Preservation of Mathematical Heritage: Interoperability into the Future

Milena Dobreva
Senior Researcher (SHAMAN project)
Centre for Digital Library Research

Head, Digital Humanities Dept., IMI-BAS
TEXTE project fellow, The Moore Institute, NUIG

University of **Strathclyde**

CENTRE FOR DIGITAL
LIBRARY RESEARCH

Talk overview

1   Why digital preservation?

2   Who is active in this field?

3   What issues can be identified for DML?

4   Where to look for further info?

Milena Dobreva, KOM, DML-CZ, 12 June 2008

The technology changes…

- Hardware
- Software
- Environment
  - E.g. network addresses
- Standards (including file formats)

People also change…

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Example

Rhind Papyrus,
c. 1650 BC
"Correct method of reckoning, for grasping the meaning of things and knowing everything that is, obscurities and all secrets."

- table of the numbers 2/n, where n=3, 5, 7, ..., 101, all expressed as a sum of fractions with a numerator 1
- about 85 assorted mathematical exercises with solutions.

CENTRE FOR DIGITAL
LIBRARY RESEARCH

The Theory of Communication Functions
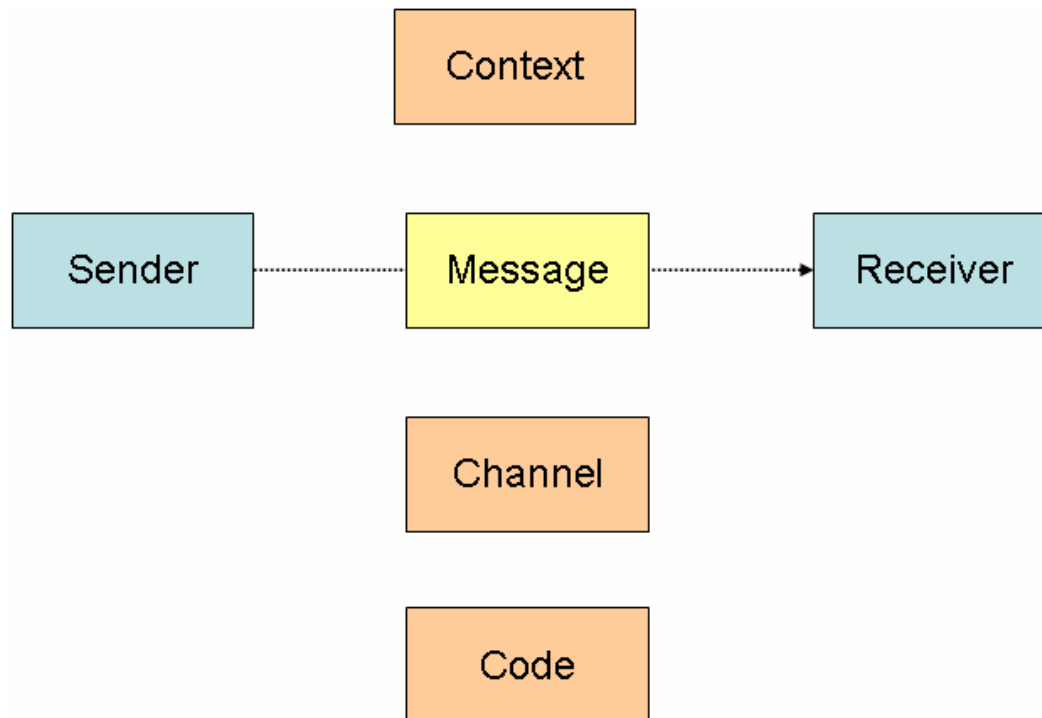
Sender, Receiver – distant in time

Message – digital object

Channel – technological medium

Context – preservation environment

Code – software, standards

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Volume of DML

- The entire mathematical literature consists of approximately 50 million pages contained in books, journals, and various other publications. (estimation of Prof. R. Keith Dennis, Ithaca)

- This is not digitized completely – in the recent EU-DML project the estimate of existing digitized volume across Europe is **> 2,500,000 pages**

- **The effort to create this collection is substantial and motivated by researchers' needs; we need to find also ways to sustain the DML for the future.**

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Typical preservation issues…

Failure of any component of the technological chain

Hardware, software or support environment change.

Outcomes of a project which is not sustained.

Problems with the "the bits"

This could happen because of a storage device or medium failure;

or if a DNS entry is no longer be resolvable.
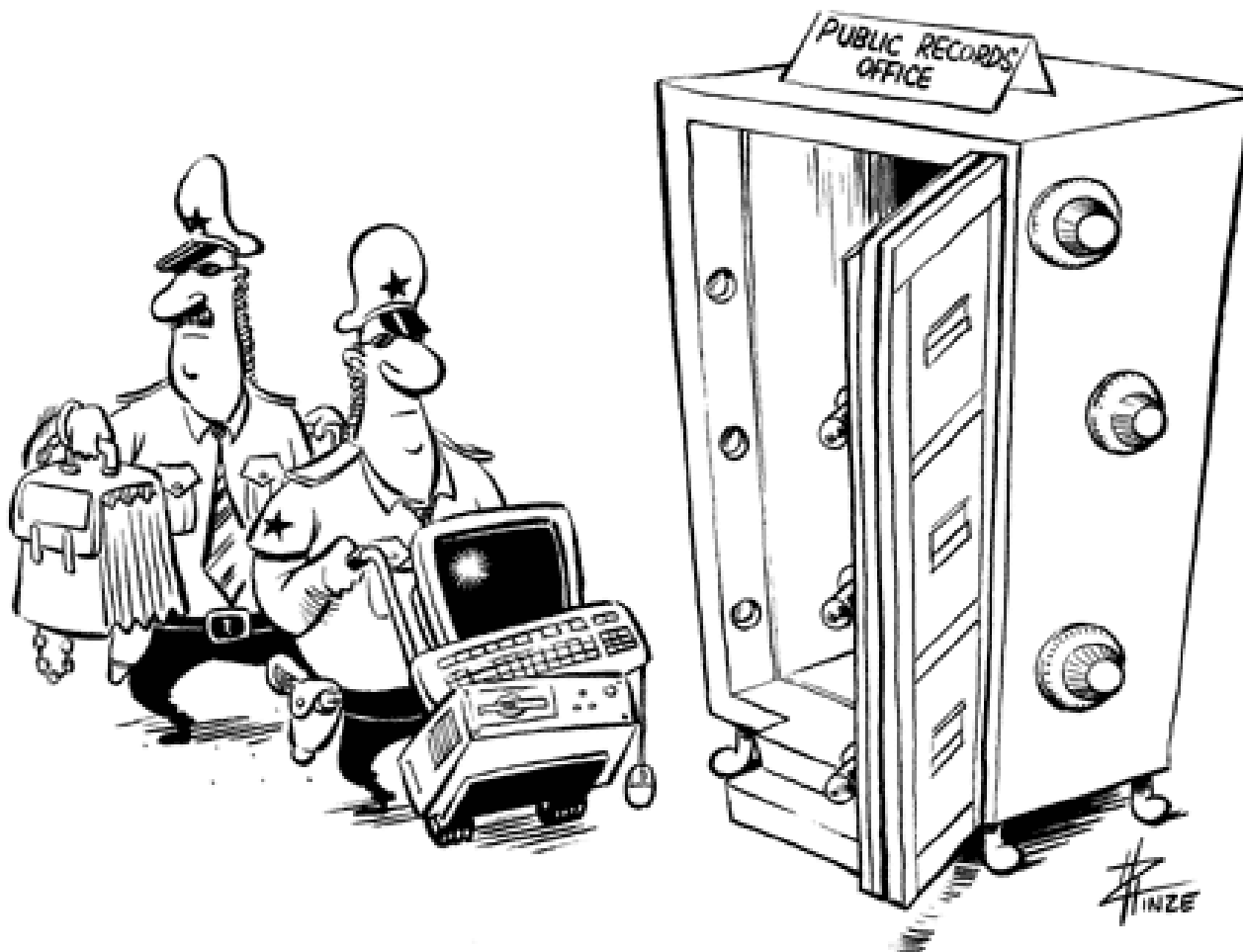
Changes in the Knowledge Base

Loss of understanding or usability

Lost data on provenance or authenticity (requirements of trusted repositories)

Record of who did what and how did they do it

Approaches: migration, simulation, digital archaeology

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Is this our best option?

© Hinze/Scratch! www.scratch.com.au

2 The International Context (who does what)

➢ Overview of the EC policy

➢ Brief presentation of recently supported projects (FP6/FP7)

Milena Dobreva, KOM, DML-CZ, 12 June 2008

The Preservation Landscape

*"… if we do not actively pursue the preservation of digital material now, we risk having **a gap in our intellectual record**. … If you allow me another historical reference, we do not want to experience the digital equivalent of the destruction of the Alexandria Library. Scientific assets are just too valuable to be put at risk".*

Ms. Viviane Reding,
EU Commissioner on IS & Media

"The digital preservation community's inability to bring firm evidence to bear in support of its contentions about data loss, coupled with the alarmist rhetoric of terms such as digital dark ages and digital black hole, leave us exposed."

Prof. Ross Harvey,

Charles Sturt University, NSW, Australia

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Policy of the EC

Action plan for digitisation, online access and digital preservation

> 2008: national strategies for long term preservation and deposit with quantitative and qualitative targets
>
> 2008: standards for digital preservation
>
> 2009: legislative framework supporting digital preservation

http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2006:297:0001:0005:EN:PDF

Milena Dobreva, KOM, DML-CZ, 12 June 2008

European projects

FP5

- **DELOS** (2004-) 'Networking of Excellence in Digital Libraries' - Digital Preservation Cluster

FP6

- **DPE** (2006-) Digital Preservation Europe
- **PLANETS** (2006-) Permanent Long-term Access through Networked Services
- **CASPAR** (2006-) Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval

FP7

- **SHAMAN** (2007-)

Milena Dobreva, KOM, DML-CZ, 12 June 2008

DELOS Preservation Cluster (2004-2007) www.dpc.delos.info


Emphasis

on

training
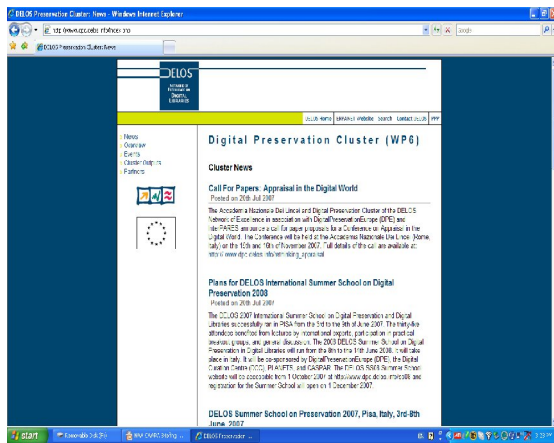
- – Coordination
- – Collection and dissemination of information, knowledge and expertise
- – Develop tools for evaluating digital preservation strategies
- – Develop testbeds and metrics
- – Relate research agenda to the development of exploitable product opportunities
- – Help eliminate the duplication of effort

Milena Dobreva, KOM, DML-CZ, 12 June 2008

DPE (2006-)
www.digitalpreservationeurope.eu

*Emphasis
on
cooperation*
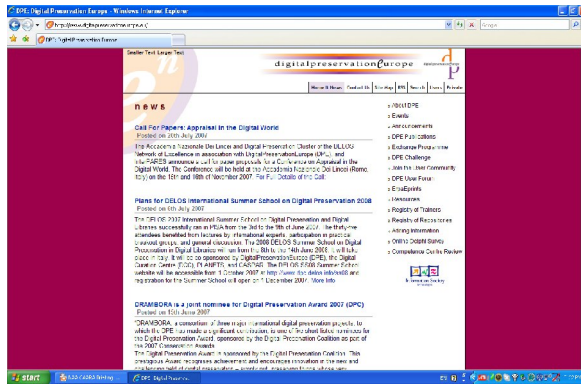
Network of excellence

Objectives:

- To create a coherent platform for proactive cooperation, collaboration, exchange and dissemination of research results and experience in the preservation of digital objects
- To increase prevalence, viability and accountability of preservation services
- To improve awareness and skills

Milena Dobreva, KOM, DML-CZ, 12 June 2008

One Outcome of DPE: Research Roadmap: 9 themes

1. Restoration
2. Conservation
3. Collection and Repository Management
4. Preservation as Risk Management
5. Preserving the Interpretability and Functionality of Digital Objects
6. Collection Cohesion and Interoperability
7. Automation in Preservation
8. Preserving the Context
9. Storage Technologies and Methods
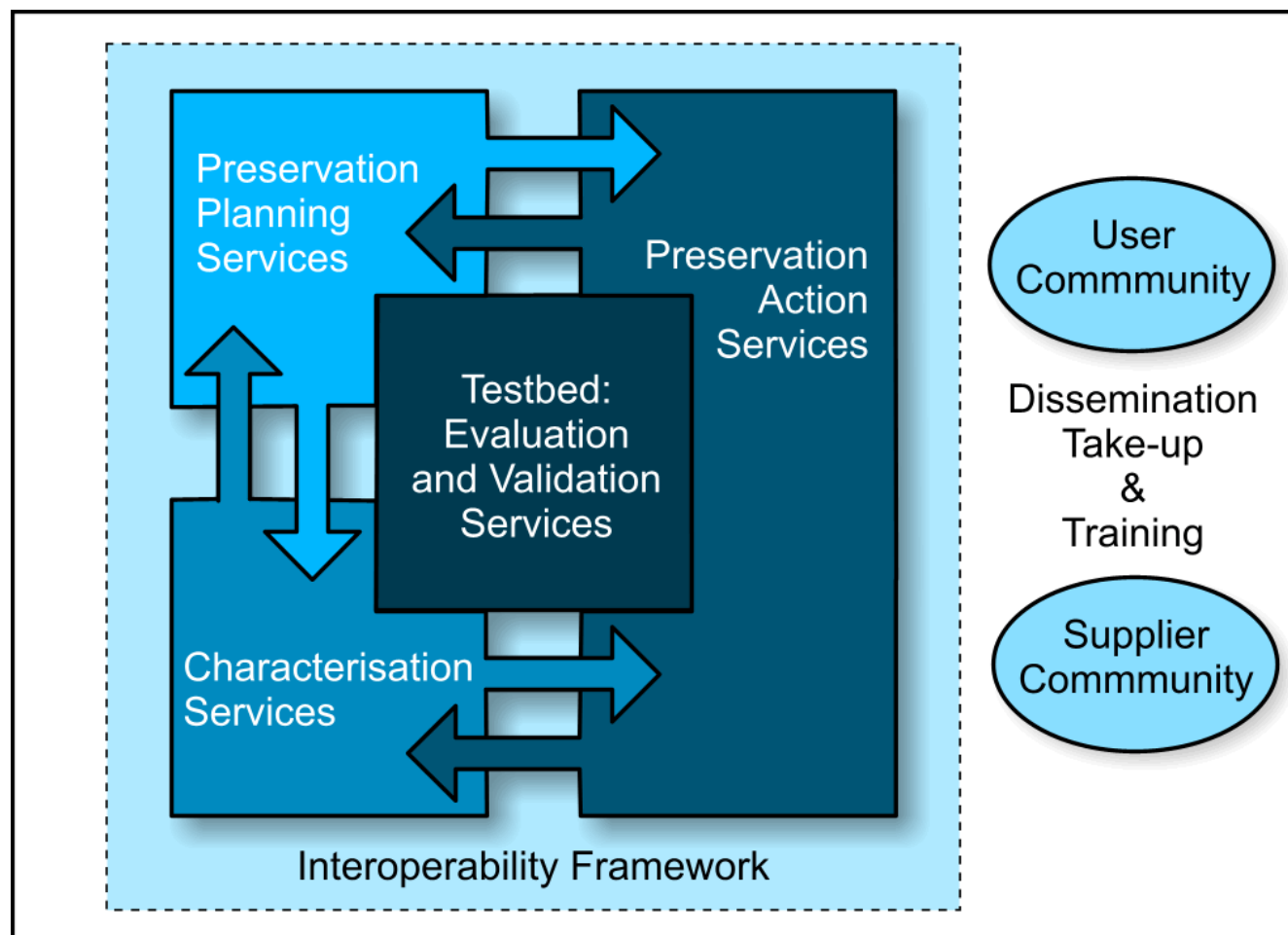
PLANETS (2006-)
www.planets-project.eu
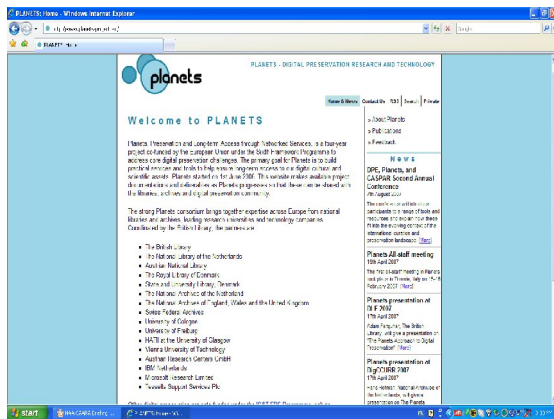
Emphasis
on
preservation
planning

Objectives – to offer:

– Preservation Planning services that empower organisations to define, evaluate, and execute preservation

– Methodologies, tools and services for the Characterisation of digital objects

– Innovative solutions for Preservation Actions tools …
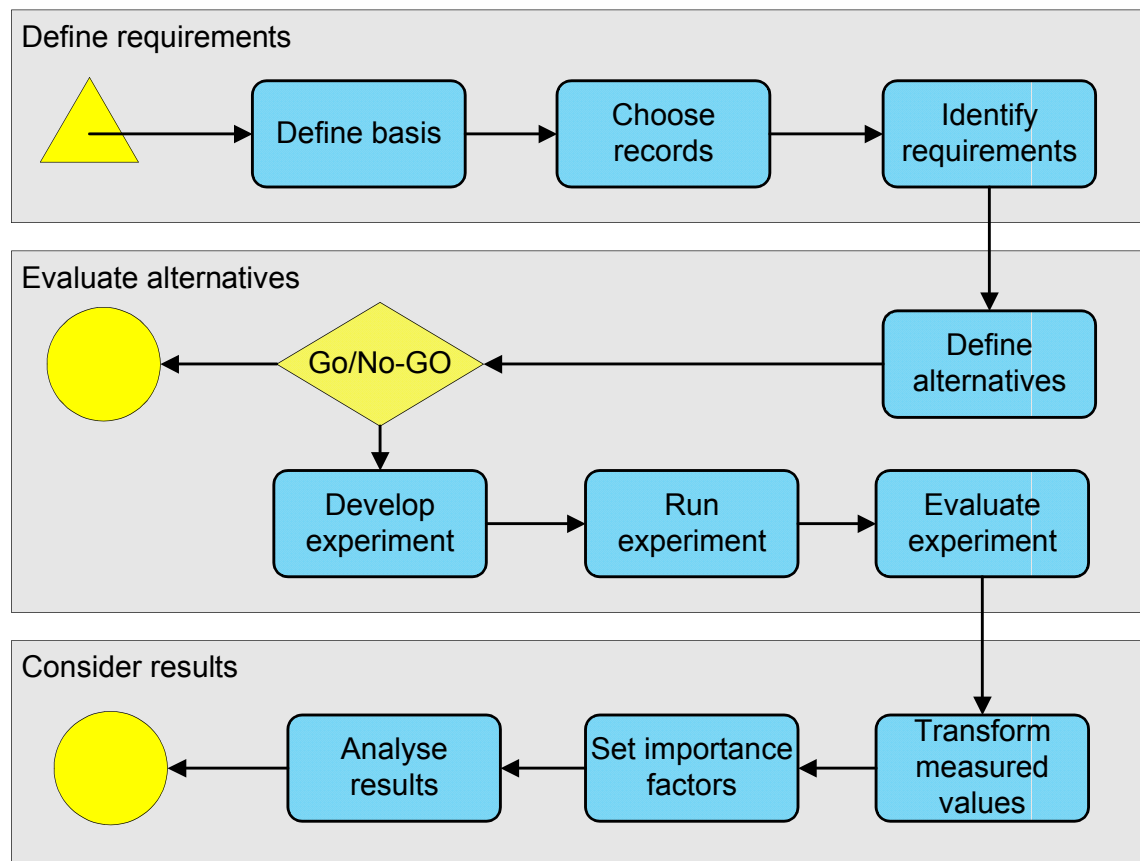
– An Interoperability Framework

– A Testbed

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Planets architecture

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Examples… process view

Preservation planning in Planets
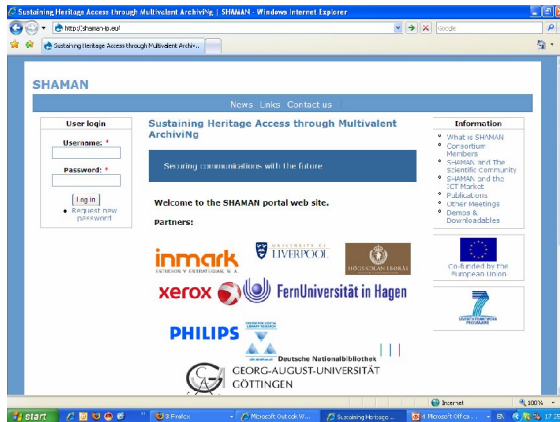
CASPAR (2006-)
http://www.casparpreserves.eu

Aims to:

- Implement, extend, and validate the OAIS reference model

- Enhance the techniques for capturing Representation Information

- Research more sophisticated access to and use of preserved digital resources

- Develop case studies

- Actively contribute to the relevant standardisation activities

Emphasis
on
extending OAIS
reference model
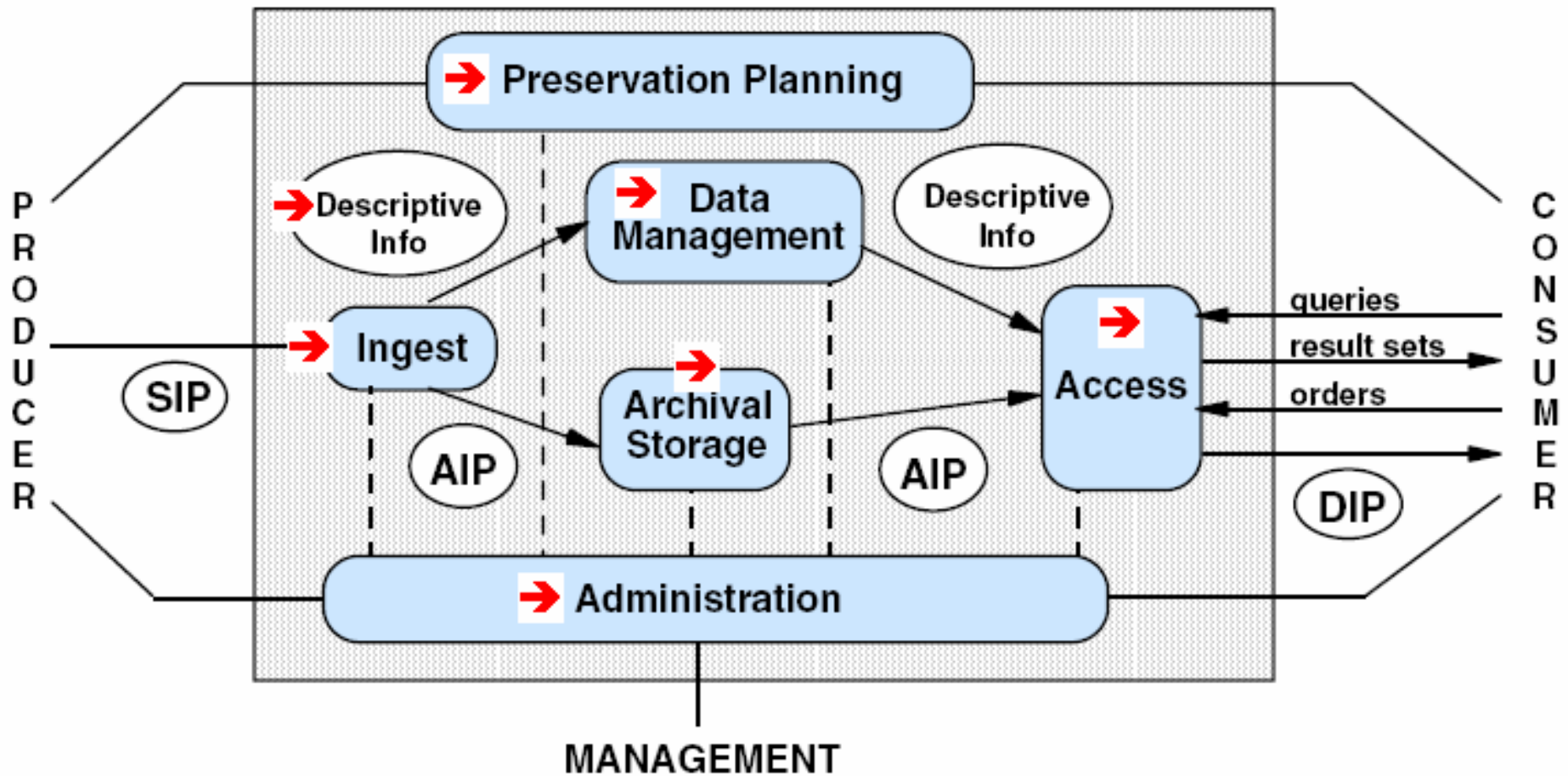
SHAMAN (2007 -)
www.shaman-ip.eu

FP7 call 1, IP

The goal of SHAMAN is to create a preservation framework where not only the original data are preserved, but also the preservation process. Such that one can take the original data and the preservation process into another system and successfully preserves the original data.

CDLR leads WP2 – reference architecture

Emphasis on **theory of preservation+ distributed environment**

Milena Dobreva, KOM, DML-CZ, 12 June 2008

OAIS
Open Archival Information System

Source: OAIS

CENTRE FOR DIGITAL LIBRARY RESEARCH

Preservation metadata

Reference information

Context information

Provenance information

Fixity information

Trustworthiness in repositories – authenticity, provenance, integrity

Audit of repositories

SHAMAN: what's new

- theoretical foundations – trying to build a systematic theory of preservation which aims to extend the OAIS (Open Archival Information System) approach

- storage - based on grid technologies for data storage (iRODS)

- automation – the architecture will integrate a workflow machine

- visualization of objects – Multivalent tool (visualization of digital objects without the original rendering software)

- domains of focus: cultural institutions, engineering, eScience

3 **What issues can be identified in DML?**

- **Motivation: as seen during this workshop**
  - The mathematical literature must be carefully archived, indexed and preserved.
  - Preservation is currently understood as *hardware maintenance, emulation, management*
- **Size of effort – typical area of work in progress, can not be estimated**
- **Who is in charge – the curent content providers**
- **Retro born digital volumes – preservation is already known**
- **What to preserve? – master image files, final electronic publications, intermidiate auxiliary digital objects?**

CENTRE FOR DIGITAL
LIBRARY RESEARCH

And some more…

Format & Quality

- PDF/DjVu600 dpi bitonal + hidden text
- PDF born digital
- PDF 300 dpi monochrome image
- TIFF/PDF600 dpi mainly bitonal
- PS

Metadata: descriptive, no preservation metadata

- Heterogeneous
  - **sets of elements,** and
  - **formats** (XML, LaTeX)

One preservation connection…

- **SHAMAN** is identified as a suitable preservation project to refer to within the most recent EU-DML proposal to eContentplus

- **EU-DML** is seen as one possible trial collections within the e-Science domain of focus of SHAMAN

Milena Dobreva, KOM, DML-CZ, 12 June 2008

Preservation - from three-legs…
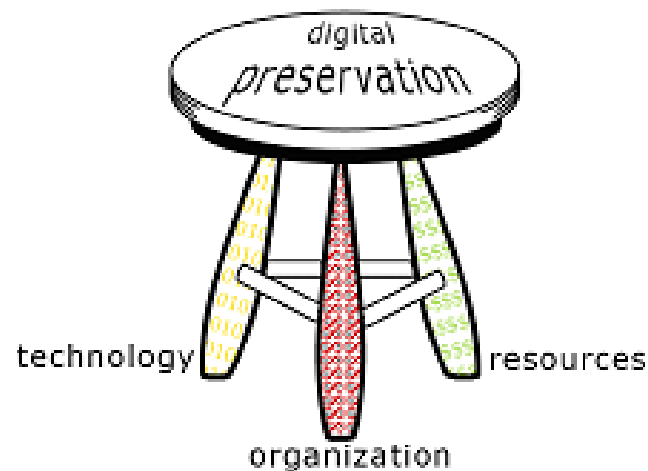
Cornell University Library

Digital Preservation
Management Workshop
series (2003-2006)

… to six pillars

Preservation: interoperability over time

Standards
  (including metadata)

Workflows
  (including approaches)

Quality

Costs

Research agenda

Automation

Milena Dobreva, KOM, DML-CZ, 12 June 2008

CENTRE FOR DIGITAL
LIBRARY RESEARCH

Where to find further info?

Websites

– Digital Libraries initiative

http://ec.europa.eu/information_society/activities/ digital_libraries/index_en.htm

– Digital Cultural Heritage, Digital Libraries and Digital Preservation

http://www.cordis.lu/ist/digicult/index.html

Summer School in DP

International Journal of Digital Curation

Conferences: ECDL, IJDL, ICADL

Questions?

Milena Dobreva, KOM, DML-CZ, 12 June 2008