

## **The 5th DELOS Summer School: Digital Preservation in Digital Libraries, San Miniato, Italy, 4-10 June 2006**

Michael Day<sup>1</sup>  
UKOLN, University of Bath  
<http://www.ukoln.ac.uk/>

July 2006

The 5th DELOS Summer School was held at the Centro Studi "I Cappuccini" (a former Capuchin monastery) in San Miniato, Italy on the 4th to 10th June 2006. The event was the second DELOS summer school to feature "digital preservation in digital libraries," and was organised by the DELOS digital preservation cluster, with additional sponsorship from the Digital Curation Centre, the Fondazione Giangiacomo Feltrinelli and the Fondazione Rinascimento Digitale. The academic directors of the school - Seamus Ross of the University of Glasgow (UK), Hans Hofman of the Nationaal Archief (Netherlands), and Maria Guercio of the University of Urbino (Italy) - put together a very interesting programme.

The school got under way on the evening of Sunday 4th June with a welcome and short overview session by Seamus Ross followed by a reception. Lectures started the following morning with the first of ten three-hour sessions, each led by one or more tutors. Each session had one or more breakout groups, giving participants the chance to debate particular issues in more detail and to report back findings to the whole school.

Heike Neuroth of the Göttingen State and University Library (Germany) led the opening session, a general overview of the role of digital preservation in digital libraries. This provided a good framework for the rest of the week's lectures, and emphasised the importance of understanding digital preservation as the set of processes needed to ensure that authentic objects remained accessible, usable and understandable in the future. As with some of the following sessions, key topics mentioned included the interaction of document lifecycles and preservation workflows, the key roles of standards and preservation policies, and the importance of determining the 'significant properties' of objects.

The following session, led by Wendy Duff of the University of Toronto (Canada), looked at the complex - but key - topic of metadata, focusing to a large extent on the roles of metadata in supporting records management processes and digital preservation (e.g. the PREMIS Data Dictionary). Important issues raised included the observation that metadata standards are rarely static, posing a significant problem for the longer-term management of associated objects. Another problem is that the many different factors that influence the choice of metadata needed - e.g., desired functionality, designated community, object granularity, professional background, etc. - means that no single metadata scheme is going to meet all needs. Duff highlighted the potential role of metadata registries in addressing both of these problems.

The next morning, David Giarretta of Rutherford Appleton Laboratory (UK) gave an introduction to the influential Reference Model for an Open Archival Information System (OAIS), with practical exercises looking at the different classes of information that would

---

<sup>1</sup> Draft report prepared for the *DELOS Newsletter*, issue 6: <http://www.delos.info/>

be needed for a designated community to use (or understand) an object in 50 years time.

The afternoon session, led by Andrew Wilson of the Arts and Humanities Data Service (UK) and Michael Day of UKOLN, University of Bath (UK), provided detailed introductions to object authenticity and approaches to digital preservation. As with the opening session on the first day, the presentation on preservation approaches emphasised the impact of different strategies on the 'significant properties' of objects being preserved. For example, asking whether it is always necessary to preserve the exact experience of using an object - assuming that this is possible - or whether some other kind of outcome would be sufficient.

Wednesday morning's session, jointly led by Andreas Rauber of Vienna University of Technology (Austria) and Hans Hofman, investigated the development of experimental frameworks for the evaluation of different digital preservation strategies and methods, based on work in part developed by the DELOS digital preservation cluster. Such test environments would help inform the selection of appropriate preservation strategies - e.g., taking into account the significant properties of objects and other factors - but also help to document the decision-making process itself. Breakout groups considered different types of object class and attempted to judge the respective importance of things like appearance, structure, behaviour and content. The feedback sessions revealed that, in practice, it is often extremely difficult to untangle these criteria.

The afternoon session, led by Ross Harvey of Charles Sturt University (Australia), was devoted to thinking in more detail about selection and appraisal issues. This looked at the range of existing practices developed by libraries and archives and emphasised the importance of documenting decisions about selection when preserving digital materials. Breakout group exercises prompted students to think about current and future uses of the different classes of record or information produced by particular types of organisation.

The next morning's session concerned ingest processes. Jointly led by Birte Christensen-Dalsgaard and Lars Clausen of the State and University Library in Århus (Denmark), the first presentation introduced some of the factors that needed to be considered when objects are transferred to a repository, citing experiences with handing research outputs and digital television content. A second presentation by Lars Clausen focused in more detail on the capture of Web content.

The afternoon session, by Andrew McHugh of the University of Glasgow (UK), introduced the linked topics of trust, audit and certification. The breakout group exercises involved a detailed look at the draft certification criteria published in 2005 by a task force supported by RLG and the US National Archives and Records Administration, criteria that are currently themselves being evaluated by projects led by the Digital Curation Centre and the Centre for Research Libraries.

The final day of the school commenced with a session looking at the vulnerability of digital objects once they are outside the relative safety of collections. This session, led by Manfred Thaller of the Universität zu Köln (Germany), encouraged students to think progressively about the persistence of object content (also context, authenticity and metadata) after ten, one hundred and one thousand years. The practical exercises were focused on the nature of the autonomous units that would best serve persistence.

The final session of the summer school, led by Yunhyong Kim of the University of Glasgow (UK), concerned the automatic extraction of semantic metadata - based on

research being undertaken by the Digital Curation Centre and the DELOS digital preservation cluster. The presentation looked at the role of machine-learning techniques for extracting semantic metadata from text-based documents, focusing initially on automatic genre identification and classification.

Further information on the school, including session plans and tutor biographies, is available on the DELOS digital preservation cluster website. A more detailed report of the summer school will be published later this year in the Italian archival science journal *Archivi e computer*.

**Further information:**

- DELOS digital preservation cluster website: <http://www.dpc.delos.info/>
- 5th DELOS Summer School 2006: <http://www.dpc.delos.info/ss06/index.php>
- Digital Curation Centre: <http://www.dcc.ac.uk/>
- Fondazione Giangiacomo Feltrinelli: <http://www.feltrinelli.it/Fondazione/index.htm>
- Fondazione Rinascimento Digitale: <http://www.rinascimento-digitale.it/>