



DELOS Summer School 2005
Digital Preservation in Digital Libraries
Sophia Antipolis, France
5th-11th June, 2005

Session Plans & Biographical Sketches of Lecturers

- ◆ **DELOS: A Network of Excellence on Digital Libraries**
Instrument: Network of Excellence
Thematic Priority: IST-2002-2.3.1.12
Technology-enhanced Learning and Access to Cultural Heritage
European Commission Project no. 507618

Co-Sponsor:

- ◆ **Digital Curation Centre (DCC) of the United Kingdom**
The DCC is funded by the JISC & the EPSRC (www.dcc.ac.uk)

Host:

- ◆ **ERCIM & INRIA**

Seamus Ross & Hans Hofman
Academic Directors of the DELOS 2005 Summer School
On behalf of the Digital Preservation Cluster of DELOS
<http://www.dpc.delos.info>

Lecturers and Session Presenters

- *Birte Christensen-Dalsgaard*, Head of Development at the State and University Library in Aarhus.
- *Michael Day*, Research and Development team at UKOLN.
- *David Giaretta*, Associate Director for Development at the UK Digital Curation Centre.
- *Mariella Guercio*, professor in archival science and electronic record management, ISTBAL, Università degli Studi di Urbino.
- *Ross Harvey*, inaugural Professor of Library and Information Management at Charles Sturt University's School of Information Studies.
- *Stephan Heuscher*, co-founder ikeep AG (Digital Archives Services).
- *Hans Hofman*, co-director of ERPANET and senior advisor at the Nationaal Archief (National Archives) of the Netherlands.
- *Anne R. Kenney*, Associate University Librarian for Instruction, Research, and Information Services at Cornell University Library.
- *Seamus Ross*, Professor of Humanities Informatics and Digital Curation, and Director HATII (Humanities Advanced Technology and Information Institute) (<http://www.hatii.arts.gla.ac.uk>), University of Glasgow.
- *Andreas Rauber*, Associate Professor at the Department of Software Technology and Interactive Systems at the Vienna University of Technology.
- *Manfred Thaller*, Professor of Historisch-Kulturwissenschaftliche Informationsverarbeitung, University of Köln.

DELOS SS2005 Timetable

Sunday, 5 June 2005

- 18:30 Overview of the Summer School and Introduction to the Lecturers
- 19:30 Opening Reception at ERCIM/INRIA

Monday, 6 June 2005

- 9:00 - 12:30 Digital Curation in Digital Libraries: Issues, Obstacles, and Possibilities (*Anne R. Kenney* is Associate University Librarian for Instruction, Research, and Information Services at Cornell University Library)
- 14:00 - 17:30: Unpacking The OAIS Model (*David Giaretta* is the Associate Director for Development at the UK Digital Curation Centre)

Tuesday, 7 June 2005

- 9:00 - 12:30: Categories, Uses and Challenges of Metadata and Process Documentation (*Michael Day* research and development team at UKOLN)
- 14:00 - 17:30: Workflow and Workflow Modelling (Stephan Heuscher, iKeep AG, Digital Archives Services)

Wednesday, 8 June 2005

- 9:00 - 12:30: Identifying, Evaluating and Selecting Preservation Methods: An Introduction to the DELOS Testbed and Utility Analysis (*Hans Hofman* is co-director of ERPANET and senior advisor at the Nationaal Archief (National Archives) of the Netherlands & *Andreas Rauber*, Associate Professor at the Department of Software Technology and Interactive Systems at the Vienna University of Technology)
- 14:00 - 17:30: Managing Ingest: Handling, Documenting, and Automating (*Birte Christensen-Dalsgaard* Head of Development at the State and University Library in Aarhus)

Thursday, 9 June 2005

- 9:00 - 12:30: Role of Registries and Representation Information (*David Giaretta* is the Associate Director for Development at the UK Digital Curation Centre)
- 14:00 - 17:30: Approaches to Audit and Certification of Repositories and Preservation Processes (*Seamus Ross*, Professor of Humanities Informatics and Digital Curation, and Director HATII (Humanities Advanced Technology and Information Institute) (<http://www.hatii.arts.gla.ac.uk>), University of Glasgow) & *Mariella Guercio*, professor in archival science and electronic record management, ISTBAL, Università degli Studi di Urbino).

Friday, 10 June 2005

- 9:00 - 12:30: Methodologies of Selection and Appraisal (*Ross Harvey* is the inaugural Professor of Library and Information Management at Charles Sturt University's School of Information Studies)
- 14:00 - 17:30: Digital Libraries as Persistent Collections of Autonomous Objects (*Manfred Thaller* Professor of Historisch-Kulturwissenschaftliche Informationsverarbeitung, University of Köln)

Saturday, 11 June 2005 (Morning Only)

- 9:00 - 12:30: Question and Answer Session (Panel: All Lecturers)

Opening Session:

Overview of the Summer School & Introduction to the Lecturers

Abstract:

The DELOS Network of Excellence is a four-year project funded by the European Commission under the Sixth Framework Programme to synergize and foster technology for the next-generation of Digital Libraries (www.delos.info). This delivery of this event is co-sponsored by the Digital Curation Centre (DCC) and ERPANET. The aim of the summer school is to assist participants in understanding how to address digital preservation challenges in the context of the digital library. The workshop also provides a networking opportunity for participants to meet with other students and researchers, international experts, and practitioners across disciplinary and national boundaries.

During this six-day summer school, internationally established lecturers will each lead half-day sessions and most will be available through the week for further discussion. By the end of the course, participants should have:

- gained an appreciation of the issues surrounding digital preservation within the context of digital library development and management;
- developed a grasp of the core research in the area of digital curation and preservation;
- developed a coherent and practical understanding of activities surrounding digital preservation;
- gained experience with issues surrounding workflow modelling, metadata definition, and ingest process management;
- acquired an appreciation of the different approaches to selecting and appraising potential digital acquisitions;
- become familiar with the OAIS model and gained a knowledge of the approaches to repository design and deployment;
- a working knowledge of the issues surrounding audit and certification of digital repositories;
- a working knowledge of the techniques and practices that underlie digital curation;
- develop a firm understanding of the issues of authenticity, integrity, and reliability in relation to digital libraries; and
- considered how digital curation and preservation requirements can be integrated into approaches to digital library development

Session (Monday AM)

Digital Preservation in Digital Libraries: Issues, Obstacles, and Possibilities

Lecturer and Session Leader:

Anne R. Kenney, Associate University Librarian for Instruction, Research, and Information Services at Cornell University Library

Abstract:

This session begins with an exploration of key issues for digital preservation, including a review of reasons for and against establishing digital preservation programs. The focus will be from an organizational perspective and participants will be introduced to the five developmental stages that institutions pass through in establishing digital preservation programs. Managerial, technical, and resource implications of digital curation will be explored through the introductory lecture and the class project. Participants will then discuss strategies for ensuring that digital preservation is well integrated into digital library programs.

Will include lecture, class project, and discussion of key issues

Session outline:

- *Lecture (60 minutes)*
- *Activity & Discussion (60 minutes):* Analysis of Institutional Readiness
- *Seminar: (60 minutes)*

Session (Monday PM)

Unpacking the OAIS Model

Lecturer and Session Leader:

Dr David Giaretta, Associate Director for Development at the UK Digital Curation Centre

Abstract:

This session aims to elucidate the Open Archival Information Systems (OAIS) Reference Model, its origins, its purpose, the reasons behind its structure, the several models within it and the follow-on standards planned or in progress. OAIS is one of the key standards concerned with the long term preservation of digital information.

Aims and learning outcomes of this session:

This session aims to:

- provide a clear understanding of the origins and role of the OAIS Reference Model
- explain the real issues being addressed by OAIS
- describe the various components of the model
- give examples of mapping existing repositories to OAIS
- indicate how the OAIS Reference Model fits into the wider context of Digital Libraries

By the end of this session students will have:

- developed an appreciation of the significance of the OAIS Reference Model
- understood the roles and interrelations between the components of the reference model
- become familiar with the limitations of the OAIS Reference Model
- grasped some of the ways in which the reference model can be applied.

Session outline

- *Lecture* (60 minutes): The key ideas behind the OAIS Reference Model, and the role of its various components
- *Discussion*(participants, facilitated by the presenter, 20 minutes): Given a stream of bits, what is needed to produce usable information? How will this change over time?
- *Seminar* (30 minutes): discussion of two or three key readings (details of these to be advised)
- *Lecture* (25 minutes): Applicability of the OAIS Reference Model.
- *Activity* (30 minutes): Mapping OAIS to existing repositories.
- *Summary* (15 minutes)

Session (Tuesday AM):

The design and use of metadata in support of digital preservation strategies

Lecture and Session Leader:

Michael Day, Research and Development team at UKOLN

Abstract:

Metadata can be defined simply as structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage resources. In recent years, there has been a growing awareness of the role of metadata in supporting the preservation and reuse of digital resources. For example, in 2003 the US National Science Foundation's Blue-Ribbon Advisory Panel on Cyberinfrastructure argued that the creation and maintenance of metadata was essential for the ongoing stewardship and curation of research data. To date, much metadata development has been fragmented, typically focused on particular implementations, domains or resource types. More recently, however, groups like the PREMIS (Preservation Metadata: Implementation Strategies) working group have begun to identify the core functional requirements of preservation metadata, basing their work partly on the increasingly influential Reference Model for an Open Archival Information System (OAIS).

Aims and learning outcomes of this session:

The session aims to:

- Introduce students to the important role of metadata in supporting digital preservation strategies and highlight some key initiatives
- Identify generic preservation metadata requirements, i.e. the functions that metadata might support within a preservation system
- Highlight the difficulty of defining a single schema that might be suitable for all resource types and all preservation strategies

By the end of the session students will have:

- Gained an appreciation of why the creation, capture and management of metadata should be a key part of most digital preservation strategies
- Developed a feel for the range of metadata initiatives and standards currently in use (or under development)
- Appreciated some of the practical reasons why generic preservation metadata schemas are difficult to develop and implement

Session outline:

- *Lecture* (30 minutes): Welcome and session outline, defining metadata, outlining its importance in supporting preservation strategies (including examples). The primary focus will not be specific schemas or implementations, but generic metadata requirements (e.g. those identified by the OAIS information model) and how these might interact with different resource types and preservation strategies.
- *Discussion* (30 minutes): Questions will include: What are the most important types of metadata needed to support preservation? What may be missing from the OAIS information model categories? How much of this metadata might exist in

other forms already, or may be able to be captured automatically?

- *Lecture* (30 minutes): Introducing selected schemas (looking in detail at the results of the PREMIS activity), implementation issues, and managing metadata.
- *Activity* (1 hour, 15 minutes; 45 minutes discussion in groups, 30 minutes reporting back time and discussion): participants (including other summer school tutors) divided into teams, each looking at a different resource type (e.g., simple dataset, online e-journal, TIFF image, etc.). Using a template, teams will need to select an appropriate preservation strategy, identify what types of metadata would be necessary to support this, consider what metadata exists already (and whether this can be captured), then mapping this on to the OCLC/RLG framework (or the PREMIS Core Metadata model, if available) The aim is to see whether such models are useful (or sufficient), and get students to think about the issues surrounding the generation of metadata in a practical way.
- *Summary and discussion* (15 minutes): Summing up and a final opportunity for questions

Session (Tuesday PM)

Workflow and Workflow Modelling in Digital Preservation

Lecturer & Session Leader:

Stephan Heuscher, ikeep Ltd (Digital Archives Services)

Abstract:

The Workflow Management Coalition (WfMC), defines a workflow as "the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules". The important terms being automation, information, routing, participants, and action, all being controlled by rules. The rules define the sequence of who has to act on the information, all this in an automated manner. Automated means that an information system is in control, i.e. that the rules are defined for and enforced by that information system (workflow management system). To obtain the rules, the desired workflow has to be modeled first. The formal modeling of workflows is very important because it enables domain experts to give their knowledge in terms of a formal definition that can be tested, reviewed, and optimized. The modeled workflows are enacted by the workflow management system, each workflow creating a log of the actions and their results during its execution.

The benefits of workflows in digital preservation (and elsewhere) lie in the need to formally describe the steps in the workflow, i.e. getting all the people involved to agree on the sequence and definitions of actions. This allows to communicate the procedures within the institution and compare it to the ones used in other institutions. Additionally, the metrics can be defined much more granular, thereby allowing a better control over the procedures.

In the session, there will be an introduction to workflows and workflow modeling, as well as practical activity of modelling and analyzing workflows.

Aims and objectives:

- Understand the basics of workflow modelling
- Gain experience in modeling workflows
- Enable critical analysis of workflows
- Provide a basis for process documentation, audit, and ingest management

Session outline:

- Lecture (60 min)
- Activity (60 min)
- Discussion of Activity and Seminar

Session (Wednesday AM)

Identifying, Evaluating and Selecting Preservation Methods:

An Introduction to the DELOS Testbed and Utility Analysis

Lecturers and Session Leaders:

Andreas Rauber, Associate Professor at the Department of Software Technology and Interactive Systems at the Vienna University of Technology.

&

Hans Hofman, co-director of ERPANET and senior advisor at the Nationaal Archief (National Archives) of the Netherlands

Abstract:

During the last couple of years, several preservation strategies have been developed, tested, rated and implemented, but until now, none of them is clearly better than all the others. Thus, many inexperienced users, but also experts are not sure which solution to apply for which data collection. Unfortunately, the decision depends not only on the types of records in a collection, but also by many other qualities, such as specific requirements, user satisfaction or costs.

Choosing the optimal preservation strategy for a given setting constitutes a complex decision, depending on a large range of possibly conflicting requirements. During the first year of DELOS the Vienna University of Technology and the Dutch National Archive developed two evaluation environments, which were joint to one common DELOS Testbed. This DELOS Testbed is a powerful tool to describe the requirements of an institution and to clearly rank possible alternatives for long-term preservation.

In this session the process of the DELOS Testbed is described in detail both theoretically and with practical examples. The process of setting requirements for long-term preservation is described as well as the setting of alternatives, the weighting of preservation goals within an organization and the evaluation and ranking of alternatives. The focus is set on the workflow, but also the foundations necessary to understand the steps of the testbed are described.

The key questions that are addressed in this session include:

What are the requirements, which a preservation solution must fulfill?

- What are the criteria, which can be considered, when choosing one or another preservation approach?
- Which alternatives are viable and how to detect inappropriate preservation solutions?
- How can different alternatives be compared and how can a suggestion for one or another of them can be made?

Aims and learning outcomes of this session:

This session aims to:

- Introduce students to the concepts of the DELOS Testbed
- Provide a detailed description of the testbed and evaluation process
- Demonstrate the usability with the help of a practical example

By the end of this session students will have:

- An improved sensibility on requirements for long-term preservation
- The knowledge to apply the DELOS Testbed in their home-institution
- The ability to make informed and accountable decisions selecting one or another preservation solution for their individual requirement.

Session outline:

Lecture (30 minutes): Theoretical introduction. Detailed description of the single steps

Discussion (participants, facilitated by the presenter, 15 minutes) and short break

Activity (45 minutes): Creation of individual objective trees in groups supported by an instructor. Definition of possible preservation alternatives.

Discussion with embedded break (45 minutes): The groups present their initial results, discussion of these ideas, short break.

Activity (45 minutes): The groups apply the other steps of the DELOS Testbed and so choose the optimal preservation solution for their individual preservation environment.

Summary discussion (30 minutes)

Session (Wednesday PM)

Managing ingest – issues relevant for creating an archive.

Lecturer and Session Leader:

Birte Christensen-Dalsgaard, State and University Library in Aarhus

Abstract:

The ingest phase involve many channelling problems, both in the selection process on what to include and in the inclusion phase, where the methodology needs to be decided. The nature of the problem is linked to the legal situation for the institution where in some cases (e.g. national archives) the institution can control the formats accepted whereas in other (e.g. deposit libraries) all material and all formats will have to be accepted.

The problem of actual collecting the material have different approaches and as a case the lecture will address the situation in Denmark, where the legal deposit law has been changed and from July 1 all material published on the net will be archived and all radio- and television broadcasts will be digitised and stored in a digital archive.

Session Outline:

Lecture: There will be a one hour lecture around the problems addressed – both in their generalised form and in the more specific form related to the web archiving project in Denmark. The lecture will also address the negotiations between the national library and the Danish Broadcasting service on how to optimise the use of taxpayer's money and at the same time optimise the quality of the digital archive.

Discussion: Discussion around some of the issues raised.

Activity: The last part of the afternoon session will be around case studies, where participants will discuss the best approach for the negotiation phase and the choice of strategies for collecting and ingesting the material to the digital archive. The work will start in groups and end in a plenum discussion.

Session (Thursday AM)

Role of Registries and Representation Information

Lecturer and Session Leader:

Dr David Giaretta, Associate Director for Development at the UK Digital Curation Centre

Abstract:

The Open Archival Information Systems (OAIS) Reference Model is one of the key standards concerned with the long term preservation of digital information. The point to notice is that this is focused on information preservation rather than simply bit preservation. Representation Information is the term used in OAIS to encompass that additional information which allows one to turn “bits” into understandable information.

Representation Information can include descriptions, either machine readable or not, standards and software. Each piece of Representation Information could itself be a digital object which itself may require its own Representation Information, and so on.

A number of fundamental questions need to be addressed. How does the recursion of Representation Information about Representation Information terminate? Where can Representation Information be found? How can it be found? How can it be created? What is the difference between Representation Information and “format”? What are the different types of Representation Information? Does it have to be preserved? Who judges whether it is adequate? Can it be revised?

These are without doubt difficult questions, and ones with which many groups are struggling.

This session builds logically on the OAIS Reference Model to address these questions and many more. Examples of the many kinds of Representation Information will be presented and discussed. Over the course of the discussion we will touch on many aspects of OAIS, infrastructure to support preservation, aspects of processing digital information and the various facets of “understanding”.

Aims and learning outcomes of this session:

This session aims to:

- provide a clear understanding of the role of Representation Information in the OAIS Information Model
- give examples of Representation Information
- indicate how Representation Information can be used
- provide a view of Registries and many of the practical issues to be faced

By the end of this session students will have:

- developed an appreciation of the importance of Representation Information
- seen the significance of the recursive nature of Representation Information
- become familiar with at least some aspects of Registries
- grasped some of the ways in which the document and data traditions are different and the ways in which they converge.

Session outline:

- *Lecture* (30 minutes): The key ideas behind “information” and “understandability” in the OAIS Reference Model, and in particular the role of Representation Information
- *Discussion* (participants, facilitated by the presenter, 30 minutes): Given a stream of bits, what is needed to produce usable information? How will this change over time?
- *Seminar* (45 minutes): discussion of two or three key readings (details of these to be advised)
- *Lecture* (30 minutes): What types of Representation Information are there? How might Representation Information be shared? What are the key features of a Registry?
- *Activity* (30 minutes): How might Representation Information change over time? A number of scenarios will be used to explore some possibilities in 5, 10 and 100 years time with a variety of potential users.
- *Summary* (15 minutes)

Session (Thursday PM)

Approaches to Audit and Certification of Repositories and Preservation Processes

Lecturer and Session Leader:

Seamus Ross, Professor of Humanities Informatics and Digital Curation, and Director HATII & *Mariella Guercio*, professor in archival science and electronic record management, ISTBAL, Università degli Studi di Urbino

Abstract:

It is widely acknowledged that the purpose of audit is to check that organisations, whether public or private, meet commonly agreed or expected standards or values, and that they are doing what they are supposed to do and performing their activities in an appropriate and acceptable manner. Audits provide insight into such factors as whether resources are used effectively and efficiently; they play a critical role in ensuring good governance and accountability. Certification should provide an independent mark as to how well the organization has done in meeting these standards.

Audit aims to provide a level of certainty and confidence that we can trust organizations with which we deal. That is why audit reports are mostly made public and organizations that fail audits frequently clearly identified. While audits can provide valuable external assessments of an organizations, many want to know how they perform against their mission, targets, and processes and procedures. An audit report can provide an overview of strong and weak points and of the effectiveness and efficiency of the organization. In the area of digital preservation audits can be used to assure the public that the information that is provided by information providers, including digital repositories, still is authentic, reliable, secure, and has been maintained to the highest standards. This is critical in a digital environment where information is no longer physical or tangible and for its survival very much dependent on the fast developing information technology.

Audits require clearly defined goals, standards and procedures against which processes, services, and products can be measured. These can either be defined externally or they can be defined internally by the organization being audited. Certification, if it follows from audit, usually requires that certain that the audit produce information which enables auditors to assess its quality and completeness and to take a view as to its accuracy. In some contexts audit may be a process that happens without warning, or may be something that is only initiated by the audited organization as a way of ensuring its readiness for external spot audits.

Depositing our digital assets with external repositories requires that we are able to trust the ability of the repository to secure our information over the long term. Audit and certification can be used to reduce the unknowns associated with our use of repositories whether as depositors or consumers of information held in the repositories. Depositors and users of resources may have legitimate concerns about repositories that do not subject themselves to audits or have failed to achieve certain levels of certification. Repositories may even decide that their business cases makes it possible for them to meet certain types of certification and not others. Return on investment may be a critical factor in deciding whether or not audit and certification are paths that repositories wish to pursue.

Aims and learning outcomes of this session:

This session aims to:

- Develop an participants understanding of why we use audit and certification mechanisms.
- Provide a perspective on what audits should achieve and what we should be auditing.
- Examine whether audit frameworks support different levels of compliance (or should they?)
- Create an appreciation of the steps involved in the audit process.
- Investigate the relationship between audit and certification.
- Define the consuming communities of audit and certification services in the econtent sector.
- Cover the ways to develop a framework that can serve as guidance in an organisation with respect to creating and managing or preserving digital entities.
- Demonstrate the objectives of certification.
- Study whether audit should be part of a process to achieve certification of a digital repository.
- Offer participants a chance to understand whether developing sustainable business models for digital repositories depends upon the establishment of a viable audit and certification environment
- Establish how audit and certification services might be provided, what can be automated and what must be done manually, and how often audits and certifications may be required.
- Sketch how organizations can prepared for audits and how they should respond to a failed audit.

Session outline:

- *Lecture:* Introduction to audit and certification services.
- *Discussion:* Focused on the RLG/OCLC study on *Attributes of a Trusted Digital Repository*, 2002, <http://www.rlg.org/longterm/repositories.pdf>
- *Breakout Session:* Defining approaches to audit and certification.

Session (Friday AM)

Selection and appraisal for preservation in digital libraries:

Why? What? How?

Lecturer & Session Leader:

Ross Harvey is the inaugural Professor of Library and Information Management at Charles Sturt University's School of Information Studies

Abstract:

Librarians and recordkeepers have long acknowledged their responsibility for preserving documents for future use, and have developed criteria and processes for identifying the documents to which they will devote resources to ensure their preservation. However, these criteria and processes have been developed and applied to documents in what have been primarily paper-based collections – they do not automatically translate to digital material. They need to be revisited and modified to ensure that they can be applied effectively to digital materials. The basis of selection is defining value – and here the challenges begin. How is value defined? Value to whom? Does value change as time passes? Will users in the future define value in the same way that we do? How do we address the preservation needs of valuable material in the way most appropriate to that value? For digital materials selection decisions are ‘not a choice made once and for all near the end of an item’s life cycle, but rather ... an ongoing process intimately connected to the active use of the digital files’ (Paul Conway). The ‘digital mortgage’ that is the consequence of the selection decisions also need to be considered: ‘Program costs don’t cease when the Web site disappears’ (Diane Vogt-O’Connor).

This session notes selection criteria traditionally applied in library practice and appraisal criteria traditionally used by archives, then indicate why these selection criteria, developed for physical artefacts, do not transfer well to digital materials. It considers additional factors to be considered when developing effective selection policies and practices for digital materials, such as considering the role of intellectual property ownership, the importance of preserving context, the place of stakeholder input, and risk management approaches. Emerging frameworks for selecting digital documents are discussed. The key questions that are addressed in this session include:

- Are selection and appraisal necessary? Why can’t we preserve everything?
- What is the mandate for collecting? Who is the audience (institutional missions, users, clientele)? Are there other stakeholders we should engage?
- What do we select for preservation? What criteria might we use?
- What techniques do we have available to assist in selection and appraisal?
- How might we make selection and appraisal a routine activity in developing and managing digital libraries?

Aims and learning outcomes of this session:

This session aims to:

- Introduce students to the key challenges of selection and appraisal of digital material for long-term preservation
- Provide students with a basic understanding of selection and appraisal principles and

practices for the purposes of digital preservation

- Indicate how selection and appraisal principles and practices can be applied to digital library development and management.

By the end of this session students will have:

- Gained an appreciation of why selection and appraisal is a necessity for digital libraries
- Developed a grasp of selection and appraisal principles and practice as they might be applied to the preservation of digital materials
- Become familiar with techniques that might apply to assist selection and appraisal
- Considered how the issues are differently articulated in different digital library environments.

Session outline:

- *Lecture* (30 minutes): Why is selection necessary? What are the key issues?
- *Discussion*(participants, facilitated by the presenter, 30 minutes): The following questions will be posed and discussed: Who are the users of digital libraries? What do they require now? What will they require in the future?
- *Seminar* (45 minutes): discussion of two or three key readings (details of these to be advised)
- *Lecture* (30 minutes): What selection criteria might we use? What techniques do we have available to assist in selection and appraisal? How might we make selection and appraisal a routine activity in developing and managing digital libraries?
- *Activity* (30 minutes): Scenarios to explore the different requirements of selection in different digital library contexts. Based on a template that participants fill in, according to their backgrounds and/or their organizations, which asks for responses to questions about the users (now? in ten years time? in 100 years time?), their requirements (now? in ten years time? in 100 years time?), and the basis on which selection to meet user needs might be made.
- *Summary* (15 minutes)

Session (Friday PM)

Digital Libraries as Persistent Collections of Autonomous Objects

Lecturer & Session Leader:

Manfred Thaller, Professor for Historisch-Kulturwissenschaftliche Informationsverarbeitung

Abstract:

As with the introduction of every technology, digital libraries started with the transfer of a well known metaphor from the non digital into the digital world. We will show on detail, that currently existing digital libraries rather carefully model a situation, where a specific institution is responsible for a specific set of holdings. The same institution is responsible for the preparation of the digital material, for the implementation of the necessary tools for the access to it by the end users, for the creation of metadata and for all aspects of preservation. Larger virtual collections, which combine holdings from many institutions are currently usually realized as an after thought. We will show, that there are good reasons to assume, that this is a transitory stage and we may in the future very well encounter wide area – e.g. national – repositories which seem totally monolithic from a user's point of view, are simply a top layer upon a network of highly distributed services. At least in the cultural heritage domain, we can expect, on top of that, that the creation of meta data will at least partially be transferred into the realm of today's end users. In many ways these possibilities for further development have great promise; with some classes of repositories they might indeed lead to fundamental changes in the accessibility of so far somewhat elusive. For all concerns of preservation, they look considerably less bright, many preservation strategies of today still built upon the metaphor mentioned initially.

The session starts with a short overview of the technological possibilities for the changes described above, summarizing the main possibilities for the creation of genuinely distributed virtual repositories of digital material. We consider, furthermore, some of the discussions about changes for the classical role models of curators v. users, which have been derived from the potential of these technologies. We conclude by giving a tentative model of both, a preservation policy for truly distributed repository systems as well as a list of preservation concerns for the "autonomous information objects" from which the may be built in the future.

Turning these themes into itemized questions:

- What are truly distributed digital repositories?
- What are the autonomous building blocks they may be made of?
- What are the consequences for traditional occupational role models?
- What are the consequences for preservation?

Aims and learning outcomes of this session:

This session aims to:

- Create an understanding for some of the consequences of recent developments in distributed systems for wide area repositories.
- Create an understanding for some of the areas, where current assumptions about the further development of digital technology may be too conservative.
- Prepare the students for a situation, where the role models of specialists in libraries, archives and museums may change even more radically than so far.
- Provide them with a starting point for the planning of preservation policies which react to these changes.

By the end of this session students will:

- Be aware of some of the implications of "distributed technologies".
- Be prepared to recognize consequences of consequential wishes of the user community for preservation issues.
- Be able to react more flexibly to the resulting challenges.

Session outline:

- *Lecture* (45 minutes): What are the technological challenges of distributed repository systems for preservation?
- *Preliminary discussion* (participants plus presenter, 15 minutes): What does that mean for current practices and how may they change?
- *Lecture* (30 minutes): Examples for scenarios where further technological developments have implications for preservation models.
- *Working groups* (participants, 45 minutes): Definition of proposals how to react to such scenarios.
- *Presentation of scenarios of working groups* (30 minutes).
- *Summary* (15 minutes)

Session (Saturday AM)

Panel Discussion

Lecturer & Session Leader:

Hans Hofman, co-director of ERPANET and senior advisor at the Nationaal Archief (National Archives) of the Netherlands.

Abstract:

This final session offers participants a chance to review the outcomes of the course and to ask questions which they did not have the opportunity to ask during the week.

Biographical Sketches of the DELOS 2005 Summer School Team

- *Birte Christensen-Dalsgaard* is head of development at the State and University Library in Aarhus and is also responsible for the groups working on digital archives, the user oriented webservices and the usability work. Her work over the last five years, after she joined the Library sector coming from a computer centre working on educational material and research networks, has focussed on areas such as general library services, IT architecture for digital libraries and on preservation issues. In connection with the latter, the work on changing the Danish Legal deposit law has been quite dominating. She was in charge of the initial projects to demonstrate the relevant activities and relevant strategies in connection with archiving the web. She was member of a national committee working on recommendations to parliament on relevant measures to secure the electronic cultural heritage. Over the last year a group consisting of members from the two national libraries has worked to develop the administrative system to initiate webarchiving
- *Michael Day* is a member of the research and development team at UKOLN, based at the University of Bath, United Kingdom. His research interests largely relate to metadata development, particularly as it relates to resource discovery, interoperability and management for long-term access and preservation. He also has a general interest in digital preservation and curation issues, for example, with regard to institutional repositories or resources made available through the World Wide Web. He has published widely in these fields. Since joining UKOLN, Michael Day has worked on a large number of metadata or preservation-related projects funded by the European Commission, the Joint Information Systems Committee, the Wellcome Trust, and others. Currently, most of his activities are based on the advisory services and research work of the UK Digital Curation Centre - of which UKOLN is a partner - and the preservation cluster of the DELOS Network of Excellence on Digital Libraries.
- *Dr David Giaretta* is the Associate Director for Development at the UK Digital Curation Centre. He is one of the co-authors of the OAIS Reference Model. His background is Astronomical research, in particular using data from astronomical satellites. He has led many archive and software projects and is currently based at the Rutherford Appleton Laboratory in the UK.
- *Mariella Guercio*, ISTBAL, Università degli Studi di Urbino, is a full professor in archival science and electronic record management at the University of Urbino where she entered in 1998. For twenty years (1978-1998) she worked as State Archivist for the Ministry of Cultural Heritage where she cooperated with the Authority for information technology in the public administration to define the Italian legislation related to the electronic recordkeeping systems. She chaired the ICA Committee on current records (1990-1992) and part of the Committee on program management of the ICA itself. She chaired the Italian team for the international project InterPARES 1 (1999-2001). She is a co-director of the European project ERPANET, a network for digital preservation. She runs many Italian projects funded by the Ministry for Research in the area of digital preservation. She takes part of the digital preservation cluster of the DELOS network of excellence. Since 2002, she is the director of the journal *Archivi & Computer*. She has published widely.

- *Ross Harvey* is the inaugural Professor of Library and Information Management at Charles Sturt University's School of Information Studies. He has held academic positions at Curtin University of Technology and Monash University in Australia, Nanyang Technological University in Singapore, and the New Zealand Library School, and has been a Visiting Professor at the University of California Los Angeles. His current research and teaching interests include the preservation of library and archival material, especially in digital form. He has published widely in the fields of preservation of library and archival material, bibliographic organisation, and library education. Ross Harvey was based at the National Library of Australia from April to June 2003 as a National Library Fellow, where he researched issues associated with the preservation of digital information. A full CV for Ross Harvey is available at <http://www.elibank.net>
- *Stephan Heuscher* has recently co-founded ikeep Ltd (Digital Archives Services), a company specialized in digital archives services. For the last three years, he was the data architect in the Archival of Electronic Data and Records project at the Swiss Federal Archives. He received his Master of Science in Information Technology and Electrical Engineering at the Swiss Federal Institute of Technology (ETH) Zurich in 1999 and is currently working on a thesis on workflows in digital archives at the University of Zurich.
- *Hans Hofman* is co-director of ERPANET and senior advisor at the Nationaal Archief (National Archives) of the Netherlands. In his position at the Nationaal Archief he is involved in e-government projects and initiatives throughout Dutch government with respect to access and management of digital records and information in general. On the international scene he is co-investigator and representative of the Nationaal Archief in the Inter Pares 2 research project, and representative of the Netherlands in the ISO TC46/SC11 on Records Management. Within this committee he is chair of the Working Group on Records Management metadata. Finally, he is a researcher in the Delos 2 project on digital libraries with respect to digital preservation. He has given numerous presentations and written many articles on topics like digital preservation, recordkeeping metadata and electronic records management. Further information on Hans Hofman is available at <http://www.erpanet.org>
- *Anne R. Kenney* is Associate University Librarian for Instruction, Research, and Information Services at Cornell University Library. For nearly fifteen years she has conducted research and led workshops on digital imaging and digital preservation of cultural heritage materials. She is the co-author of three award-winning monographs, including *Moving Theory into Practice: Digital Imaging for Libraries and Archives*, and over fifty articles and research reports. She served on the RLG/OCLC Trusted Digital Repository Task Force and the joint NSF/DELOS Working Group on Digital Preservation. Kenney is a Fellow and former president of the Society of American Archivists. She currently serves on the Joint Committee on Libraries and Archives on Cuba, sponsored by the Social Science Research Council.
- *Seamus Ross*, Professor of Humanities Informatics and Digital Curation, and Director of Humanities Computing and Information Management at the University of Glasgow, runs HATII (Humanities Advanced Technology and Information Institute) (<http://www.hatii.arts.gla.ac.uk>) of which he is the founding director. Currently he is also Principal Director of ERPANET (Electronic Resource Preservation and Network)

(IST-2001-32706) a European Commission activity to enhance the preservation of cultural heritage and scientific digital objects (<http://www.erpanet.org>). He is an Associate Director of the Digital Curation Centre in the UK (<http://www.dcc.ac.uk>) and a co-principal investigator in the DELOS Digital Libraries Network of Excellence. He is a lead partner in The Digital Culture Forum (DigiCULT Forum, IST-2001-34898), which works to improve the take-up of cutting edge research and technology by the cultural heritage sector (<http://www.digicult.info>). Before joining the University of Glasgow he was Head of ICT at the British Academy and a technologist at a company specialising in knowledge engineering. He earned a doctorate from the University of Oxford.

- *Andreas Rauber* is Associate Professor at the Department of Software Technology and Interactive Systems at the Vienna University of Technology. He furthermore is head of the iSpaces research group at the Electronic Commerce Competence Center (EC3). He received his MSc and PhD in Computer Science from the Vienna University of Technology in 1997 and 2000, respectively. In 2001 he joined the National Research Council of Italy in Pisa as an ERCIM Research Fellow, followed by an ERCIM Research position at the French National Institute for Research in Computer Science and Control, at Rocquencourt, France, in 2002. He actively participates in research projects in the field of Digital Libraries, focusing on the organization and exploration of large information spaces, as well as Web archiving, and digital preservation. His research interests cover the broad scope of digital libraries, including specifically text and music information retrieval and organization, information visualization, and data analysis and neural computation. For further information on Andreas Rauber see <http://www.ifs.tuwien.ac.at/~andi>
- *Manfred Thaller* holds a chair for "Historisch-Kulturwissenschaftliche Informationsverarbeitung" – roughly: Computer Science applied to Humanities and Cultural Heritage – at the University at Cologne. He has previously worked at the Max-Planck-Institute for History, Göttingen, Germany and is a former director of the Humanities Information Technology Program at the University of Bergen, Norway. He has been visiting professor at the Hebrew University, Jerusalem, Queen Mary and Westfield College, London, and the European University Institute, Firenze. His research interests have been derived from a general definition of the requirements of Humanities' research in computing. This has in more recent years led to his involvement in a large number of digital library projects in the cultural heritage area. Manfred Thaller is also a member of the library and information systems committee of the German National Research Council (DFG). A full CV for Manfred Thaller is available at <http://www.hki.uni-koeln.de/people/thaller/mt.html>