

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

BioCentre, York Science Park (<http://www.yorksciencepark.co.uk>) York, Friday 22nd April, 2005.

The DPC Meeting on Mass Storage Systems was held in York on 22nd April. The meeting was open to DPC members only and was intended to be an informal discussion of mass storage systems, structured around the latest DPC Technology Watch report, [Large Scale Archival Storage](#) and authored by four members of the DOM team at the British Library. Richard Masters, Sean Martin, Jim Linden, and Roderic Parker led discussion of the decision-making and planning which led to development of their storage system. The [PP slides \(in PDF 433KB\)](#) for the meeting are available.

The presentation on the storage system included the importance of having a clear mission statement for the DOM Programme, and the pragmatic decision to adopt a generic, cost-effective, and incremental approach. Major drivers for the programme were discussed, including legal and voluntary deposit and Richard Masters referred to the e-journal pilot being undertaken with volunteer publishers, to test how legally deposited e-journals will be delivered to the BL. Other categories of material includes the BL's digitised collections, sound archive, web archiving, and Ordnance Survey material. This comprises both a large volume of digital material and also a wide variety of formats.

While the decision was to purchase off-the-shelf products wherever possible, it had not been possible to purchase a storage system which met all of the BL's requirements. Principles which needed to be considered included the need for material to be invariant over time (which proved to be a fundamental difference with many commercial approaches); the need to assign an internal, unique identifier; the need to ensure that there would be no extended loss of service; and the need to ensure both integrity and authenticity. The latter needs to be more than simply checking that a file hasn't changed and the team had conducted a key generation ceremony to ensure this condition was met. This provides a trust model which ensures that a bit-stream remains unchanged after decades, despite changes of hardware during that timeframe.

Resilience of the system will be provided by having multiple sites (initially there will be one at Boston Spa, one at St Pancras), which can currently hold 12TB of storage, and a third "dark archive" to be held in another location. The multiple site design provides disaster tolerance by enabling the service to continue despite the loss of a storage site. The role of the dark archive is to provide the ability to recreate the DOM store in the extreme case that all sites are destroyed - this would be done by re-ingesting all objects from the dark archive into a new site.

The concept of total cost of ownership was outlined, Jim Linden led the meeting through elements of total cost, including initial purchase, the cost of operations (where staff costs are significant), data centre costs and application support and enhancement. It was decided that performance of commodity storage was adequate for preservation storage. It had been necessary to plan and decide on features that did not add value for the BL's needs (even though several commercial vendors felt they would provide benefits, it was necessary to articulate the BL's specific requirements, where many of these added extras were not required). Issues still needing to be considered were emerging technologies, such as the MAID concept of power saving. There are also a number of placeholders for future work, for example the assumption that the same 80/20 rule for accessed material which holds true in the print world, needs to be tested in the digital world.

It was a very informative and stimulating session and I'm grateful to the authors for taking the time to talk through their approach. One suggestion on the feedback forms for additional themes for similar meetings was preservation metadata and it may be of interest that the next Technology Watch report has recently been commissioned from Brian Lavoie of OCLC and Richard Gartner and Michael Popham of the University of Oxford and is on Preservation Metadata. This report should be ready for peer review in July 2005.