



Digital Libraries and Digital Preservation

Marcin Werla

Poznan Supercomputing
and Networking Center

Poznań Supercomputing and Networking Center

- Established in 1993
- Affiliated by the Institute of Bioorganic Chemistry, Polish Academy of Sciences
- Currently
 - 5 divisions
 - Over 250 employees
- Participated/participates in over 50 EU-funded projects



Poznań Supercomputing and Networking Center

- High performance computing center
- Center for security of computer networks and systems
- Poznań city network operator (POZMAN)
- Operator of the Polish Optical Internet PIONIER network
- Research and development center in:
 - Next generation networks
 - Grid systems and high performance computing
 - Portals and content management systems
 - PSNC Network Services Department
 - PSNC Digital Libraries Team (<http://dl.psnc.pl/>)



Digital Libraries



What is a digital library?

“A *digital library* is an **online collection of digital objects**, of assured quality, that are created or collected and managed according to internationally accepted principles for collection development and made accessible in a coherent and sustainable manner, **supported by services** necessary to allow users to retrieve and exploit the resources.”

IFLA Manifesto for Digital Libraries

<http://www.ifla.org/publications/ifla-manifesto-for-digital-libraries>



What is the mission of the digital library?

“The mission of the digital library is **to give direct access to information resources**, both digital and non-digital, in a structured and authoritative manner and thus to link information technology, education and culture in contemporary library service.”

IFLA Manifesto for Digital Libraries

<http://www.ifla.org/publications/ifla-manifesto-for-digital-libraries>

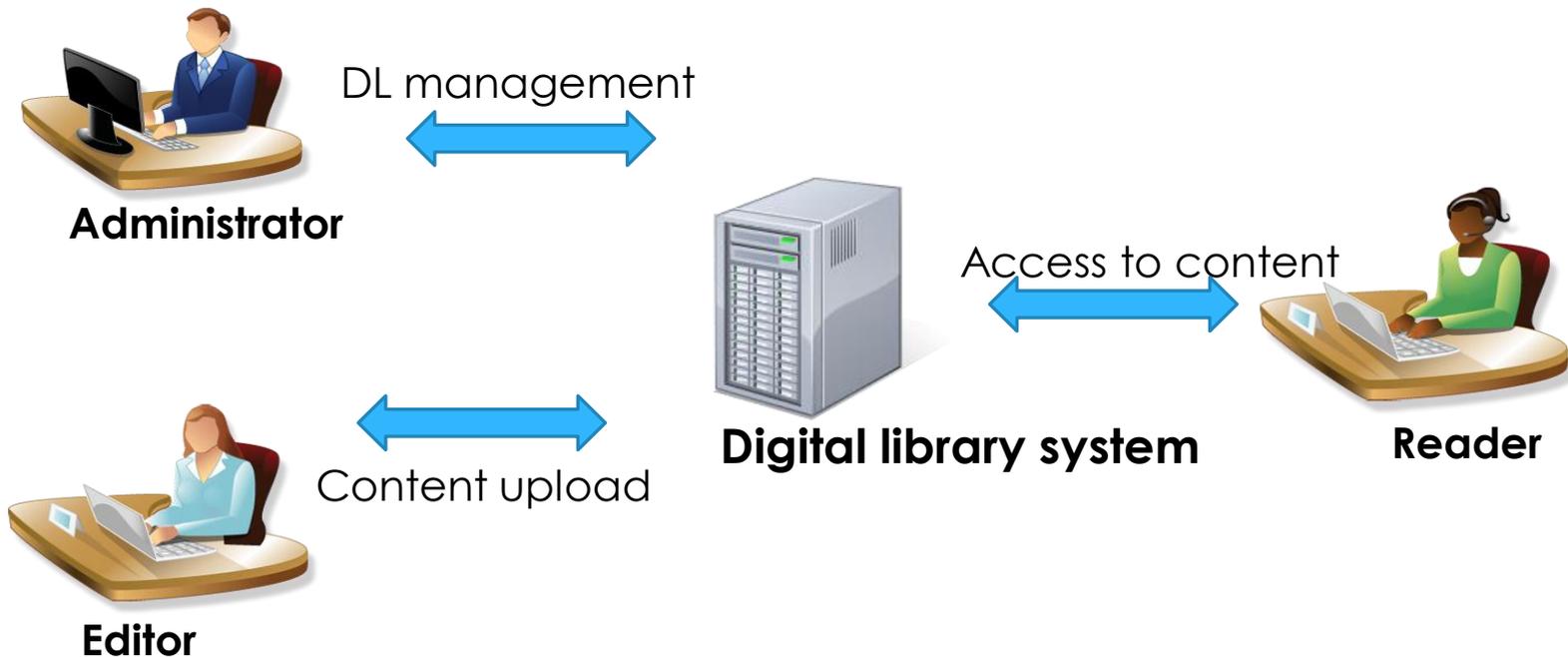


Benefits of the digital library

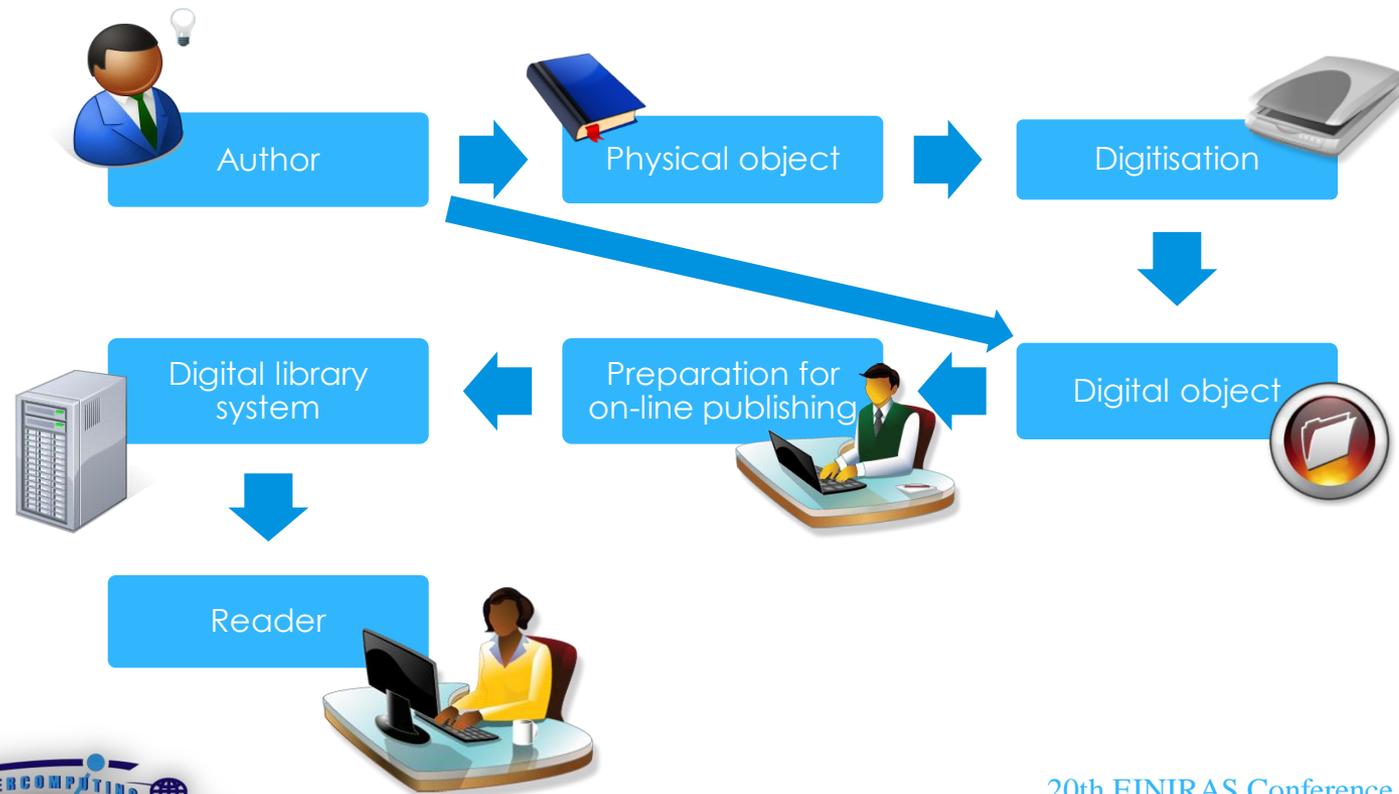
- Remote and continuous access
- Easier searching and browsing
- Shared access to resources
- Safety of the original object
- Easier information updates
- New forms of information
- ...



Users' roles in the digital library system



Basic digital library workflow



...digital and non-digital information resources...

- Physical objects have to be digitised
 - Different digitisation techniques for
 - different types of objects (text, images, audio/video, 3D)
 - different purposes
 - on-line access for “general users”
 - on-line access with focus on professional use (e.g. for research purposes)
 - preservation of the original object

...digital and non-digital information resources...

- Objects created in last few years are very often available in digital form since the beginning of their existence
 - So called “digital-born” objects
- In some cases the original digital form of digital-born objects was lost and such objects also have to be digitised

Preparation for on-line publishing

- Transformation of the digital object to its target delivery format
 - E.g. several TIFF files to one PDF document
- Enrichment of the original content
 - Optical Character Recognition (OCR)
 - Automated / possibly “dirty”
 - Corrected by humans
 - Speech to text (subtitles), calculation of MPEG features, ...

Preparation for on-line publishing

- Creation of digital object metadata
 - Descriptive metadata (title, creator, subject etc.)
 - Technical, structural and administrative metadata
- Metadata can be partially created automatically, imported from external information systems or created manually
 - Automated extraction of title, creator or keywords
 - Import of metadata records from library catalogues



Levels of description

What is the type? And dimensions?



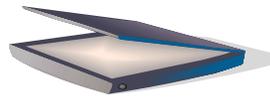
monument
10m x 3m x 3m



photograph
10cm x 15cm



image/jpeg
1200px x 1600px



made from stone,
carved, created to
praise...

monument or
person?

What to describe?



Preparation for on-line publishing

- Legal issues
 - Each published object should be accompanied with clear IPR statement
 - What is the license on which the object is published on-line?
 - Who is the holder of IPR?
 - A lot of things to think about
 - Strict copyright licenses vs. open licenses
 - Public domain
 - Orphan works
 - Open Access movement
 - ...

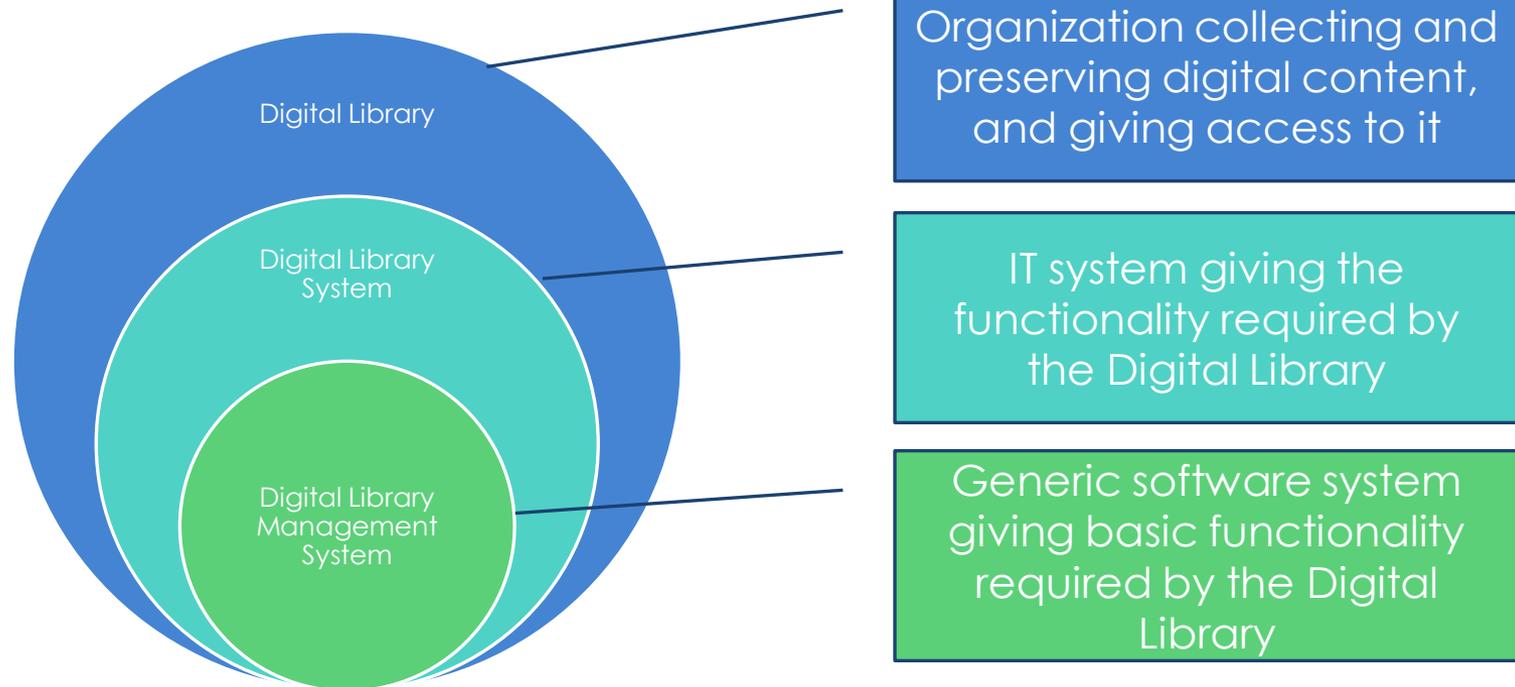


Publishing objects on-line

- Connecting digital objects with metadata and “services necessary to allow users to retrieve and exploit the resources”
- A number of digital library management systems to choose
 - Free, open source:
 - DSpace, EPrints, Greenstone
 - Non-for-profit, but not free:
 - dLibra
 - Commercial:
 - VTLS Vital, ALEPH DigiTool



DELOS Digital Library Reference Model



Digital library organizational models (in Poland)

- Regional digital libraries
 - One digital library system
 - One leading institution, many cooperating
 - Technical infrastructure and support often provided by local computing/networking centre
 - Significant amount of objects available in the digital library is related to particular region of Poland
 - Examples: Digital Library of the Wielkopolska, Silesian Digital Library



Digital library organizational models (in Poland)

- Institutional digital libraries
 - One digital library system
 - One institution
 - Responsible for entire content and technical infrastructure
 - Objects available in the digital library are related to history and present activity of the institution or collections owned by this institution
 - Examples: e-Library of the Warsaw University, Public Digital Archive of Agnieszka Osiecka



Digital library organizational models (in Poland)

- Mixed model
 - One digital library system
 - One leading institution, many cooperating
 - Technical infrastructure and support provided by the leading institution
 - Different reasons for cooperation
 - Thematic scope (Maritime Digital Library)
 - Institution profile (Digital Library of the Institutes of the Polish Academy of Sciences)



Digital libraries in Poland

Overall number of digital objects:

- ✓ over 460 thousand

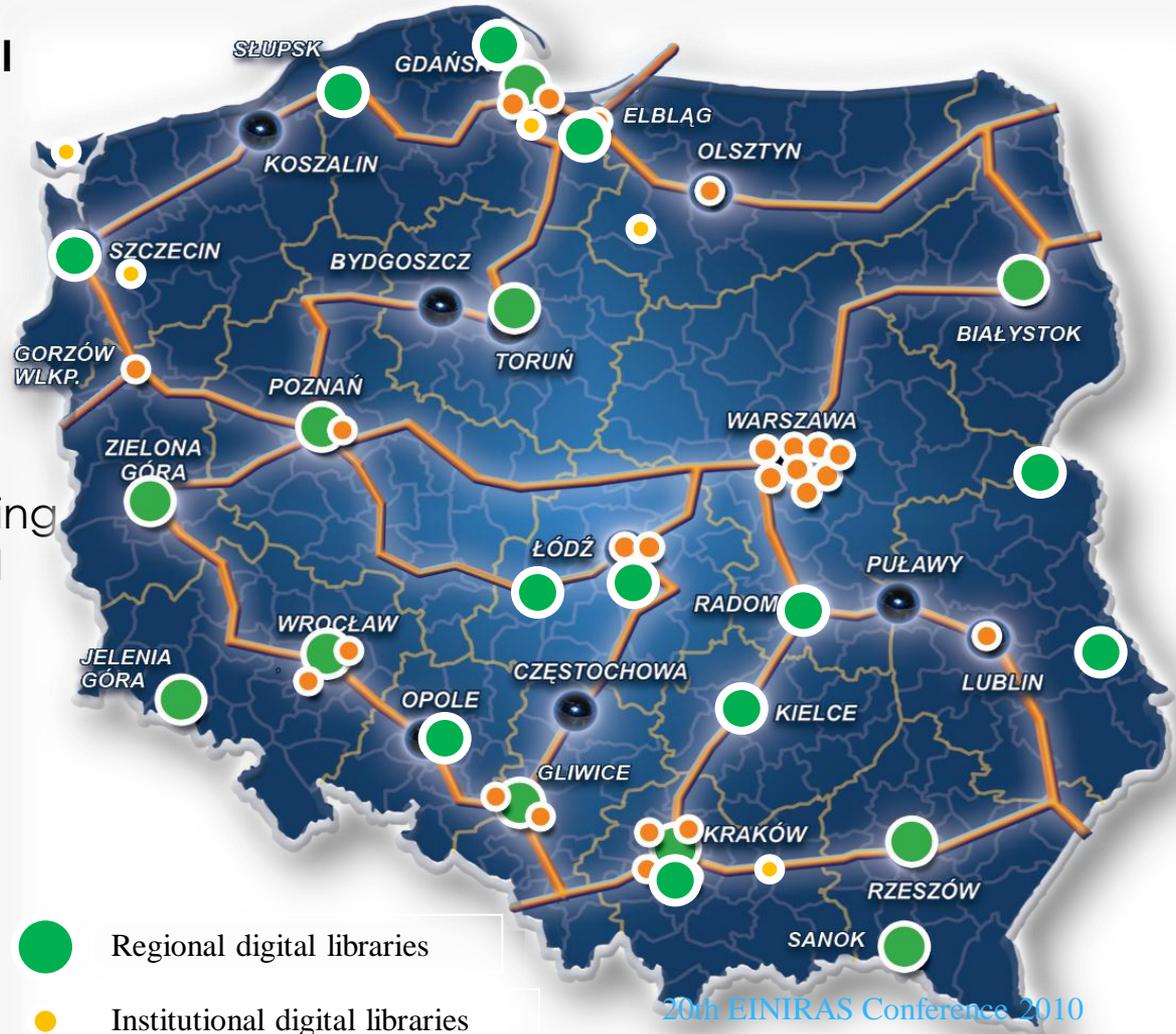
Number of active digital libraries:

- ✓ 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.



PIONIER Network

Digital Libraries Federation

- Internet service available publicly since June 2007
- Collects descriptions (aggregates metadata) of objects from Polish digital libraries
- Based on open communication standards
- Information updated each night
- Created, maintained and developed by PSNC
 - <http://fbc.pionier.net.pl/>





Wg opisu Wg identyfikatora Zaawansowane

Opis publikacji Szukaj

publikacje dostępne publikacje planowane

[Wybierz biblioteki cyfrowe](#)

Dostępnych publikacji: 421 399

Planowanych publikacji: 9 825

Zasoby "Gdańskiej Biblioteki Cyfrowej"

Od dzisiaj udostępniamy dane na temat publikacji z "Gdańskiej Biblioteki Cyfrowej". »

[20th EINIRAS Conference 2010](#)

Federacja Bibliotek Cyfrowych, 2010-06-21

PIONIER Network

Digital Libraries Federation

- Basic functionality
 - Search in the aggregated metadata
 - Digitisation plans
 - Searching
 - Reports
 - Access via API
 - Resolving of OAI identifiers of objects from connected digital libraries
 - Database of Polish digital libraries
 - <http://fbc.pionier.net.pl/owoc/libs-map>
 - Statistics and reports
 - Add-ons for the promotion of the DLF and connected libraries
 - Exposing aggregated metadata to other services



EUROPEANA



Europeana

- One of the key initiatives of the European Commission
- Portal which gives access to cultural heritage of Europe
- Information aggregated from:
 - Museum
 - Archives
 - Libraries
 - Audiovisual collections



20th EINERAS Conference 2010



Europeana

- First prototype of the portal was made available on the 20th of November 2008
- At the moment Europeana gives access to 13 million of objects distributed across entire Europe
- Europeana is a “metadata directory” with thumbnails and pointers to original objects
 - The access to full content is made via source digital libraries

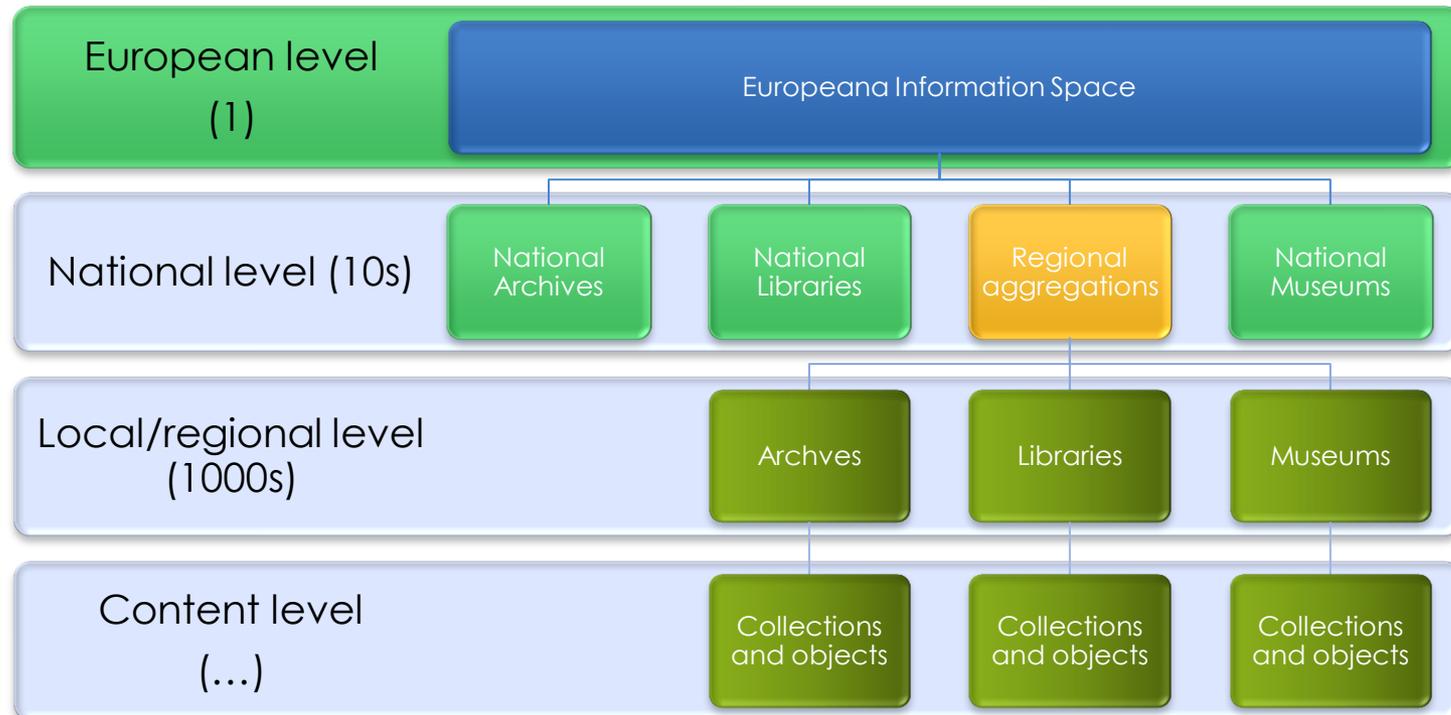


Europeana:

<http://europeana.eu>

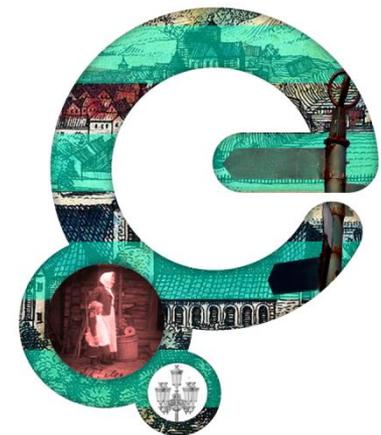


Structure of Europeana content sources



EuropeanaLocal

- European project under the eContent*Plus* program
- Duration – 3 years
 - Since June 2008 to May 2011
- Project type
 - Best Practice Network



20th EINIRAS Conference 2010
europeana
local



Main aims

- Improvement of the interoperability of digital content
 - Automated reuse
 - Creation of regional aggregations
- Creation of a network of regional repositories being able to communicate with Europeana

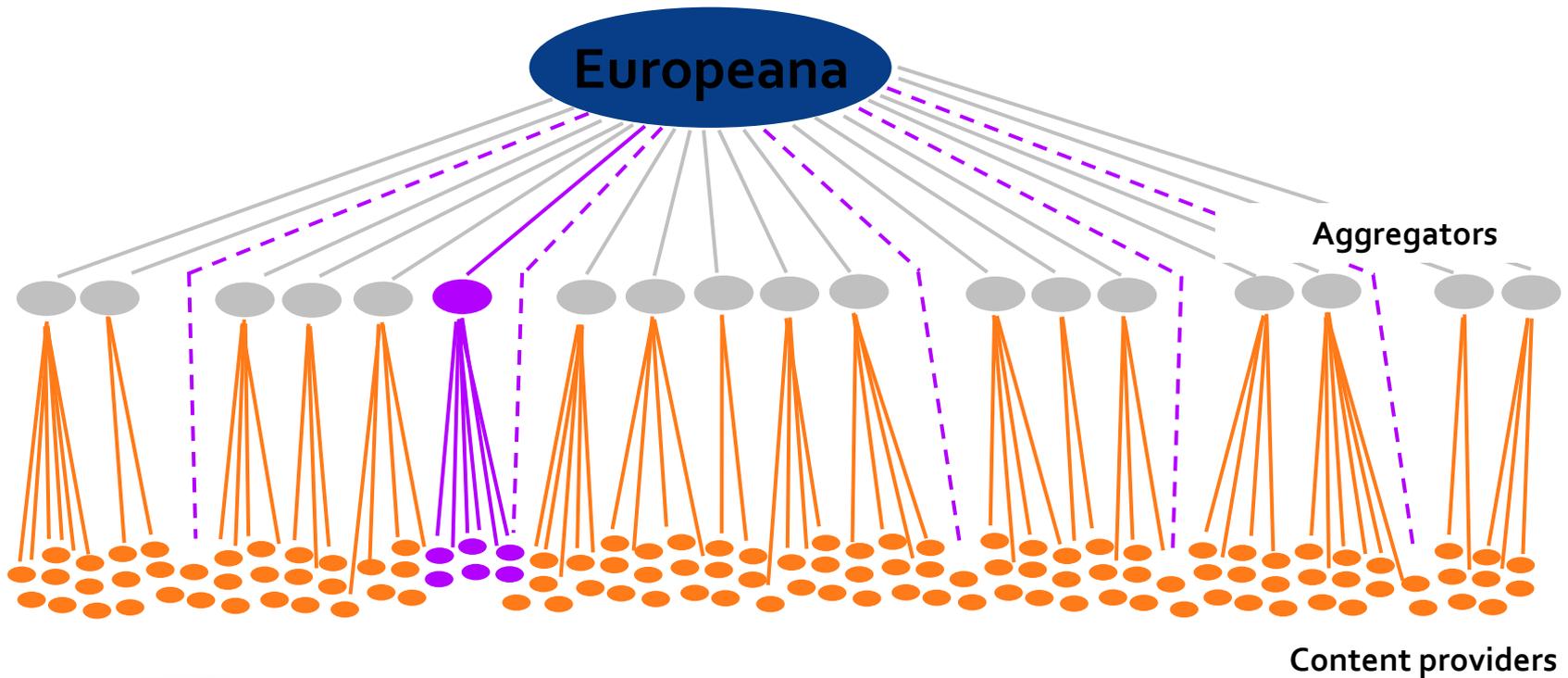


Participants (32)

- Coordinator
 - Sogn og Fjordane County Municipality (NO)
- Management support and scientific cooperation
 - MDR Partners (UK)
- Technical partners
 - EDL Foundation (NL) – main source of guidelines
 - Technical support (SK, 2 x NO)
- Country coordinators
 - AT, BE, BG, CY, CZ, EE, ES, DK, FI, DE, FR, GR, HU, IE, IT, LV, LT, MT, NL, NO, PL, PT, RO, SK, SI, SE, UK



Target model for Europeana content ingestion

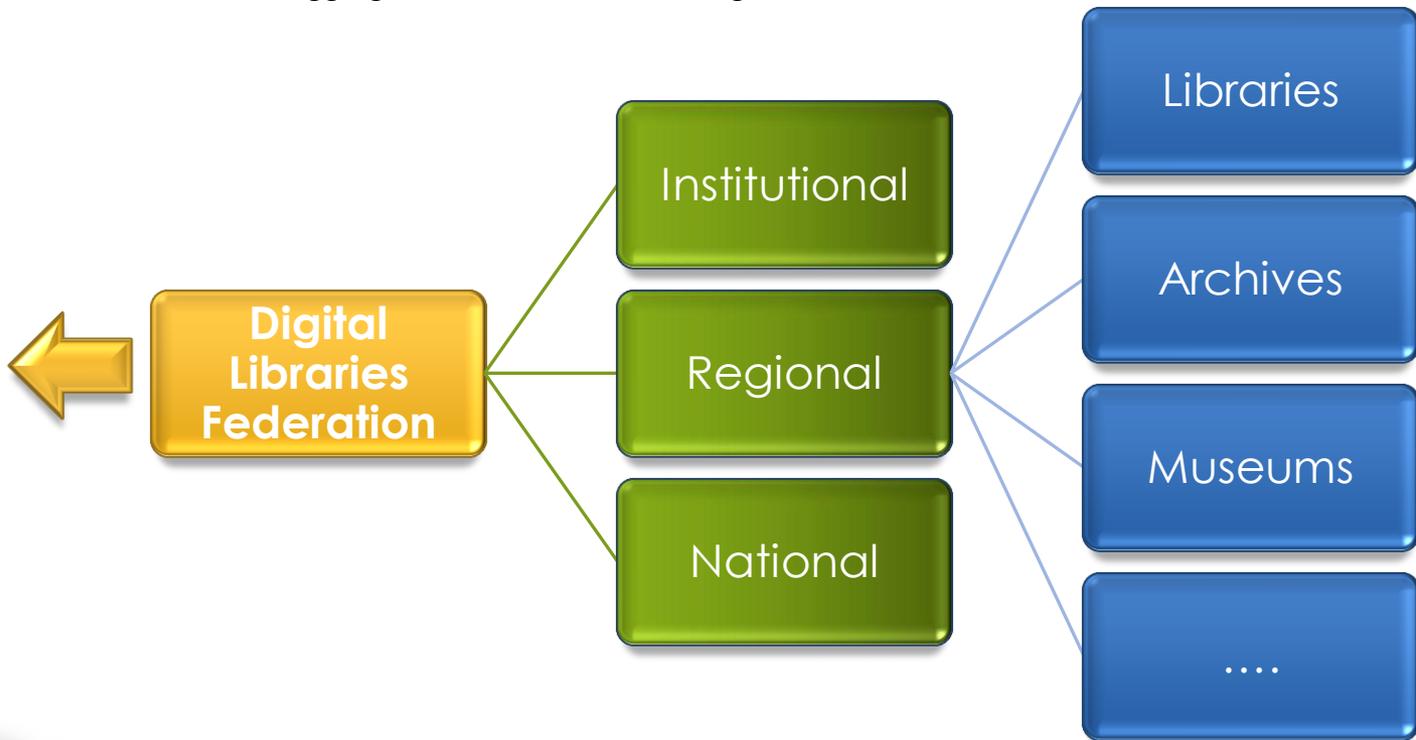


Federation as a metadata aggregator for Europeana

Metadata aggregator

Digital libraries

Institutions



Data from National Digital Library POLONA are sent to Europeana via The European Library.

20th EINIRAS Conference 2010

Digital Preservation



The Past is Prologue

- Digital Preservation encompasses a broad range of activities designed to:
 - extend the usable life of computer files
 - protecting files from media failure
 - physical loss, and obsolescence.
- Information must be intact and readable whenever user needs it



The Past is Prologue

- Mentioned accessibility can be divided to:
 - Content renderability
 - Content can be viewed by humans or processed by computers
 - Understandability
 - Content can be interpreted by humans
- This implies main issues:
 - Bitstream preservation
 - Preservation of content, form, style and functionality
- There is also an issue of authenticity of information



Digital Preservation

- There is no universal solution which could be used for all data types and situations
- There are many different content preservation elements
- The most important includes:
 - Bitstream refreshing
 - Replication
 - Technology preservation
 - Reliance on Standards
 - Migration
 - Emulation



Bitstream refreshing/copying

- Bitstream refreshing and copying is more commonly known as “backing up your data”
- Protects data from decay, media failure, malicious destruction etc.
- It should be considered as a minimum maintenance strategy

Replication

- Intention is to preserve documents through copying and the use of multiple storage locations
- Bitstream copying is a form of replication
- LOCKSS (Lots of Copies Keeps Stuff Safe)
 - Peer-to-peer data trading, open, free-market form of replication

Technology preservation

- Technology museum
- The idea is to preserve the technical environment that runs the system
 - Including media drives, original applications, OS
- It offers the potential of coping with media obsolescence
 - Assuming the media hasn't decayed beyond readability



Technology preservation

- Technology preservation is ultimately a dead end, since no obsolete technology can be kept functional indefinitely
-it is also very expensive



Reliance on Standards

- Information about format of a file are crucial for renderability
- What is in a file format specification?
 - e.g. the role of each byte in file header
- Specification is bare minimum, we will also need some software
- Without file format specification the only chance lies in digital archeology

Migration

- The goal of migration is to copy data, or convert data, from one technology to another preserving the essential characteristics of the data
- It is not always possible to make an exact digital copy or replica of an object
- Migration can deal with obsolescence of the physical storage medium, encodings and formats

Emulation

- Combines software and hardware to reproduce character of another computer, allowing old programs or media to operate in newer environment
- Emulation requires special software (emulators)

Emulation

- Win 7 is a good example of how useful emulation can be
- KEEP – Keeping Emulation Environments
 - <http://www.keep-project.eu>
 - Will develop an Emulation Access Platform to enable accurate rendering of both static and dynamic digital objects
 - Including text, sound, image files; multimedia documents, websites, databases, videogames etc.



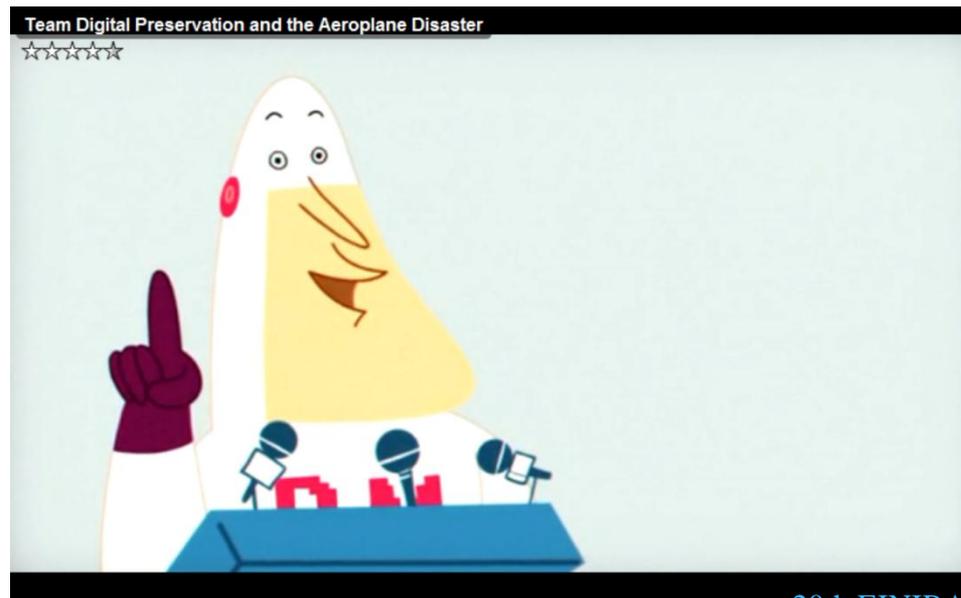
Sustainability

- Mentioned concepts and actions are only one aspect of the problem
- Another (maybe even more important) issue is sustainability of institution, repository and funding



Team Digital Preservation

- Look for it on YouTube 😊



Digital Libraries and Digital Preservation

Marcin Werla (mwerla@man.poznan.pl)
Poznan Supercomputing and Networking Center

