Scientific Evaluation of Digital Library Management Systems (DLMS):
general issues and outcomes on information access components

Maristella Agosti

University of Padova, Italy

Department of Information Engineering (DEI)

Information Management Systems Research Group (IMS)

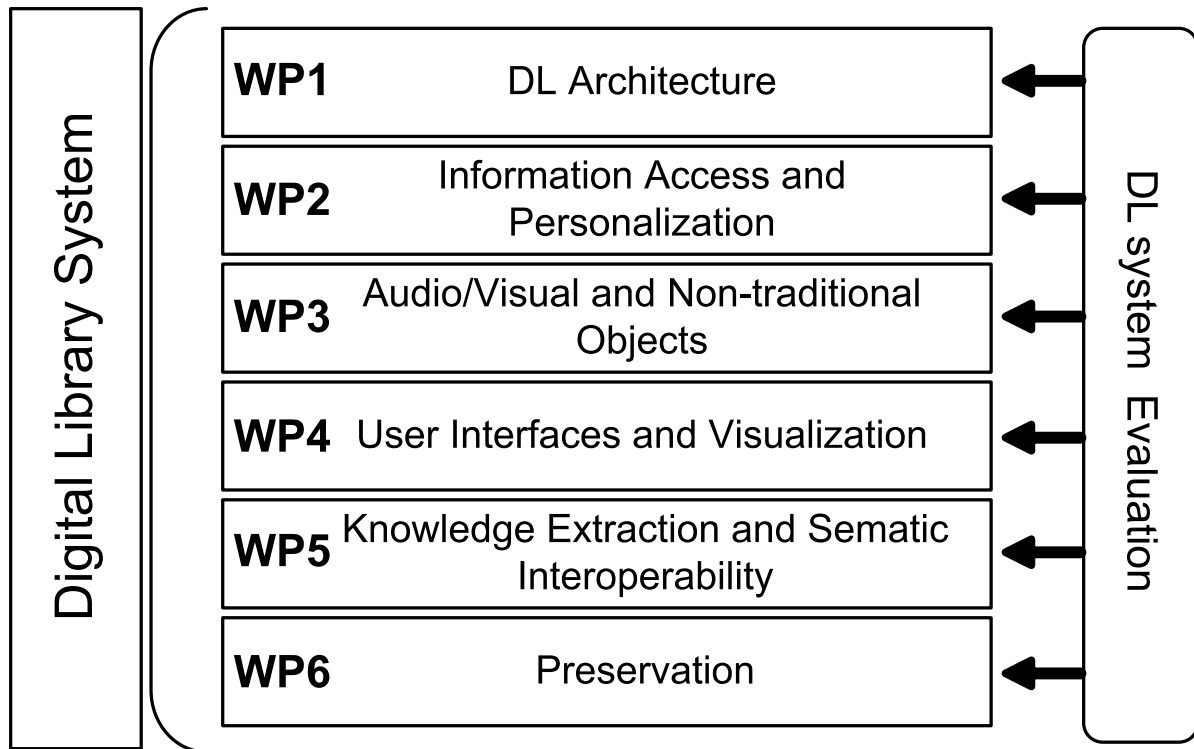DEPARTMENT OF INFORMATION ENGINEERING
UNIVERSITY OF PADOVA

Outline

❑ General aspects for DL evaluation:
  ○ Main aspects of a Digital Library Management System (DLMS): the DELOS view
  ○ Classification Scheme of a DL
  ○ DLMS as Information-technology Infrastructures.

❑ A new approach:
  ○ Focus on revision of current evaluation methodologies in order to fit to a new way of thinking about the development of a DLMS
  ○ A new proposal: a Scientific Reflection DLMS
  ○ A curation service of the scientific reflection DLMS for supporting the evaluation of information access and retrieval.
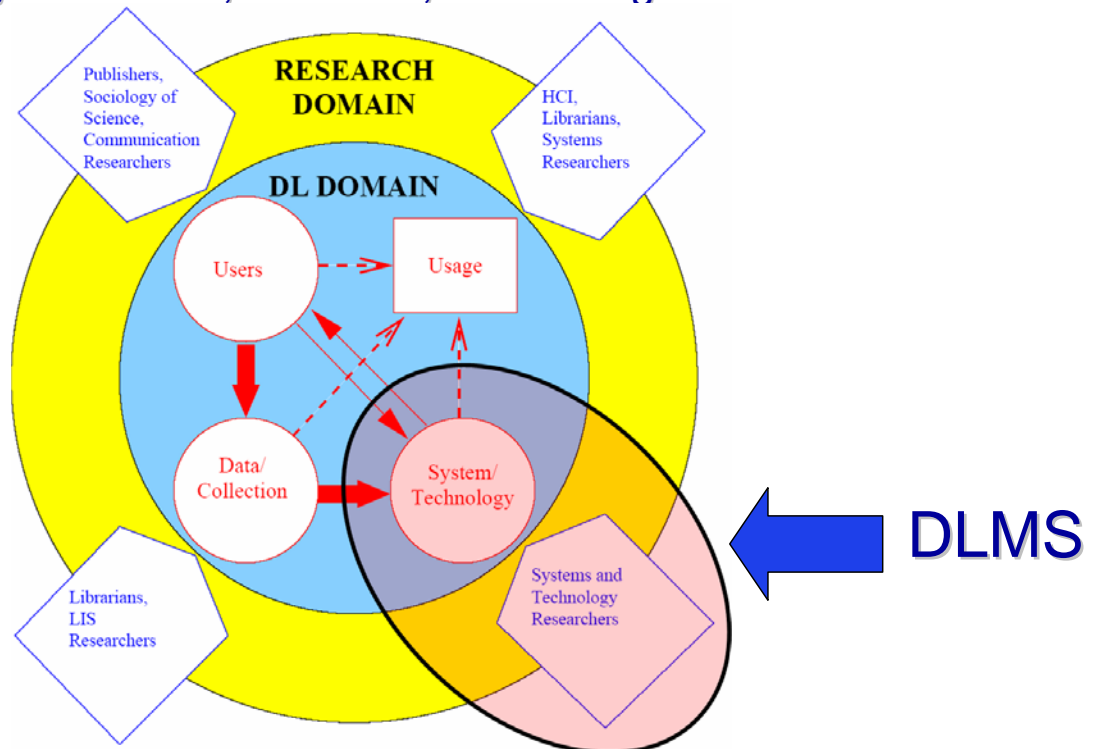
Main aspects of a DL Management System

WP1	DL Architecture	←
WP2	Information Access and Personalization	←
WP3	Audio/Visual and Non-traditional Objects	←
WP4	User Interfaces and Visualization	←
WP5	Knowledge Extraction and Sematic Interoperability	←
WP6	Preservation	←

Digital Library System

DL system Evaluation

Classification Scheme of a DL

N. Fuhr, P. Hansen, A. Micsik, I. Solvberg - ECDL2001

DLMS

DLMS as Information-technology Infrastructures

Y. Ioannidis, D. Maier, S. Abiteboul, P. Buneman, S. Davidson, E.A. Fox, A. Halevy, C. Knoblock, F. Rabitti, H.-J. Schek, G. Weikum, 2005.

❑ "DL development must move *from an art to a science* in order to give rise to industrial-strength DLMS, based on reliable and extensible services".

❑ Shift from DL to service-based DLMS.

❑ Requirement for improved reliability points out, among other issues, the need of proper evaluation methodologies in order to assess a DLMS along different dimensions.

❑ In the light of this scientific development, the evaluation methodologies should not be perceived as something external to the design and development process of a DLMS but they should be tightly integrated into it.

Evaluation of a DLMS

❑ The evaluation of a DLMS turns out to be a scientific activity whose outcomes, such as performance analyses and measurements, constitute a kind of *scientific data*
that need to be properly considered and used for

  ❍ the design and
  ❍ development

of DLMS components and services.

Evaluation of Information Access

❑ The evaluation of the

  ○ information access and retrieval component

  of a Digital Library Management System can be conducted in different ways, one of the more effective way is with the approach of using test or experimental collections.

What is a Test or Experimental Collection?

❑ A test or experimental collection is:

  a triple C = (D, Q, J)

  where:

  D is a set of documents, called also collection of documents
  Q is a set of topics, from which the actual queries are derived
  J is a set relevance judgements

❑ The definition and construction of a test-collection are based on the so-called "Cranfield methodology".

Experiments with Test-collections

❑ International evaluation forums such as:

    ❍ TREC, http://trec.nist.gov/

    ❍ CLEF, http://clef.isti.cnr.it/

    ❍ NTCIR, http://research.nii.ac.jp/ntcir/index-en.html

    ❍ INEX, http://inex.is.informatik.uni-duisburg.de/

have traditionally carried out statistical analyses, which provide participants with an overview analysis of the submitted experiments.

Evaluation Forums and Statistical Analysis

❑ The participants of International evaluation forums may conduct statistical analyses on their own experiments by using either ad-hoc packages, or generally available software tools with statistical analysis capabilities, like R, SPSS, or MATLAB.

❑ However, the choice of whether performing a statistical analysis or not is left up to each participant.

❑ Moreover, when participants perform statistical analyses using their own tools, the comparability among these analyses could not be fully granted.

A New Proposal: a Scientific Reflection DLMS

❑ A Scientific Reflection DLMS is a system that is designed to store and manage

  ○ test-collections and
  ○ related scientific data

  produced and used over the years in one or more of the international forums that have been set up to evaluate information access and extraction components in different contexts.

Role of a Scientific Reflection DLMS

❑ A Scientific Reflection DLMS designed to manage test-collections and scientific data produced in the international evaluation forums can play a central role as a service able to support:

  ○ the test-collection management in general, and specifically

  ○ the evaluation of the information interaction in context.

Scientific Reflection DLMS: curation service

❑ A Scientific Reflection DLMS can support a curation service able to

  ○ keep,

  ○ manage, and

  ○ enrich

  context data, test-collections, and research data built and produced in evaluation forums of information access and extraction components of an information management system.
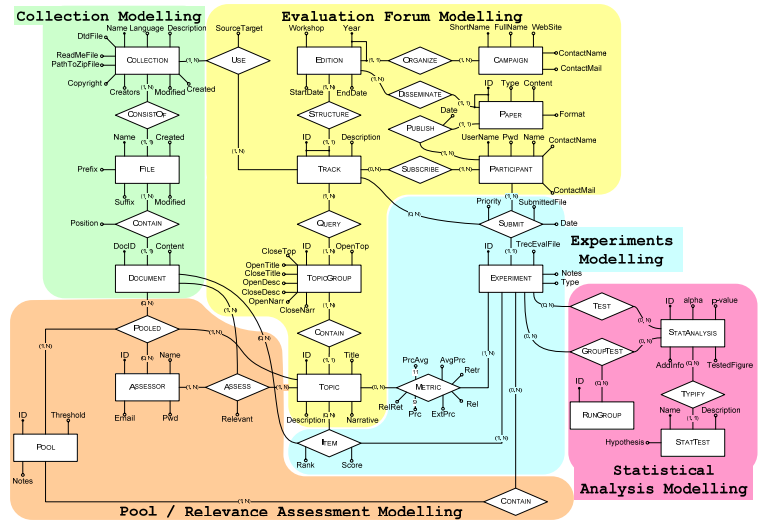
Design of the Curation Service

❑ The requirements analysis process for such a service has been conducted, the results are: the conceptual model, and the architecture for the curation service.

❑ Design of an Information Access Evaluation Service for a Scientific Reflection DLMS:

  ○ Conceptual Model for the Service

  ○ Architecture of the Service.

Conceptual Schema of the Service

❑ Areas that are represented in the conceptual schema:

  ❍ Collection modelling

  ❍ Evaluation forum modelling

  ❍ Pool/relevance assessment modelling

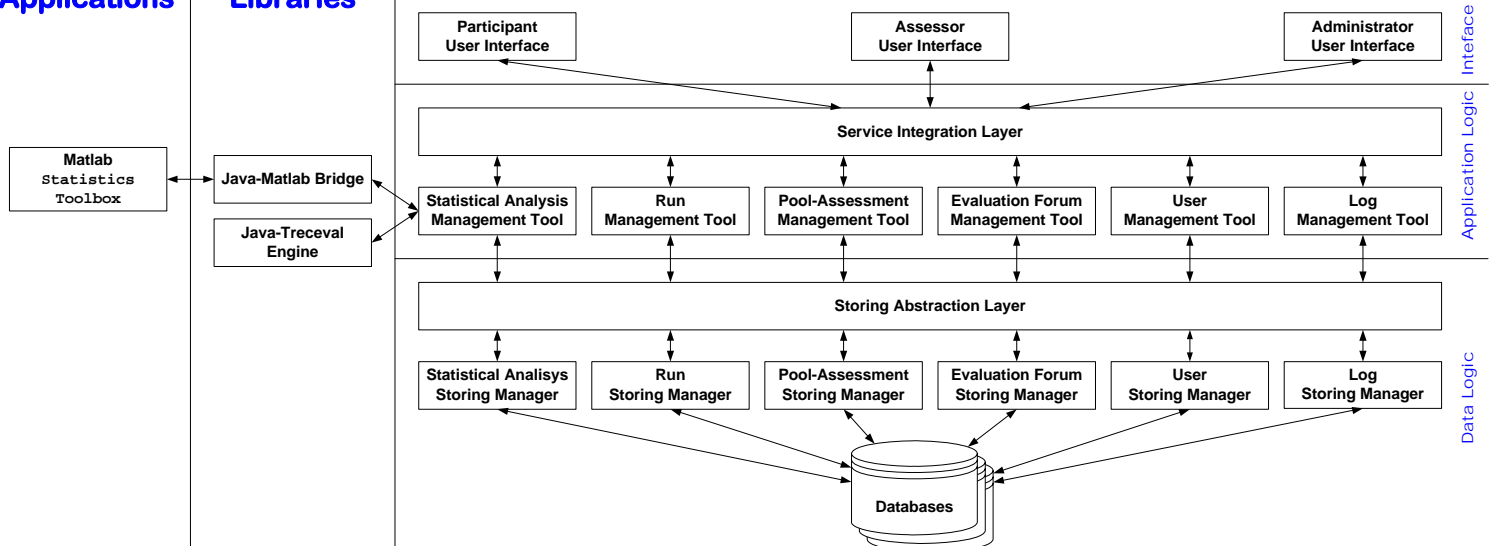 ❍ Experiments modelling

  ❍ Statistical analysis modelling

Architecture of the Service

Acknowledgements
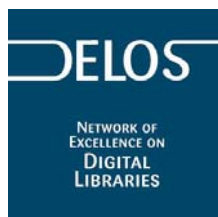
Further information on DELOS activities can be found in the NoE Web site at the URL: http://www.delos.info/

References

1. M. Agosti, G.M. Di Nunzio, N. Ferro. Evaluation of a Digital Library System. In: M. Agosti, N. Fuhr (Eds). *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*, 2004, pp. 73-78.

2. G.M. Di Nunzio, N. Ferro. DIRECT: a System for Evaluating Information Access Components of Digital Libraries. *Proc. 9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005),* pp. 483-484. LNCS 3652, Springer, Heidelberg, Germany, 2005.

3. N. Fuhr, P. Hansen, A. Micsik, I. Sølvberg. Digital Libraries: A Generic Classification Scheme. *Proc. 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001),* pp. 187-199. LNCS 2163, Springer, Heidelberg, Germany, 2001.

4. Y. Ioannidis, D. Maier, S. Abiteboul, P. Buneman, S. Davidson, E.A. Fox, A. Halevy, C. Knoblock, F. Rabitti, H.-J. Schek, G. Weikum. Digital library information-technology infrastructures. *International Journal on Digital Libraries*, 5(4):266-274, 2005.