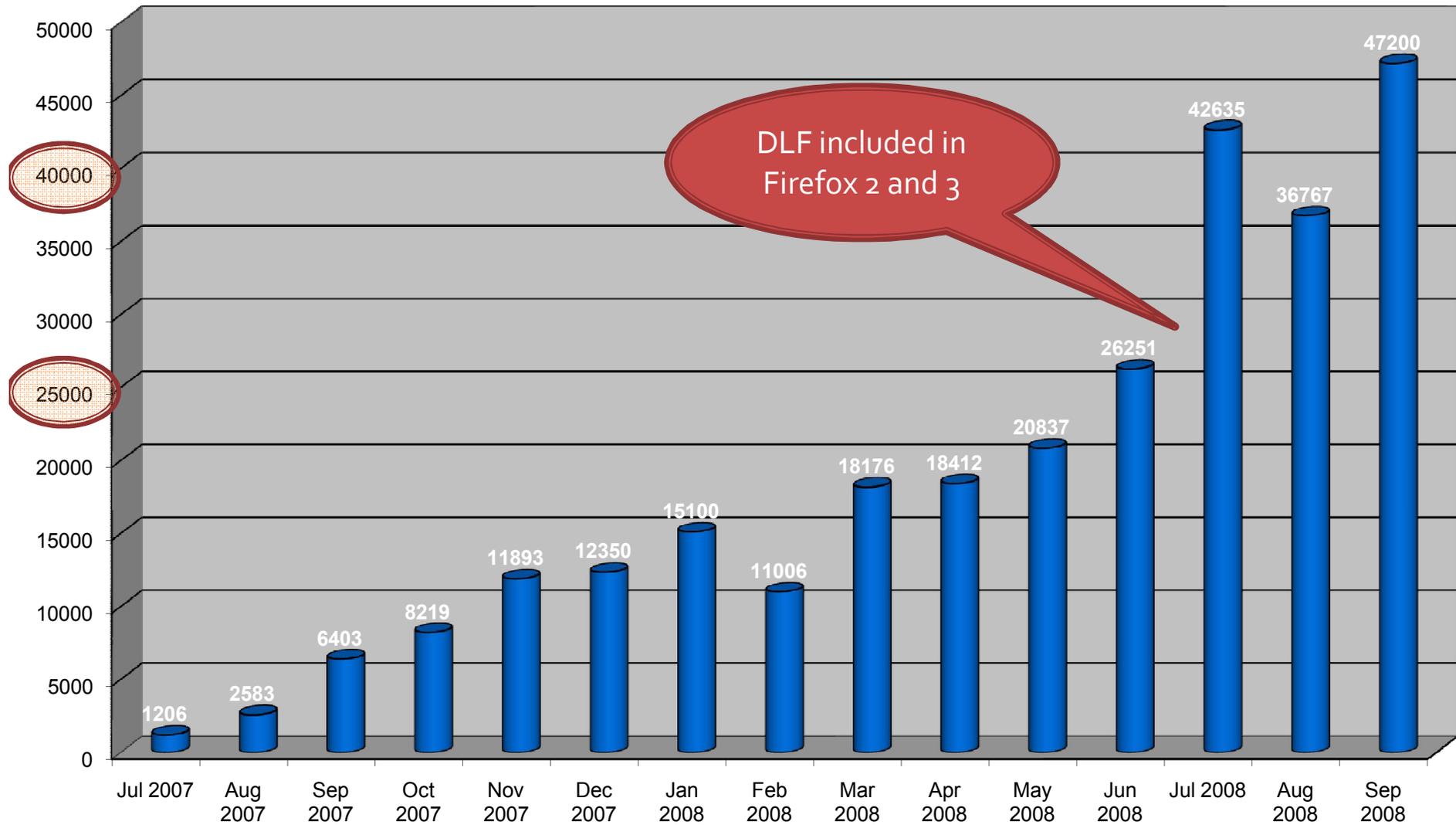


# Promotion of Polish digital resources

- June 17th, 2008 – Firefox 3.0 released with DLF plug-in included in the Polish version
- July 3rd, 2008 – Firefox 2.0.15 released with DLF plug-in included in the Polish version
- Resources of Polish digital libraries available for Polish internet users together with Google, Wikipedia etc.
  - Currently Firefox 2 and 3 have together 41% of web browsers market in Poland
- Measurable result?

# Promotion of Polish digital resources

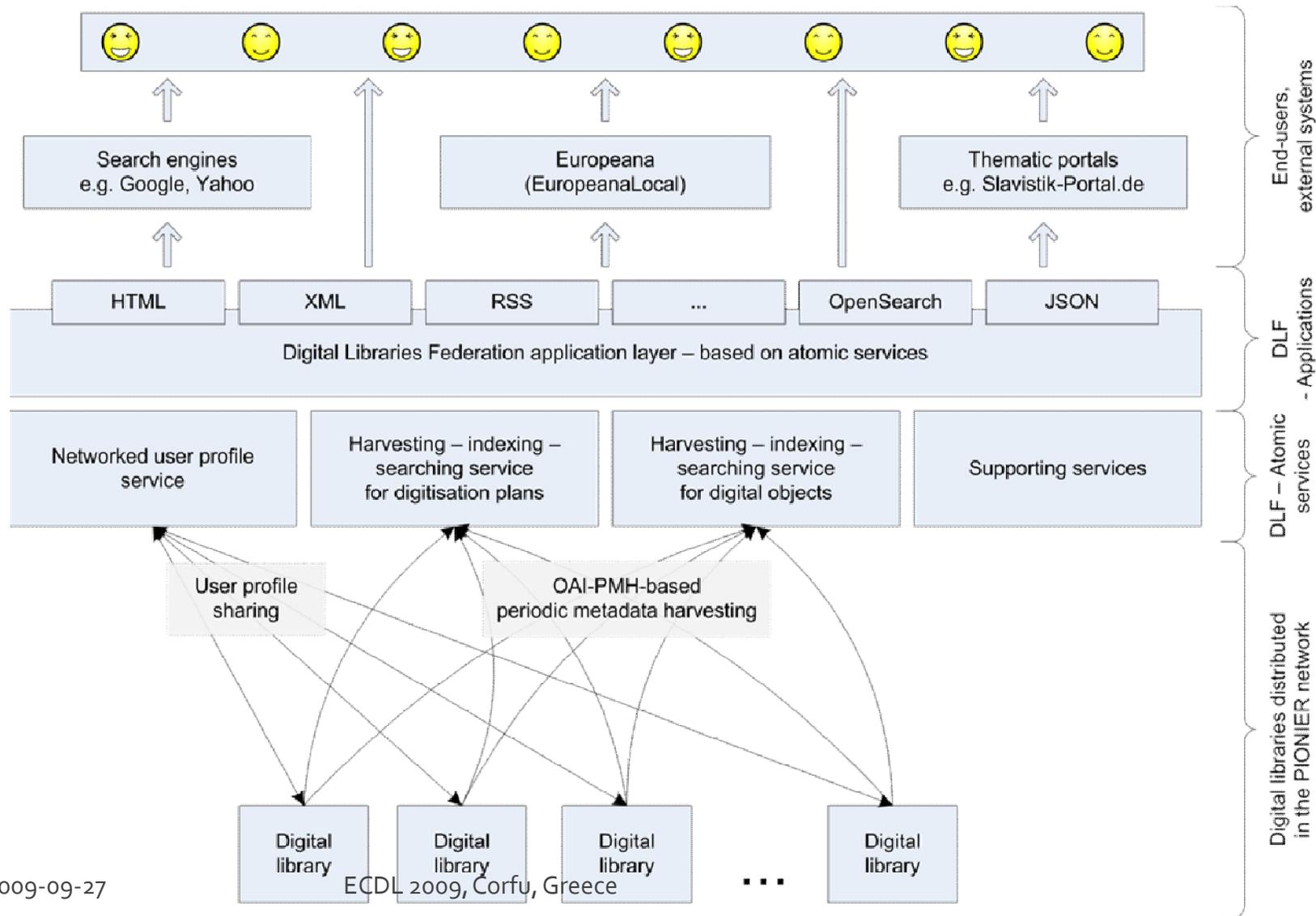
## - Total number of visits in the DLF each month



2009-09-27

ECDL 2009, Corfu, Greece

# The Digital Libraries Federation architecture



# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# Open standards in the DLF

- To join the DLF a digital library must have the OAI-PMH interface
- Since January 2008 this protocol was used in Poland only by the dLibra-based digital libraries
- In 2008 four digital libraries not based on dLibra created OAI-PMH interfaces and joined the DLF
  - **Cracow University of Technology Digital Library**  
(in DLF since 06.02.2008)
  - **Academic Digital Library ABC**  
(in DLF since 21.05.2008)
  - **ICM – DIR – Polish resources**  
(in DLF since 25.07.2008)
  - **Digital Library of the Diocese Archive in Tarnów**  
(in DLF since 22.10.2008)

# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# Automated coordination of digitization in Poland

- Supported via
  - Possibility to search the digitization plans
  - Report with the digitization plans
  - Report with the list and a summary of potential duplicates already existing
  - Prevention of the duplicated digitization efforts
    - DLF via dedicated API allows to check if there already are (digitized or planned) objects being potential duplicates to an object with the given description (metadata)

# Automated coordination of digitization in Poland

- Prevention of duplicated digitization efforts on the stage of digitization planning
  1. Editor of a digital library prepares the metadata of an object planned for digitization
  2. The digital library system submits this description to the DLF and checks if there are objects with similar metadata already digitized or planned for digitization in other institutions
  3. On the basis of such information the editor can remove particular position from its digitization plans

# Infrastructure of digital libraries

- Most important aspects
  - Strongly distributed infrastructure
    - Around 40 digital libraries available now
  - Single access point for end-users and external services
    - Digital Libraries Federation
  - Heterogenic systems
    - Open communication standards
  - Coordinated management of resources
    - Coordination of digitization
  - Common services
    - Networked user profile

# Networked user profile

- Digital libraries belonging to the DLF may trust each other and share the end-user profile
- Readers can log-in to several digital libraries having the account in just one digital library
- With this mechanism, having just one username and password, the end-users can use their profile data across several digital libraries
  - Eg. One list of favourites books from several digital libraries

# Networked user profile

- Technical realization
  - Based on the Shibboleth project developed as a part of the Internet2 programme in the US and similar European project named Guanxi
  - SAML open protocol is the communication basis
  - The Shibboleth model (Identity Provider and Service Provider) was extended by a Profile Provider

# Summary

- The last few years was a period of intense development of the digital libraries infrastructure in Poland
- Now it is time to connect this infrastructure to appropriate European services
- In this tutorial we will show our experience for the preparation to such connections
- We hope that you will find it useful