POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

SUPERCOMPUTING
POZNAŃ
NETWORKING

MISSION

CAPACITIES

COOPERATION

PEOPLE

IDEAS

INFORMATION SOCIETY

MILESTONES

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

ACCESS IT Plus – Training workshop in Veria (March '12)
Session 4: Digital libraries basics

Speaker: Marcin Werla (mwerla@man.poznan.pl)

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

ACCESS IT Plus – Training workshop in Veria (March '12)
Session 4: Digital libraries basics
Part 1: What should I know about digital libraries?

Speaker: Marcin Werla (mwerla@man.poznan.pl)

ACCESS IT+

What is a digital library for end user?

- Library gives users access to its collections of physical objects
    - borrowing (and returning)
    - reading in the reading room
    - is located in certain physical place
    - has certain opening times

- Digital library gives users access to its collections of digital objects
    - copying (no "returning")
    - reading on-line (optionally with some security constraints)
    - is located in virtual place (some WWW address)
    - is open 24/7

Physical object vs digital object

Function	Physical object	Digital object
Discovery	Catalogue card	Metadata record
Identification	Call mark	URI
Use	Physical content	Digital content

Digital library as an enterprise

- Digital library elements
  - Content
    - Digital objects and collections
  - People
    - Staff
    - Users
  - IT infrastructure
    - Hardware
    - Software
  - Organizational elements
    - Plans, procedures, guidelines, …
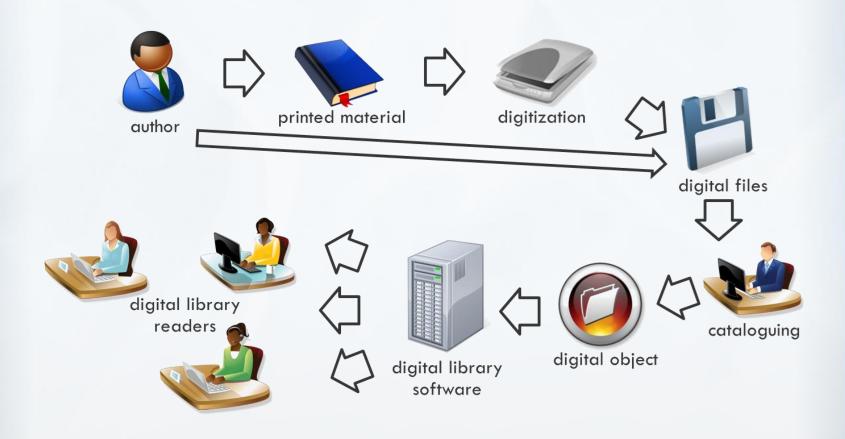
- "Cowboy coding" is not the best approach

How does a digital library work?

author → printed material → digitization → digital files → cataloguing → digital object → digital library software → digital library readers

What are the benefits
of using a digital library?

- 24/7 availability of digital resources

- Easy browsing and searching (incl. content)

- Preservation of the digital materials

- Easy update of the information

- New forms of information

- Automated information exchange with external software systems
  - Why this is important?
    - Send the data where the users already are

Who interacts with a digital library?

According
to OAIS model:

- Producers – specialized staff who is responsible
  for submitting new digital objects to a digital library
- Consumers – also called readers,
  with no real duties, use materials available in a digital library
  (searching, browsing)
- Management – defined strategy, rules and procedures
  that need to be followed in relation to a digital library

Examples of popular digital libraries

- Project Gutenberg (http://www.gutenberg.org/)
  - It is commonly presupposed that it was the first digital library ever created
  - Started by Michael Hart in 1971 with the digitization of the United States Declaration of Independence.
  - The mission is to encourage the creation and distribution of eBooks
  - Powered totally by volunteers

- World Digital Library (http://www.wdl.org/)
  - Initiative of UNESCO and Library of Congress, launched in April 2009
  - Mission: "The World Digital Library makes available on the Internet, free of charge and in multilingual format, significant primary materials from countries and cultures around the world"

- More examples are in the course

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

ACCESS IT Plus – Training workshop in Veria (March '12)
Session 4: Digital libraries basics
Part 2: Case study: Digital libraries infrastructure in Poland

Speaker: Marcin Werla (mwerla@man.poznan.pl)

Digital libraries in Poland: Case study

- Organizational models
  - Institutional digital libraries – single institution is giving on-line access to its resources
  - Consortial digital libraries – several institutions are sharing single digital library to give access to (selection) of its on-line resources
    - Scope of cooperation may be different
      - Regional
      - Thematic
      - Institutional
    - What are possible benefits of such cooperation?
      - Cost effectiveness
      - Exchange of knowledge
      - New ideas and projects
    - What are possible negative effects?
      - Lower institutional visibility
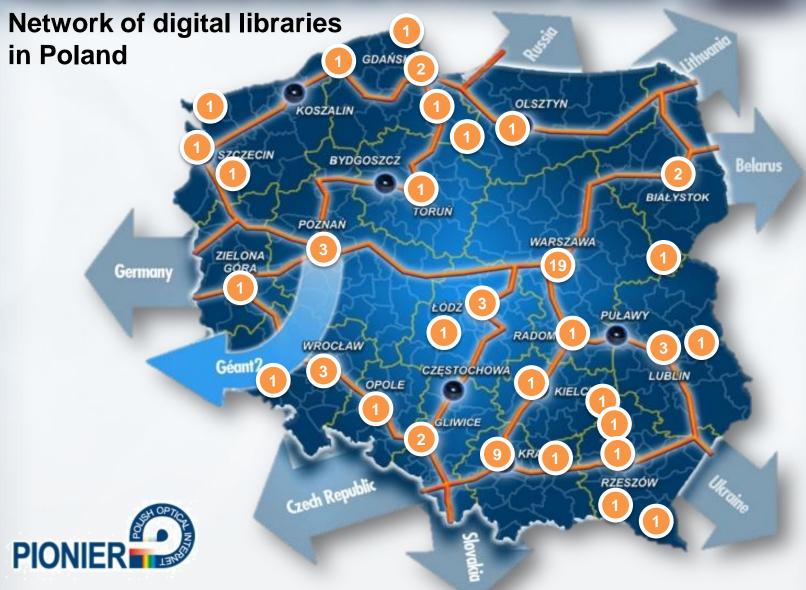   - Difficulties with significantly different kinds of collections

Digital libraries in Poland: Case study

- ## Short history
  - First works started in 1999
  - First digital library deployed in 2002
    - Wielkopolska Digital Library (http://www.wbc.poznan.pl/)
      - Regional digital library
      - The largest digital library in Poland (almost 160 000 objects)
  - Second digital library deployed at Wroclaw University of Technology in 2004 after digital libraries workshop organized by PSNC
    - Later it was transformed into Dolnoslaska Digital Library (http://www.dbc.wroc.pl/)
  - In 2005 first digital library supported from EU funds was launched: Kujawsko-Pomorska Digital Library (http://kpbc.umk.pl/)

Digital libraries in Poland: Case study

- Short history
  - In 2007 a Polish librarians community named Library 2.0 was established
    - Its discussion forum quickly became a place for knowledge exchange about digitization and digital libraries
  - In 2007 Polish metadata aggregator was launched – the Digital Libraries Federation
  - In 2009 cooperation with Europeana was established

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

Network of digital libraries in Poland

Digital libraries in Poland: Case study

- Examples:
  - Non-government institution:
    - Public Digital Archive of Agnieszka Osiecka
    - http://www.archiwumagnieszkiosieckiej.pl/dlibra
  - Institutional
    - e-library of the Warsaw Universtiy Library
    - http://ebuw.uw.edu.pl/dlibra
  - Regional
    - Pomeranian Digital Library
    - http://pbc.gda.pl/dlibra

Digital libraries in Poland: Case study

- Examples:
  - Thematic
    - Design Your Benefit

    http://biblioteka.zsz.com.pl/dlibra

    - Pedagogical Digital Library

    http://dlibra.up.krakow.pl:8080/dlibra/dlibra
  - Consortial
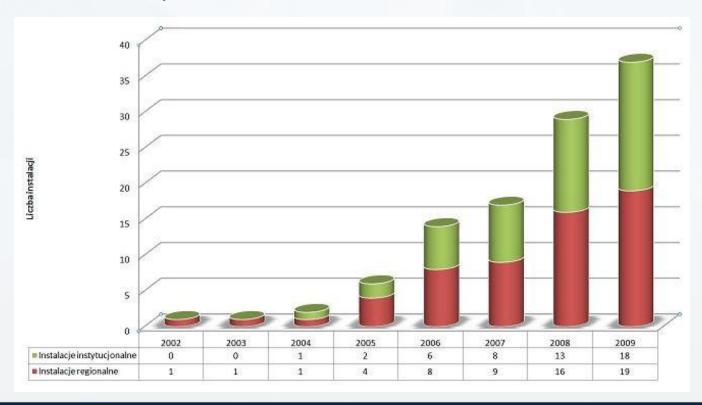   - Digital Repository of Research Institutes of the Polish Academy of Sciences

    http://rcin.org.pl/dlibra

    - Virtual Library of the FIDES Federation

    http://digital.fides.org.pl/dlibra

Digital libraries in Poland: Case study

- Polish Digital Libraries Federation – History
  - The number of digital libraries in Poland was growing quite intensively

	2002	2003	2004	2005	2006	2007	2008	2009
Instalacje instytucjonalne	0	0	1	2	6	8	13	18
Instalacje regionalne	1	1	1	4	8	9	16	19

Digital libraries in Poland: Case study

- Polish Digital Libraries Federation – History
  - The number of digital libraries in Poland was growing quite intensively
  - There was a need for single access point to resources of these libraries
    - Why Google was not enough?
      - Mix of high quality trusted content and… junk
      - Lack of deeper analysis of digital libraries data
  - In 2007 the Digital Libraries Federation portal was made available publicly by PSNC under http://fbc.pionier.net.pl/
  - Now it gives access to over 930 000 of digital objects from over 70 different digital libraries, museums and archives
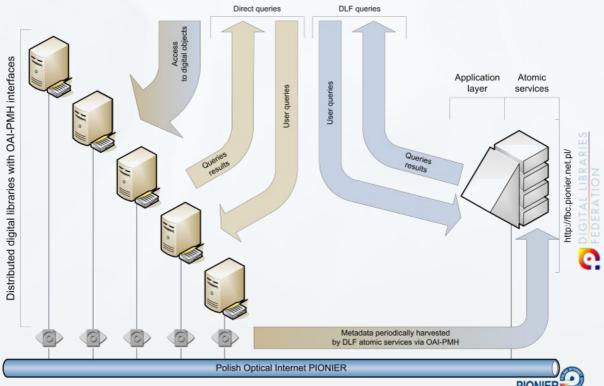   - The number of participating institutions is over 300

Digital libraries in Poland: Case study

- Polish Digital Libraries Federation – Functionality
    - Basic and advanced search
    - Search by identifier
    - Browse
        - Recently added
        - Recently displayed
        - Most frequently displayed
        - Recommended
    - Source of information about Polish digital libraries
    - Source of knowledge for Polish digital librarians
    - Coordination of digitisation

Digital libraries in Poland: Case study

- Polish Digital Libraries Federation – Architecture

Digital libraries in Poland: Case study

- Polish Digital Libraries Federation – Cooperation
  - DLF is a service designed to provide data both for end users and external services
    - Europeana

    http://europeana.eu/portal/search.html?query=*:*&qf=PROVIDER:Federacja+Bibliotek+Cyfrowych
    - DART-Europe

    http://www.dart-europe.eu/browse-results.php?dsource=27
    - Google

    https://www.google.pl/search?q=site%3Afbc.pionier.net.pl%2Fid
  - This makes life of Polish digital librarians easier ☺
    - Why?

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

ACCESS IT Plus – Training workshop in Veria (March '12)
Session 4: Digital libraries basics
Part 3: When should I create my own digital library and how to do it?

Speaker: Marcin Werla (mwerla@man.poznan.pl)

ACCESS IT

Building digital library

- Should I build my own digital library or join some existing one?
  - What requirements does the digital libraries building software have?
    - Is the institution rich enough to keep servers satisfying this need and a proper network connection?
  - What is the amount of materials, which are planned to be available in the digital library?
  - How many users will potentially be using the library?
  - Whether the institution do employ (or possess financial resources to employ) a qualified IT staff?

- What are the alternative options? Are there any public or commercial bodies I can join instead?

- Maybe I should establish new consortium and build digital library for it?

Building digital library

- IT infrastructure scalability – parameters to consider
  - The amount of users using the system simultaneously
  - The amount of publications in the digital library
  - The pace of releasing new publications
  - The size of publication files and number of those files
  - Whether the publications have good text information, which will be indexed (e.g. OCR)
  - The complexity of publication description and the level of minuteness of this description

  Example: http://kpbc.umk.pl/stats/

- The system load will increase in time and can radically change e.g. in association with events

- There is no easy formula to calculate what hardware should you buy

- So: what you can do when your servers are too busy?
  - Hardware side
  - Software side

Building digital library

- Hardware infrastructure –three basic costs variants
  - Fully local infrastructure – local server(s) in local server room
  - Co-located infrastructure
    - Your server in commercial server room
    - Hired server in commercial server room
  - Cloud infrastructure
    - E.g. Amazon S3 + EC^2
      - http://aws.amazon.com/ec2/pricing/
      - http://calculator.s3.amazonaws.com/calc5.html

Building digital library

- Choosing software
  - Digital library software
    - Greenstone (http://www.greenstone.org/)
    - Dspace (http://www.dspace.org/)
    - Eprints (http://www.eprints.org/)
    - dLibra (http://dlibra.psnc.pl/)
    - …
  - Operating system
  - Database

- Which criteria use to choose software components?
  - The most important one is: use those which you will be able to configure and maintain

Building digital library

- Planning, organizing, management
    - Identifying, planning and organizing all tasks which are necessary for activating and keeping the digital library
        - Start up stage
        - Reguar activity
    - Aspects which have to be covered include
        - staff
        - tasks division
        - areas of responsibility
        - work organization and efficiency
        - monitoring
        - risks
        - …
    - „Not plan but planning" – constant monitoring of plan execution and plan adaptations if necessary
    - THIS PART IS NOT COVERED BY OUR COURSES

Building digital library

- Designing information architecture
    - Metadata schema is a set of elements which will be used for describing digital objects
        - There are different types of metadata:
            - descriptive
            - administrative
            - technical
    - Administrative and technical metadata are less problematic in a sense that it is easier to define needs which should be fulfilled by such metadata
    - Aim of descriptive metadata is to allow discovery of digital object and initial assessment about particular object usefulness
        - Therefore descriptive metadata schema must take into account not only needs of librarians but also needs of users
        - Clear definition of such needs may problematic and such needs may evolve

Building digital library

- Designing information architecture
    - Definition of a metadata schema for use in digital library includes
        - Set of elements with definitions of their meanings
        - Set of rules e.g.
            - Which elements are obligatory
            - How to encode specific elements (creator, date, language etc.)
        - Set of guidelines e.g.
            - How to choose title if there is more than one (cover, title page etc.)
            - How to choose subject keywords
            - How to express size of the object
    - In case of digitisation: which object to describe original or digitized?
    - It is worth to publish such guidelines on-line for comments and collaboration

Building digital library

- Designing information architecture
  - Good starting point is to use existing well established metadata schema like Dublin Core Metadata Element Set or Dublin Core Terms
  - Is there possibility to import already existing metadata?
  - Multilingualism of metadata is also worth to consider

Building digital library

- Designing information architecture
    - Above the digital object level the information architecture should include
        - Complex objects (e.g. journals as a whole)
        - Collections
            - Especially useful for users which do not know what to search for
                - » How many of such users can be?
            - Different scopes of collection
            - One object may belong to several collections

Building digital library

- Selection criteria
  - There can be different selection criteria for materials to be placed in digital library e.g.:
    - Condition of objects
    - Popularity of objects
    - Cultural value of objects
    - Thematic scope of a digital library
    - Relation to some important event
    - Requirements of the digital library/digitization funding body

Building digital library

- Digital objects formats
  - Again selection criteria may vary
  - What to take into account
    - Users capabilities
      - What software users have installed on their PCs?
    - Users needs
      - What functionality users need? E.g.
        » Zooming
        » Rotating
        » Text searching
    - Digital library ecosystem
      - What formats are understandable by external systems?
        » Google Scholar

Building digital library

- ## Releasing digital library
  - Besides digital objects you have to prepare
    - Auxiliary materials – help pages, license information, about information
    - Possibility of sending feedback
  - VERY IMPORTANT: Remember about usability testing of the end user interface, preferably with the participation of end users representatives
  - Finally promotion is also an important aspect which should be planned, e.g.:
    - Leaflets, posters, gadgets, articles on thematic conferences, news releases
    - Official opening of a digital library with media and VIPs participation
    - Social networks like Facebook
    - Your local Wikipedia

Things to discuss

- Are there any existing digital libraries in participants' countries? What is the source of funding?

- What IT infrastructure is used in your libraries to handle digital libraries?

- Are users aware of digital libraries existence? What do they think about them? What are they used for?