

Report on the BL/CURL DPC Forum held at the British Library Conference Centre, Tuesday 19th October 2004.

To open PDFs you will need [Adobe Reader](#)

The 9th DPC Forum was a collaboration between CURL and the British Library. The theme of institutional repositories was proposed by CURL as being very timely as the move from theory to practice is likely to accelerate, requiring more emphasis on sustainability and lessons learned from the practical experience of early adopters. Clifford Lynch's quote from a recent RLG DigiNews : 'An institutional repository needs to be a service with continuity behind it.....Institutions need to recognize that they are making commitments for the long term.' Clifford Lynch, 2004 http://www.rlg.org/en/page.php?Page_ID=19481#article0 was used in promoting the Forum and several presenters used other pertinent Lynch quotes. Themes emerging from the day were that there were many challenges, but it was important to continue to gain practical experience and build on experience and expertise. Some speakers also referred to the current need to provide mediation for depositors of content but that this was not scaleable. Ways and means of enhancing efficiency included shared tools and services, such as the PRONOM file format registry, and automating parts of the ingest processes.

In opening the Forum, Richard Ovenden, Keeper of Special Collections at the Bodleian Library, set the institutional repository scene, as one in which there is a gradual progression from theory to practice but uptake has been slow ([Introduction](#) PDF 108KB). The purpose of this Forum would be to hear from the early adopters, and listen and learn from them. The role and commitment of CURL to institutional repositories and digital preservation was seen at task force level, in individual CURL institutions, and through consortial activity. The role of the DPC in setting the digital preservation agenda was now well known and its value in training, information exchange and providing advice and guidance was a valuable asset.

Delegates were referred to the JISC press releases contained in their packs, which provided details of the successful proposals from the recent 4/04 Call on Digital Preservation and Institutional Asset management and also the forthcoming Repositories programme call, which will be the subject of two further calls in 2005 and indicated a major step forward and a major investment by JISC.

Session 1 was chaired by Paul Ayris, Director of Library Services, University College London, who introduced the first presentation by William Nixon, Deputy Head of IT Services at the University of Glasgow who presented a paper ' [From ePrints to eSPIDA: Digital Preservation at the University of Glasgow](#) ' (PDF 822KB). A number of questions had been raised by the Glasgow experience, which had started as a pilot service in 2001. Digital preservation was not the primary focus as there was no content to preserve, but was becoming more of an issue and providing the greatest challenge. We need rigorous, robust preservation options if we are to move to the non-print world. William also suggested that this may well prove to be a selling point for academics in encouraging them to

deposit their papers with the repository. In reviewing progress to date, Nixon said that there was a need to transition from project funding to embedding repositories into the bottom line of institutions so that they can make a stewardship commitment without dependence on project funding and move towards becoming a trusted digital repository.

John MacColl, Sub-Librarian, Digital Library, University of Edinburgh, and and Jim Downing, Preservation Development Manager, DSpace@Cambridge provided two perspectives of DSpace, as a manager of a repository service, and as a developer of the preservation aspects of DSpace. John MacColl drew attention to the services arising from project funding but which could potentially fall into disrepute unless they are properly managed over time ([DSpace MacColl Presentation](#)

[MacColl Presentation](#)

PDF 655KB). Digital preservation could be regarded as a high cost for individual institutions to undertake and it might be necessary to make use of other facilities. Advice and guidance were needed by the library community and the Edinburgh would be looking to the DCC as a source of that technical and practical guidance.

Jim Downing described the DSpace at Cambridge repository in which there are no mandates on type of material or file formats but they do actively provide advice on good practice ([DSpace Downing Presentation](#)

[Downing Presentation](#)

PDF 166KB). Better preservation metadata was needed to support preservation planning. Tools such as PRONOM, which are already available, are proving valuable in helping to provide monitor technological obsolescence. Cambridge have been advised to retain human readable action plans and to add automation, wherever feasible/appropriate, but to retain human validation of automated steps. Currently DSpace at Cambridge records all item and metadata changes but this would not be scaleable. It would be necessary to refine policy and implementation.

The final session of the morning was a joint presentation on Storage Resource Broker (SRB) at the AHDS ([SRB Presentation](#) PDF 1.2MB). Hamish James provided an overview of what SRB is and its role at AHDS. The SRB software assists in managing digital objects scattered around multiple locations, a clear benefit for a distributed service such as AHDS, which was moving from a loose federation of repositories to a much more centralised preservation service, while still maintaining its distributed nature. The collection was expected to grow to 10 TB within the next two years, so any service must be scaleable. Andrew Speakman then outlined some of the practical issues involved in installing SRB. Andrew drew attention to a frequently recurring theme in any discussion of digital preservation, that of collaboration and the need to take advantage of related effort which has already occurred. He also went on to outline the pros and cons of SRB, pros included the ability to handle large networked data volumes and high user acceptance. On the negative side, technical support is not well advanced so there is a requirement for significant in-house expertise as it is quite complex to install. In concluding Andrew said that SRB has the potential to simplify day-to-day operations and also to simplify distributed management of data and indicated that the AHDS was looking for partners using SRB.

The afternoon session was chaired by Richard Boulderstone, Director eStrategy, the British Library and began with a presentation ' [Preserving EPrints:Scaling the Preservation Mountain](#) ' (PDF 144KB) on the SHERPA project presented by Sheila Anderson and co-authored with Stephen Pinfield. Sheila outlined the SHERPA project objectives and partners Nottingham (lead), Edinburgh, Glasgow, Leeds, Oxford, Sheffield, York, the British Library, and AHDS. SHERPA is primarily concerned with e-prints, i.e. a digital duplicate of an academic research paper that is made available online as a means of improving access to the paper.

Differing views have been expressed on whether it is necessary to preserve these documents but there is an opportunity here to move beyond saving and rescuing digital objects to building the infrastructure required to manage them from the start. A good start has been made in identifying properties of e-prints, looking at selection and retention criteria, preferred formats, rights issues etc. but none of these are 'doing' preservation. Using the OAIS model as a guide, a preservation storage layer and preservation planning (e.g. policies and procedures, risk assessment) needs to be added, with preservation and administration metadata and preservation protocols and processes in place.

A new two-year project, known as SHERPA DP, which is being led by AHDS in partnership with Nottingham and 3-4 SHERPA partners and funded under the recent JISC 4/04 Call has recently been announced. The aim of SHERPA DP will be to develop a persistent preservation environment for SHERPA partners based on the OAIS model and to explore the use of METS for packaging and transferring metadata and content. A Digital Preservation User Guide would be another practical deliverable from this project. The preservation community would be looking to the DCC for support, particularly in functions which are most appropriately centralised, such as technology watch.

The final presentation was from David Ryan, Head of Archives Services and Digital Preservation at the National Archives, 'Delivering digital records: towards a seamless flow'. David described the development of the Digital Archive and key points needed for its success ([TNA Presentation Part 1](#)

PDF 96KB), which were a strong business case linked to core organisational aims, a good team, and the need to sell the fact that this is not an insuperable problem. It has taken three years for the Digital Archive to become a comprehensive service delivery and all business targets have been met but it is critical to recognise that stewardship is a long-term evolving business. In recruiting staff it was essential to have the right technical skills, combined with the ability to sell the work to others within the organisation (

[TNA Presentation Part 2](#)

PDF 90KB). The reality is that we must collect e-records. The Digital Archive should be scalable to 100TB, which is way beyond current storage requirements though it is rapidly growing (

[TNA Presentation Part 3](#)

PDF 1MB). TNA works with government departments but the current procedures, which tend to be case-by-case and handcrafted, was not scaleable (Editor's note: a similar point was made by William Dixon in Glasgow's experience of building their repository). Preservation planning is a key feature of the Digital Archive, which must be able to accommodate changes in preservation management over time. The main thing is to ensure that the bitstream remains unharmed incase a different preservation strategy is adopted (the current strategy is migration). Other TNA digital preservation effort includes the PRONOM service (

[TNA Presentation Part 4](#)

PDF 613KB), which is now on Version 4 and is designed to be the primary file format registry. PRONOM can be used to help decisions about migration planning because it can indicate when a file format is likely to become unsupported. The UK Central Government Web Archive has captured c. 60 web sites to date and is currently held separately from the Digital Archive but it was intended to bring the two together. An issue is the size of the government website domain. Finally the work of NDAD was described, and their role as contractor for TNA in preserving data sets. Next steps would include a comparison of the NDAD data model and the digital Archive data model. In closing, David said that trusted digital repository certification was a key issue and there was a need for a process to allow a federated system of preservation and access.

A final panel session allowed delegates to put questions to all the speakers.