

Digital Library of Wielkopolska: Storage issues in the largest Polish digital library

Marcin Mielnicki

marcinm@man.poznan.pl

Poznań Supercomputing and Networking Center

Poznań, Poland

7th TERENA TF-Storage Meeting 9 September 2010, Poznań, Poland

Agenda

- Goals of digitization
- Functions of digital library
- Digital libraries in Poland
- PSNC activities in the digital libraries area
- Digital Library of Wielkopolska
- Digitization workflow
- Future plans

Goals of digitization

- Access

- > over the Internet
- > search capability
- > for learning, teaching, research

- Long-term preservation

- > protection against deterioration or damage

Functions of digital library

- Storage
 - > various digital objects
 - > metadata
- Management
 - > structure: grouping and collections
 - > metadata
 - > rights management
- Presentation
 - > browsing/searching
 - > viewing
- Communication/Interoperability
 - > search engines
 - > digital libraries
 - > other systems

Digital libraries in Poland

Overall number of digital objects

✓ over 450 thousands

Number of active digital libraries:

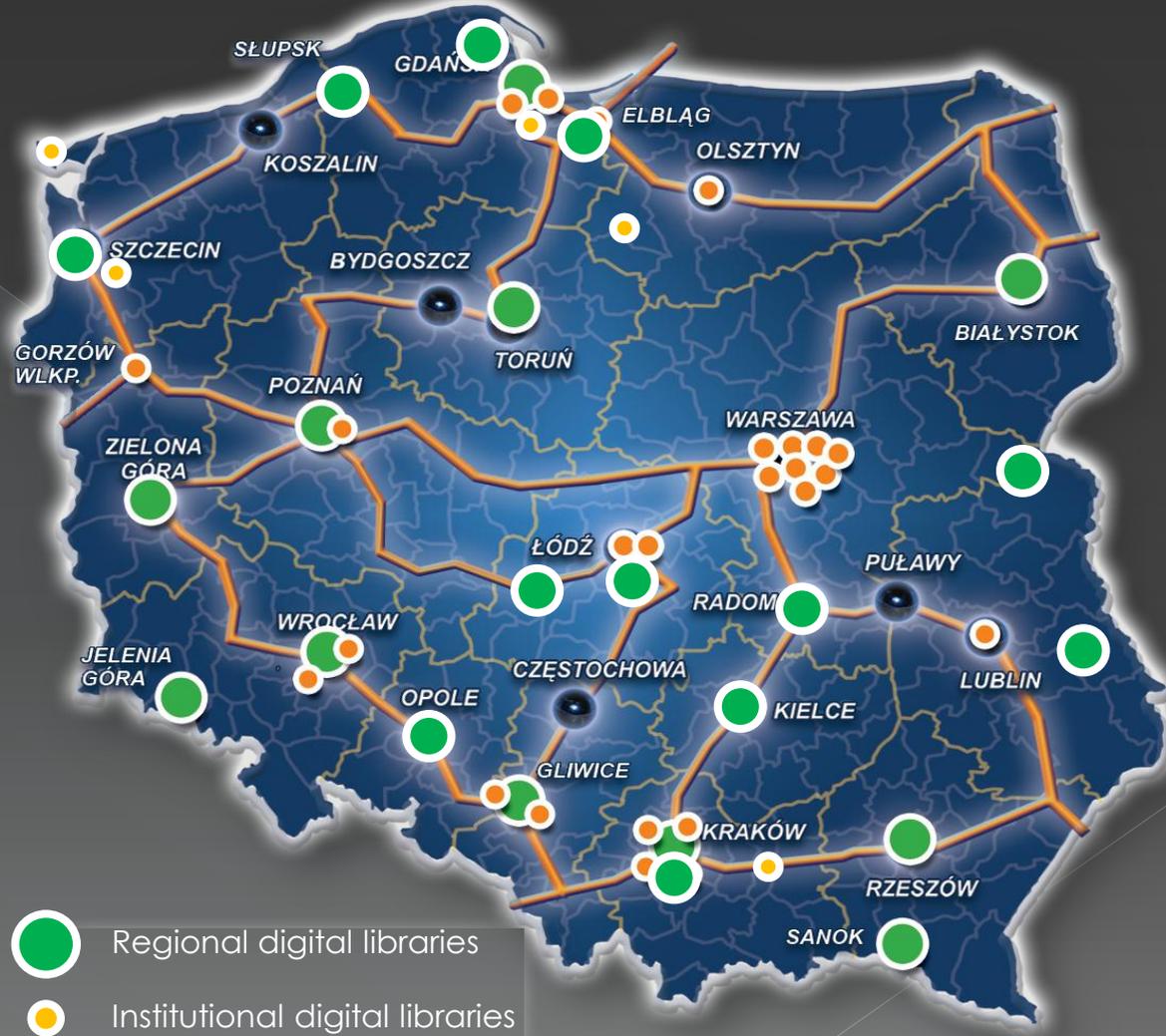
✓ 59 digital libraries

+ several other digital libraries

in the phase of planning, configuration or initial content uploading

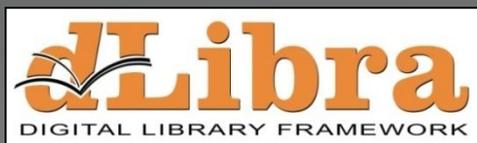
Number of cooperating institutions:

✓ ca. 300 universities and public libraries, archives, NGOs, etc.



PSNC activities in the field of digital libraries

- **dLibra** - software for building digital libraries (developed since 1999)
- **dMuseion** - software for building digital museums (developed since 2009)
- **PIONIER Network Digital Libraries Federation (DLF)** - integration point of majority of Polish digital libraries (on-line since 2007)



PSNC activities in the field of digital libraries

- **EuropeanaLocal** – making local and regional content accessible through the European Digital Library (2008-2011)
- **Interedition** – promoting the interoperability of the tools and methodology used in the field of digital scholarly editing and research (2009-2012)
- **Access IT** – works to Accelerate the Circulation of Culture through Exchange of Skills in Information and Technology (2009-2011)
- **IMPACT** – IMProving ACcess to Text (PSNC joined for 2010-2011)

Types of digital libraries

- Institutional
- Regional

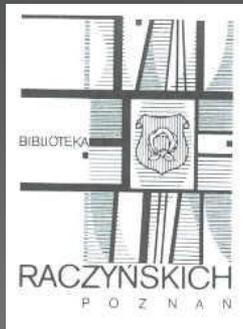


Digital Library of Wielkopolska

- Goal: to facilitate access to scientific publications, publications for education, resources for research and indirectly – protection by digitization
- Launched 1 October 2002
- Coordinated by
 - > Poznan Foundation of Scientific Libraries (<http://www.pfsl.poznan.pl/>)
- Hosted and technically supported by
 - > Poznan Supercomputing and Networking Center
- dLibra based

Participants

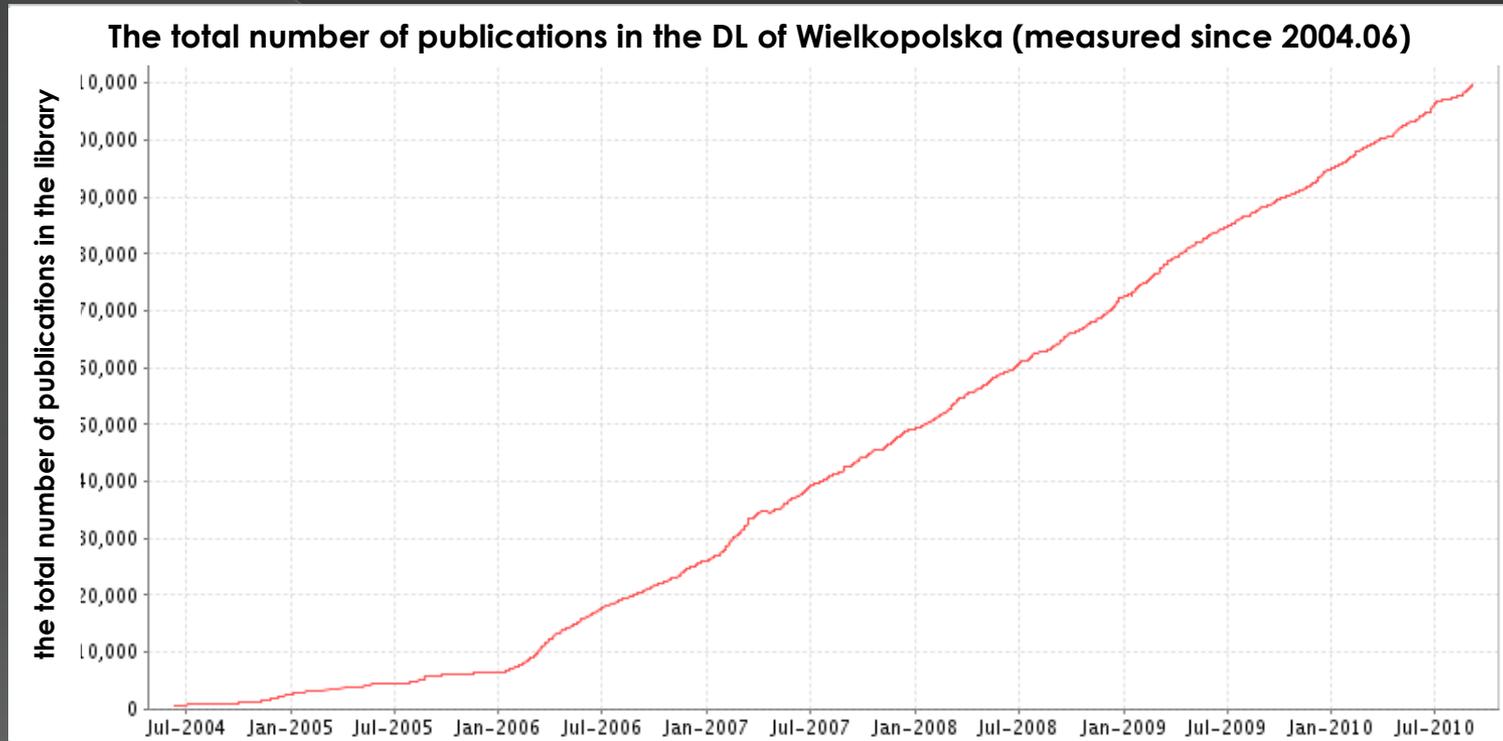
- 16 scientific and public libraries from region
- 40 editor accounts
- More than 250 contributing institutions and private persons



Content

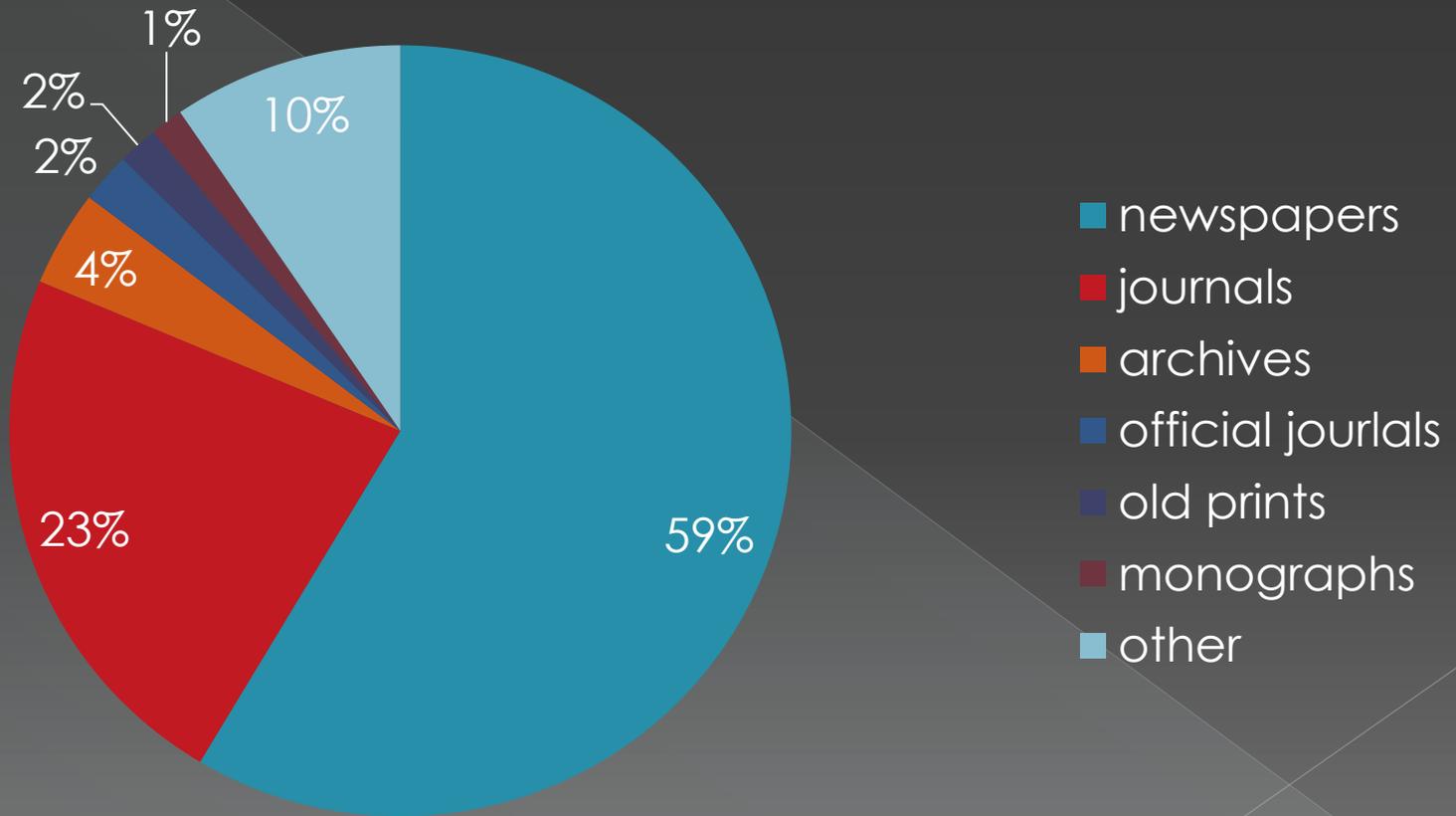
- 1 10 000 various digital objects
- 2 800 000 presentation files
- 780 GB presentation files
- DjVu (99%), PDF, HTML

Content



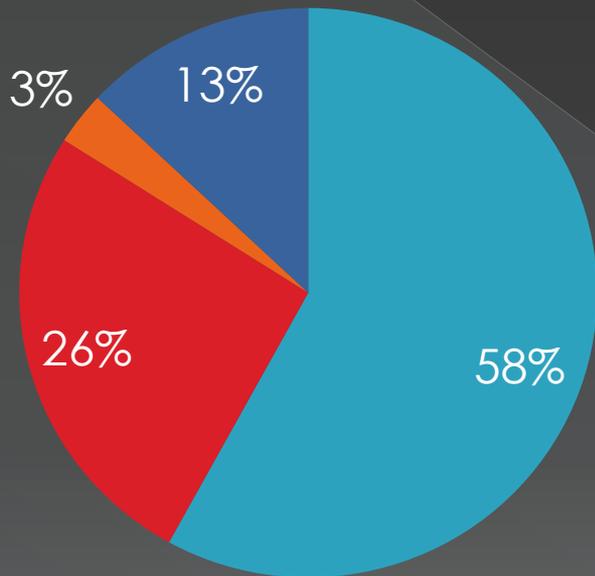
- 1800 new publications every month on average (since 2006)

Content - type of publication



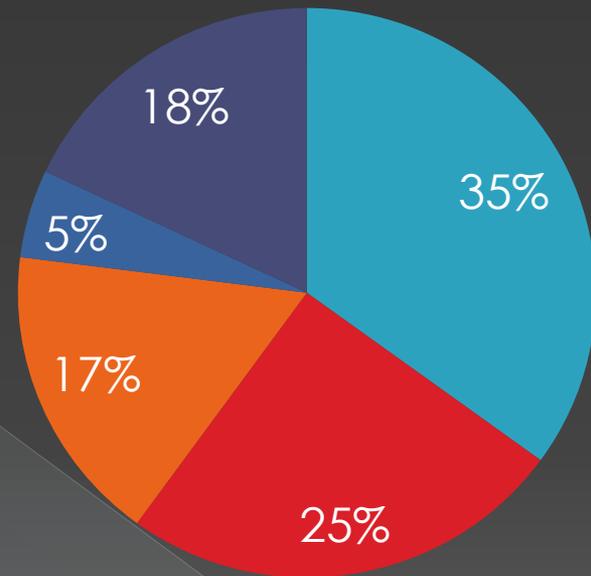
Right management statistics

Available publications



- Biblioteka Kórnicka PAN
- Biblioteka Uniwersytecka w Poznaniu
- Archiwum Państwowe w Poznaniu
- Other

Viewed publications



- Biblioteka Kórnicka PAN
- Biblioteka Uniwersytecka w Poznaniu
- Biblioteka AWF w Poznaniu
- Radostaw Muszkieta
- Other

Traffic stats (August 2010)

- 26 000 visitors
- 51 00 visits
- 61 000 searches
- 113 000 viewed publications
- 600 000 HTML interface pages generated by dLibra

Short demonstration

The screenshot shows the homepage of the Wielkopolska Biblioteka Cyfrowa. At the top, there is a logo of an open book and the text "WIELKOPOLSKA BIBLIOTEKA CYFROWA". Below the logo, there are navigation buttons: "readers on-line: 194", "MAIN PAGE", "COLLECTIONS", "NEW ACCOUNT", "LOGIN", and "CONTACT".

On the left side, there is a "Library" menu with the following items:

- ▼ Digital Library of Wielkopolska
 - ▶ Cultural heritage
 - > Educational materials
 - > Music notes
 - > Regional materials
- Kolekcje wydzielone:
 - > Abums
 - > The Unity of the Brethren Archive
 - > Archiwalia miast wielkopolskich
 - > Journals and newspapers
 - > Bibliografie online
 - > Jordaniana (Henryk Jordan's heritage)
 - > Directory
 - > Rozprawy doktorskie i habilitacyjne
 - > Polish Army
 - > Powstanie styczniowe w zbiorach Wielkopolskiej Biblioteki Cyfrowej
 - > Varia

The main content area features a search bar with the text "Digital Library of Wielkopolska" and "Network of Polish digital libraries". Below the search bar, there is a search button and a search icon. The search options are: "only in text", "only in description", and "in text and in description". There is also a link to "Advanced search..".

Below the search bar, there is a section for "Exhibitions" with the following items:

- Kolekcja pamiętników BUAM. Druga wojna światowa (1939-1945)
- 90 lat Uniwersytetu im. Adama Mickiewicza w Poznaniu
- Powszechna Wystawa Krajowa - 80 lat temu
- 90 lat Politechniki Poznańskiej
- Polonia Maior-fontes

There is a "More..." link below the exhibitions list.

At the bottom, there are two sections: "Recently added" and "Frequently read publications".

Recently added:

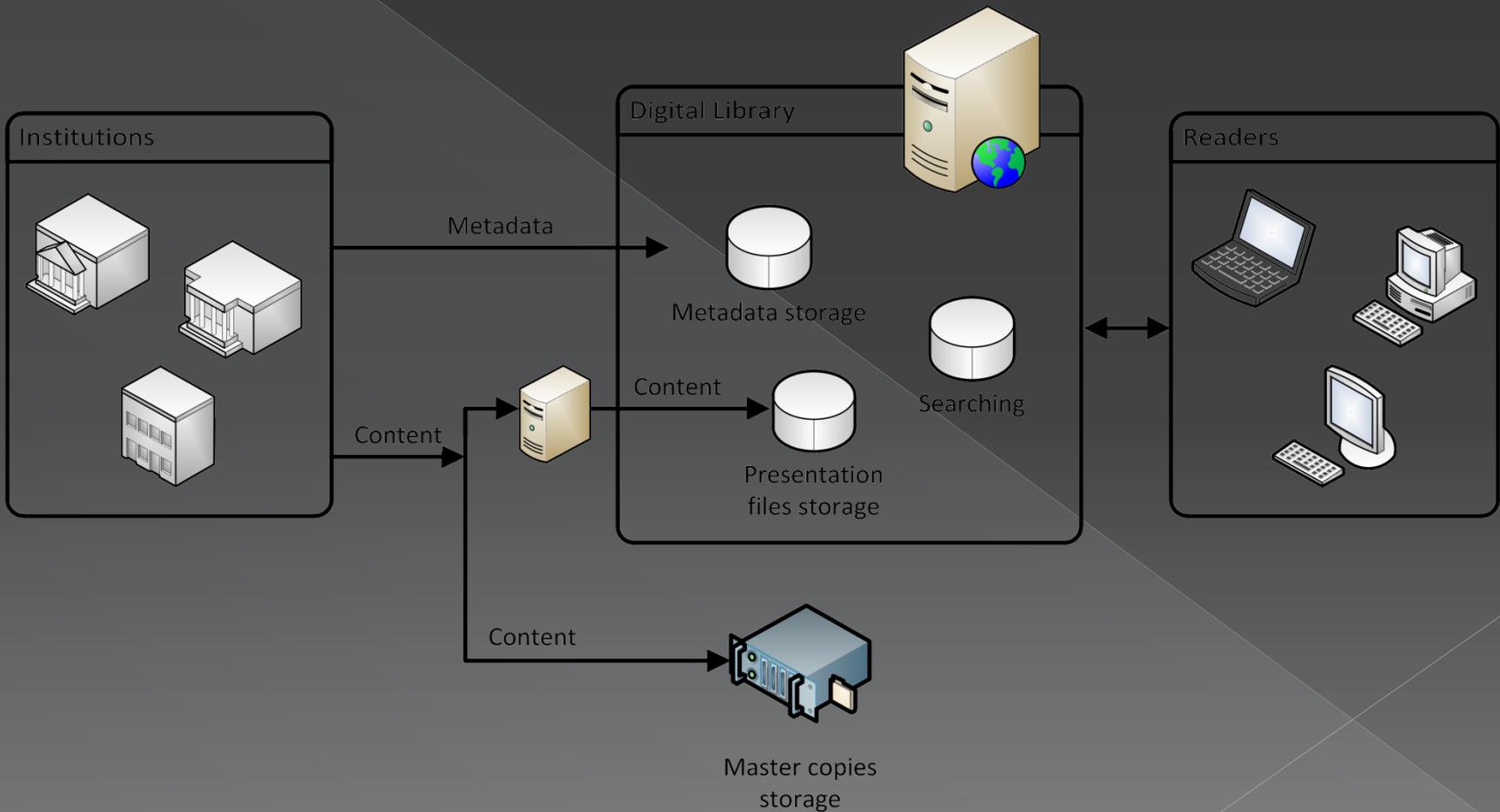
1. Wiadomości Brukowe. 1821 nr236
2. Wiadomości Brukowe. 1821 nr235

Frequently read publications:

1. Ocenianie osiągnięć uczniów przez nauczycieli wychowania fizycznego [395604]

<http://www.wbc.poznan.pl>

IT infrastructure



Digitization workflow

- Planning – choosing objects for digitization
- Creating metadata for planned publications
- Preparing digital version of objects
- Conversion master copies to presentation files
- Publication: content + planned publication



Planned publication

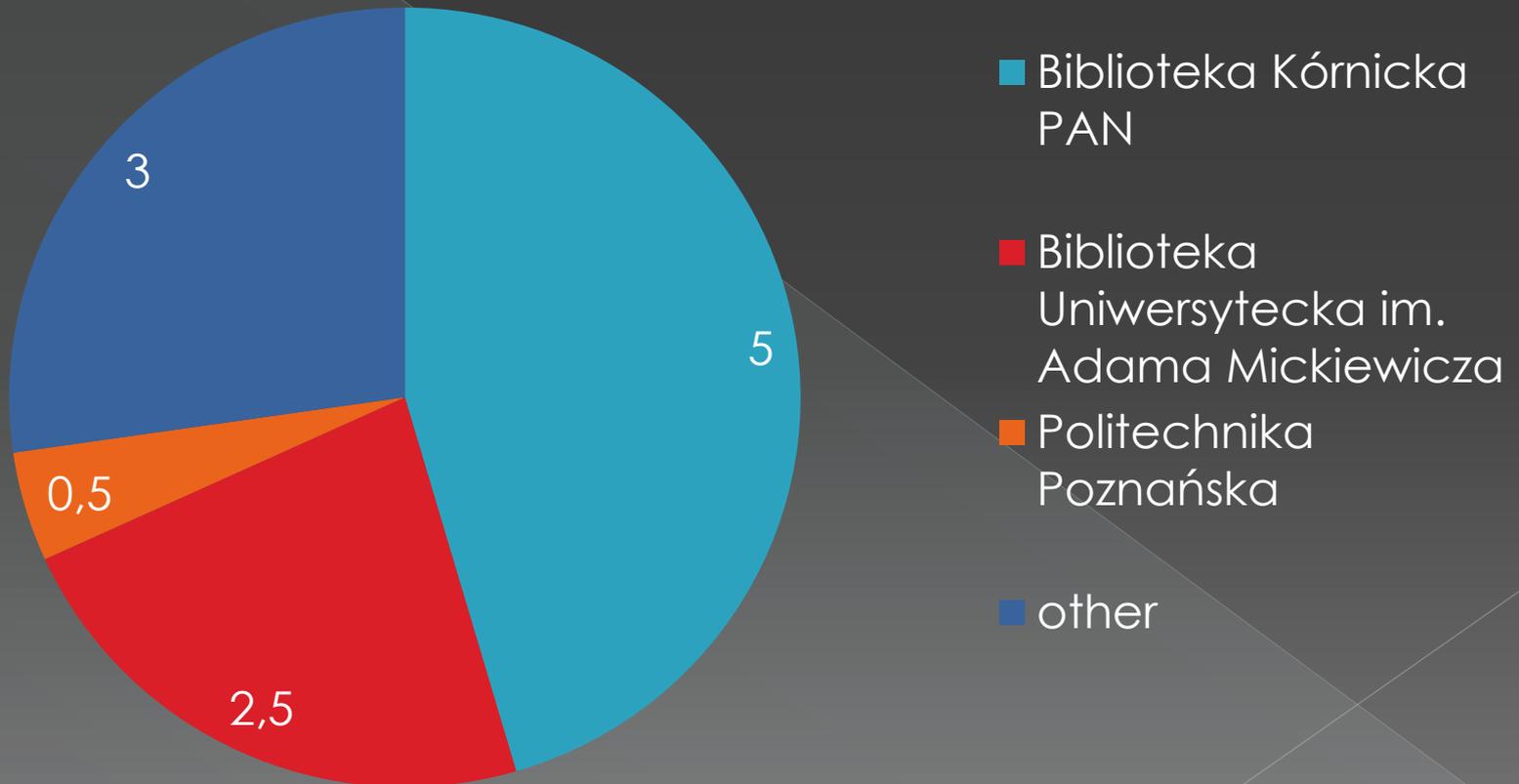
- ◉ Descriptive metadata
 - > created manually
 - > or imported from MARC (library catalogues)
- ◉ Persistent identifier

Master copies storage

- No single solution!
- Each participating library archives on its own
 - > portable HDD, DVD
- Master copies size - 10 TB
- Average increase of master copies size - 1,3 TB/year

Master copies storage

Master copies size (TB)



Creating presentation files

- Fully automated
- Conversion
 - > input: TIFF, JGP, GIF and PDF
 - > output: DjVu
- OCR
- DjVu INDIRECT multi-page format
- One virtual server
 - > DocumentExpress Enterprise

Storage

- ◎ SUN StorageTek 2540 FC 2540 Array
 - > 12 x 300GB SAS
- ◎ + extension array
 - > 12 x 300GB SAS
- ◎ 1 TB of 4,5 TB dedicated for WBC and currently used in 70%

Publication

- Planned publication + content (presentation files)
- dLibra system
 - > 4 virtual servers
 - > 2 Oracle databases (1 as a live backup copy)

Future plans

- More storage – ST2450 12x600 GB SAS (underway)
- Blu-ray based optical disc archiving
 - > for every participant!
 - > 370x50GB (18,5 TB)
 - > master copy archiving

Summary

- ◎ Digital Library of Wielkopolska
 - > the first and the largest Polish digital library
 - > regional – more than 250 contributors
 - > highly automated process of digitization
 - > in the near future solution for long-term preservation of master copies

Acknowledgements

- Thanks to Krzysztof Ober from Poznań Foundation of Scientific Libraries for preparation of materials for presentation

Thank you for your attention. Any questions?

Digital Library of Wielkopolska
<http://www.wbc.poznan.pl>

Marcin Mielnicki
marcinm@man.poznan.pl
Poznań Supercomputing and Networking Center
Poznań, Poland