

## The 4th DELOS Summer School: Digital Preservation in Digital Libraries, Sophia-Antipolis, France, 5-11 June 2005

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

July 2005

The 4th DELOS international summer school on digital library technologies was held in Sophia-Antipolis Science Park (France) on 5-11 June 2005. The event was concerned with the issues surrounding the long-term preservation of information in digital form, based on a wide-ranging and informative programme put together by Seamus Ross of the University of Glasgow and Digital Curation Centre (UK) and Hans Hofman of the Nationaal Archief (Netherlands). Students came from all parts of Europe, but also from North America, Australasia and Asia, and included both computer scientists and practitioners with experience of dealing with digital objects. The event was organised by Digital Preservation Cluster of the DELOS Network of Excellence on Digital Libraries (<http://www.dpc.delos.info/>) and co-sponsored by the Digital Curation Centre and ERPANET.

The school got under way on the evening of Sunday, 5 June, with an overview and introduction by Seamus Ross. This was followed by a drinks reception that gave the chance for tutors and students to meet each other in an informal setting.

The serious work of the school commenced the following morning with the first of ten three-hour themed sessions, each one led by one or more of the tutors. David Giaretta of the Rutherford Appleton Laboratory and Digital Curation Centre (UK) led the opening session, an introduction to the increasingly influential *Reference Model for an Open Archival Information System* (OAIS) (CCSDS 650.0-B-1, 2002). After a quick outline of its development into an ISO standard and its adoption by a growing number of digital preservation initiatives, Giaretta used the bulk of his session to provide a brief technical overview of the OAIS model itself. This included an introduction to key underlying concepts - like that of the 'designated community' - as well as a detailed outline of the OAIS information model and functional entities. While most of the entities defined by OAIS - i.e., ingest, archival storage, administration, access, data management - could be seen to be applicable to other types of service, Giaretta noted that the preservation planning function was unique, representing the *active* tasks of monitoring the technical environment and designated community to ensure that information stored in an OAIS remains accessible.

The overview of the OAIS information model was useful background for the afternoon session, which was my introduction to the role of metadata in preservation. While acknowledging that metadata was not always a meaningful concept, the initial presentation argued that collecting and maintaining metadata was essential to the long-term preservation of digital information and elaborated some of the different preservation functions that it could support, in part based on the information types identified by the OAIS information model. After a short discussion, the remainder of the session looked at

---

<sup>1</sup> A shorter version of this report appeared in the *DELOS Newsletter*, issue 4, September 2005: <http://www.delos.info/>

a range of metadata initiatives with relevance to digital preservation and gave a more detailed overview of the recently published PREMIS *Data Dictionary for Preservation Metadata* (PREMIS Working Group, 2005).

Anne R. Kenney of Cornell University (USA) led the morning session on Tuesday, focused on the development of institutional responses to digital preservation. The opening part of the presentation highlighted the key point that the digital preservation problem is not just a technological matter, but is largely a matter of organisational will and resources. After an overview of a survey of institutional readiness undertaken in 2003-2004, Kenney explained how Cornell had merged the OAIS functional entities with the attributes of trusted digital repositories identified by a working group sponsored by the Research Libraries Group (RLG) and OCLC Online Computer Library Center (Working Group on Digital Archive Attributes, 2002), and then mapped Cornell's processes to the merged model. The next part of the session outlined five organisational stages of digital preservation, from the first acknowledgement of a problem, through the response to specific threats through the initiation of projects, the increasing co-ordination and consolidation of these projects, their incorporation in the wider institutional environment, to the final realisation that preservation is dependent on inter-institutional collaboration and co-operation. These stages could be seen as a maturity model and as a way of measuring progress. The practical exercise involved pairs of students undertaking a survey of institutional readiness on behalf of one of their host organisations.

Stephan Heuscher of iKeep Digital Archives Services (Switzerland) led the afternoon session on workflows and workflow modelling. The initial presentation provided some general background on workflows and argued that they were useful for supporting automation, traceability and the sharing of implicit knowledge. The exercises involved different groups developing textual descriptions of processes, working these into workflow diagrams using a simple notation based on UML (<http://www.uml.org/>), and their refinement.

Andreas Rauber of Vienna University of Technology (Austria) and Hans Hofman jointly led the first session on Wednesday, based on work undertaken by the DELOS Digital Preservation cluster on the development of testbeds for the consistent measurement and evaluation of digital preservation strategies (Hofman, *et al.*, 2004). They noted that such testbeds could inform the selection of preservation strategies and help to document the decision-making process, as well as provide insight on issues like authenticity, costs and metadata requirements. Above all, they argued, testbeds provide a consistent experimental environment that encourages those selecting strategies to consider (and document) preservation objectives and to decide which evaluation criteria are most important. The group-based practical exercises focused on the creation of objective trees for collections of different types of digital object, based on the utility analysis-based evaluation techniques developed by Vienna University of Technology (Rauch & Rauber, 2004; Rauch, *et al.*, 2005).

Birte Christensen-Dalsgaard of the State and University Library in Århus (Denmark) led the afternoon session on the management of ingest. This focused in the main on Web resources, first providing background on the development of the Web from simple linked HTML pages to the current generation of complex, integrated online services based on content management systems. One key question was asking exactly *what* we want to preserve, e.g. whether we just want to maintain the underlying content or the user's *experience* of a site. Christensen-Dalsgaard then introduced different methods of ingest

with reference to various Danish initiatives, including the use of harvesting techniques by Web archives and institutional repositories.

David Giaretta returned to lead the next session, on the key OAI concept of representation information. The OAI model defines this as the "information that maps a Data Object into more meaningful concepts," giving the example of how the ASCII specification describes how a sequence of bits can be mapped into a particular symbol. With exercises based on the idea of a traveller from the far future asking questions about today's digital objects, Giaretta led the school through the multiple layers of representation information (e.g., media, stream, structure, object and application) and argued that the existence of persistent and trusted registries of representation information (not just file format registries) would support the sharing of effort.

Seamus Ross of the University of Glasgow and Digital Curation Centre (UK) led Thursday afternoon's session on the audit and certification of repositories. The session began with reminder of the importance of trust and authenticity in digital contexts followed by an outline of current activities, including the ongoing work of an international Task Force on Digital Repository Certification supported by RLG and the US National Archives and Records Administration ([http://www.rlg.org/en/page.php?Page\\_ID=367](http://www.rlg.org/en/page.php?Page_ID=367)). The practical exercises enabled small groups to consider in more detail the nature and characteristics of the certification process, including the identification of questions that could be included in a self-assessment prior to full audit.

Ross Harvey of Charles Sturt University (Australia) began his session on selection and appraisal for preservation with a recapitulation of some of the week's common themes, emphasising the importance of making documented decisions about selection when we preserve digital materials. The session outlined current thinking on selection and appraisal, which tends to assume that maintaining access to *all* digital material is neither practical nor desirable, and which favours a cautiously selective approach based on the preservation of material deemed to be of some continuing value. Harvey noted that digital materials were different from traditional ones because preservation decisions needed to be made early in their lifecycle, also noting the problems of quantity and high maintenance costs. Group exercises focused on the current and future usage of specific collections of digital materials and the development of selection criteria.

Manfred Thaller of the University at Cologne (Germany) led the final session on digital libraries as persistent collections of autonomous objects. This extremely thought-provoking session emphasised the importance of autonomy, i.e. that digital objects must be able to survive even if the system of which it was a part is destroyed. Practical exercises prompted small groups to consider the properties of 'self-defining' files and to structure a particular knowledge domain.

From the tutor's point of view, the 2005 DELOS Summer School was a useful opportunity to spend a week discussing digital preservation issues with highly informed and motivated set of students. Initial feedback also suggested that the school had been useful for students, raising the possibility that the event could be repeated. The venue was slightly less successful. The geographical nature of Sophia meant that evening social events depended very heavily on the few opportunities available within the park itself and on the limited public transport links with nearby Antibes and Juan-les-Pins. Despite this, all summer school participants forged a very successful social life.

Throughout the week, there were a number of common themes. Perhaps the most important of these was the importance of documenting processes, whether these be related to selection, ingest, or the evaluation of preservation strategies. Other key

themes related to the importance of the user perspective, the need for better co-operation between institutions, and the problem of managing massive (and rapidly growing) quantities of heterogeneous metadata. Further information on the summer school, including session plans and tutor biographies, is available on the DELOS Digital Preservation cluster Website (<http://www.dpc.delos.info/>).

## References

CCSDS 650.0-B-1. (2002). Reference model for an Open Archival Information System (OAIS). Retrieved July 22, 2005, from <http://www.ccsds.org/documents/650x0b1.pdf>

Hofman, H., *et al.* (2004). "Framework for testbed for digital preservation experiments." DELOS deliverable D6.1.1. Retrieved July 22, 2005, from <http://www.dpc.delos.info/outputs/index.php>

PREMIS Working Group. (2005). *Data dictionary for preservation metadata: final report of the PREMIS working group*. Dublin, Ohio: OCLC Research. Retrieved July 22, 2005, from <http://www.oclc.org/research/projects/pmwg/>

Rauch, C., & Rauber, A. (2004). "Preserving digital media: towards a preservation solution evaluation metric." In: Proceedings of the 7th International Conference on Asian Digital Libraries (ICADL 2004), Shanghai, China, December 13-17, 2004. Lecture Notes in Computer Science, 3334. Berlin: Springer-Verlag, pp. 203-212. Retrieved July 22, 2005, from <http://lib.ccnu.edu.cn/icadl2004/papers/3334/33340203.pdf>

Rauch, C., Pavuza, F., Strodl, S., & Rauber, A. (2005). "Evaluating preservation strategies for audio and visual files." In: Proceedings of the 9th DELOS Network of Excellence thematic workshop: digital repositories: interoperability and common services, Heraklion, Crete, May 11-13, 2005. Retrieved July 22, 2005, from <http://delos-wp5.ukoln.ac.uk/forums/dig-rep-workshop/rauch.pdf>

Working Group on Digital Archive Attributes. (2002). *Trusted digital repositories: attributes and responsibilities*. Mountain View, Calif.: Research Libraries Group. Retrieved July 22, 2005, from <http://www.rlg.org/en/pdfs/repositories.pdf>