# POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

# PREMIS and METS: dArceo case study

Tomasz Parkoła (tparkola@man.poznan.pl)
Network Services Department, PSNC

# Digital libraries infrastructure in the PIONIER network

- Developed since 2002
- Consists of around 100 digital libraries
- Collaborative effort of several hundreds of research and cultural institutions including local networking and computing centers
- Over 1.7M digital objects available on-line for the information society

# Toolset for Polish digital libraries

**Developed and provided by PSNC**

**tools**

dLibra
DIGITAL LIBRARY FRAMEWORK

dMuseion
DIGITAL MUSEUM FRAMEWORK

dLab
DIGITISATION WORKFLOW MANAGEMENT

dArceo
LONG-TERM PRESERVATION SERVICES

---

**projects**

SCAPE
SCAlable Preservation Environments

Improving Access to Text
IMPACT

Interedition

impact
digitisation.eu
centre of competence

ENRICH PROJECT

EUROPEANA LOCAL
connecting cultural heritage

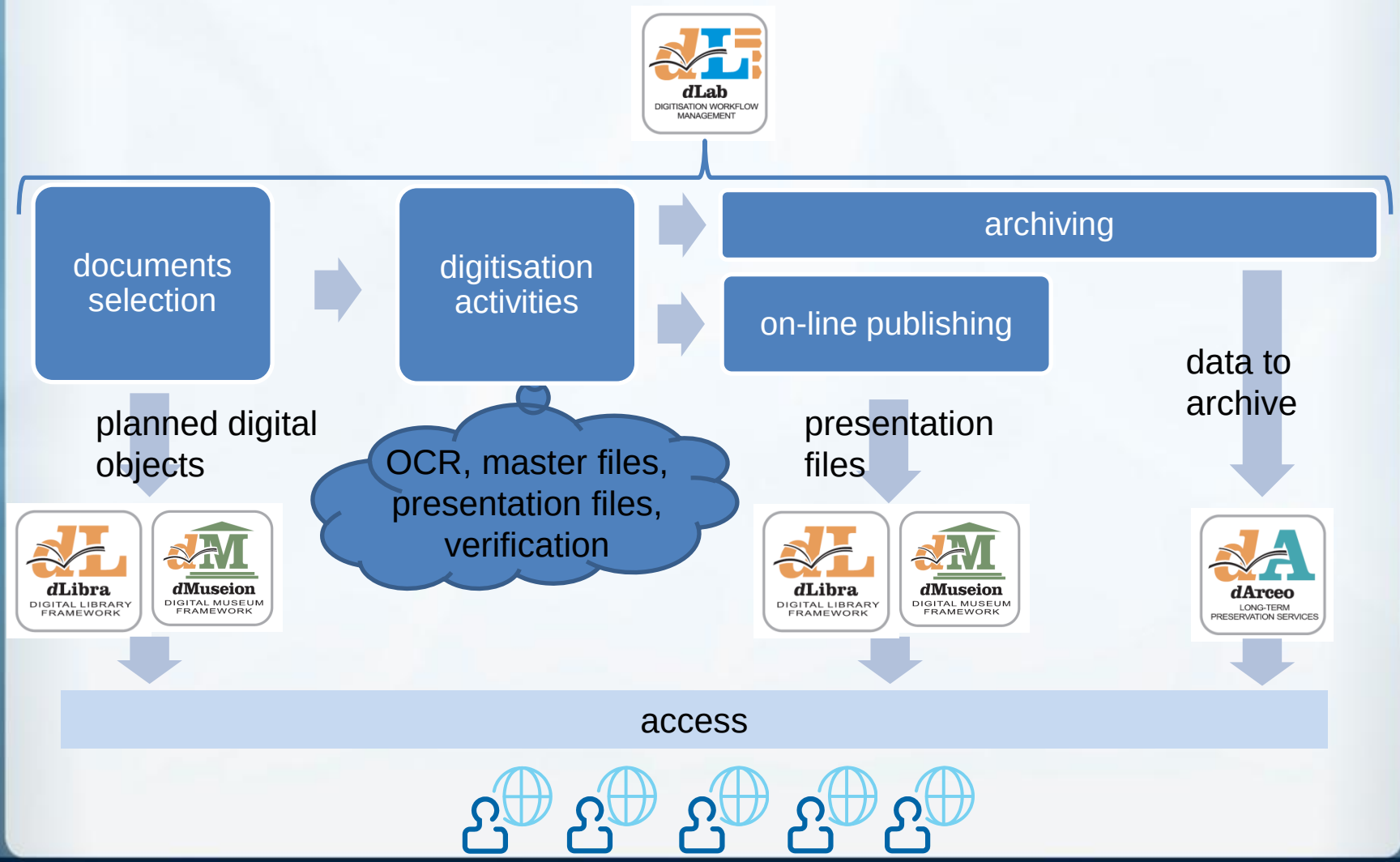
Workflow 4Ever

europeana
think culture

ACCESS IT

synat
system nauki i techniki

# Digitisation process



# dArceo

- Long term preservation system for images, text and a/v content
- Uses well know standards and formats
  - METS as a metadata container
  - PREMIS for recording events related to files and objects
  - TextMD, DocumentMD, MIX, AES57 for technical metadata
- Extendible in the context of migration, conversion and delivery services
- Used in production mode in Poland

# PREMIS in dArceo

- Object level PREMIS metadata are embedded into METS
  - OAI identifier
  - Relation to other objects
  - Associated events (e.g. creation, transformation)
  - List of all associated files in the AIP
- File level PREMIS metadata are located in separate XML files
  - Identifier in the context of object
  - Characterisation information
  - Associated events (e.g. creation, validation)

# PREMIS and METS issue

- Issue: technical metadata duplication
  - METS reference to external file
  - PREMIS characterisation section
  - PREMIS characterisation extension (external file)
  - Guidelines for using PREMIS with METS do not state clear rules
- Risk: inconsistency in the AIP
- Solution: clear guidelines, e.g. to use only one approach of all possible

# PREMIS issue 1

- Issue: technical metadata inconsistency within PREMIS
  - size
  - formatDesignation
  - formatRegistry
  - objectCharacteristicsExtension

# PREMIS issue 1

- Issue: technical metadata inconsistency within PREMIS
  - size
  - formatDesignation
  - formatRegistry
  - objectCharacteristicsExtension

# PREMIS issue 1

- Issue: technical metadata inconsistency within PREMIS
  - size
  - formatDesignation
  - formatRegistry
  - objectCharacteristicsExtension

# PREMIS issue 1

- Issue: technical metadata inconsistency within PREMIS
  - size
  - formatDesignation
  - formatRegistry
  - objectCharacteristicsExtension
- Risk: inconsistency in the AIP
  - If present content analysis is required in order to identify the correct metadata
- Solution: clear guideline to use only one approach of all possible

# PREMIS issue 2

- Issue: vocabulary for eventType
  - No vocabulary enforced, suggested to build own
  - Starter list proposed, but not aligned with OAIS terminology (migration instead of transformation)
- Risk: decreased interoperability and misunderstandings
- Solution: introduce vocabulary or align starter list with OAIS terminology

# POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

## Thank you!

tparkola@man.poznan.pl
http://dl.psnc.pl/

Poznań Supercomputing and Networking Center
affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences,
ul. Noskowskiego 12/14, 61-704 Poznań, POLAND,
Office: phone center: (+48 61) 858-20-00,
fax: (+48 61) 852-59-54,
e-mail: office@man.poznan.pl, http://www.man.poznan.pl