

E-journal Archiving and Preservation Workshop

Overview of an event organized by the British Library, the Digital Preservation Coalition and JISC at the BL Conference Centre on 27 March 2007

EMMA CASS

Business Manager, British Library

HELEN HOCKX-YU

Formerly Programme Manager, Joint Information Systems Committee
now Planets Project Manager, British Library

CAROL JACKSON

Administration Manager, Digital Preservation Coalition

PHILIP POTHEN

PR Manager, Joint Information Systems Committee

SAMANTHA TILLET

Product Development Manager, British Library

In March 2007, the Digital Preservation Coalition (DPC), the Joint Information Systems Committee (JISC) and the British Library invited attendees to a workshop focusing on the archiving and preservation of e-journals. Speakers from the publishing and library environments came together at the British Library to discuss developments and requirements in the area of e-journal archiving. The event was attended by over 100 delegates and the speakers' presentations were followed by a lively debate enabling all attendees to put their ideas and questions forward. The speakers' presentations for the 2007 workshop are available on the Internet¹ and it is intended that the event will be an annual occurrence.

The uncertainty over the continuing access to scholarly journals has been highlighted as a major issue, accentuated as subscribers move to e-only access. As a first of its kind in the UK, the E-Journal Archiving and Preservation Workshop brought together international stakeholders from the information world, including publishers, librarians and representatives from archiving solution providers. The workshop, chaired by Hazel Woodward (Cranfield University), provided an opportunity to review the initiatives that are currently in place and discuss the developments which are required.

In the first session, Anne Kenney (Cornell University) provided an overview of the e-journal archiving landscape, drawing on work that was done in 2006 to pull together the opinions of library directors and also providing a detailed review of 12 archiving solutions. The resultant report, *Journal*

*Archiving Metes and Bounds: A Survey of the Landscape*², highlighted three emerging themes:

- sense of urgency
- libraries have resource commitment and competing priorities
- need for a collective response

Kenney believes that legislation needs to be implemented to ensure that all content is preserved and that none is left at risk. She stressed that it is also important that more than one archiving solution is developed, as it is impossible for one service to cater for all requirements. Indeed current content coverage by the 12 archiving solutions which were reviewed is certainly not all-inclusive: each service has its own motivation, which can mean that the most at risk content is not preserved. Publishers and libraries need to work together

with e-journal archiving programmes to ensure that scholarly works are not lost and that future needs are met.

The keynote presentation was followed by presentations from four different providers of archiving and preservation services.

Victoria Reich, Director of the LOCKSS programme, gave an overview of LOCKSS and CLOCKSS. The LOCKSS (Lots of Copies Keep Stuff Safe) programme is a method whereby libraries can work together to ensure that content is perpetually available using the LOCKSS software which enables libraries to collect and preserve content with permission from the publisher. Each library involved in LOCKSS is part of a network so that multiple copies of an article are preserved. Reich described "how the mesh of the whole is stronger than any single silo" and thought that the library should continue to be the custodian of the content in a digital age. CLOCKSS (Controlled LOCKSS) uses the LOCKSS software but is overseen by a board of publishers and libraries. Reich also demonstrated that the content most at risk is not necessarily the most read or cited and this type of content is what needs to be included in archiving and preservation programmes.

Eileen Fenton, Executive Director of Portico, described the service provided by Portico which archives peer review scholarly journals and to date includes more than 5,800 journals from 30 publishers. Portico's approach can be described as an 'archive of last resort' where material is made available following certain trigger events, for example if a publisher ceases operation or if back issues are removed from a publisher's offering and are not available anywhere else. Portico is funded by annual contributions from publishers and libraries and offers a service which meets the needs of both. Fenton showed that the intellectual property of the content is seen as the most important thing to preserve; for example, the text, tables, images and supplemental files, rather than the look and feel of the original product. Content is transformed into a version of the NLM Archiving DTD for storage purposes and, to ensure future availability, it will be migrated to new formats as old ones become obsolete.

Erik Oltmans, Head of Acquisitions and Cataloguing at the National Library of The Netherlands (Koninklijke Bibliotheek), spoke of the KB's e-Depot service which, from a starting point in 1995, is now ingesting 50,000 items a day from

10 publishers. Preservation and archiving principles have become embedded in the library's practices and procedures since this started over a decade ago. He also highlighted that the service includes a variety of material from e-journals to e-books to CD ROM content and if a title stopped being available then it would be made available to all via e-Depot. Oltmans believes that as publishers are unlikely to deposit content at a large number of places, a network of trusted partners is essential. As e-journals do not necessarily have a natural country of provenance, there is not always an obvious guardian of such content. This means the current geographical model for preserving content is insufficient to guarantee long-term access.

The UK national library perspective was given by Richard Boulderstone, Director of e-Strategy at The British Library, who showed that the traditional scholarly communication process is changing and that this requires new processes and durable technical architecture to ensure that archiving solutions are able to satisfy requirements now and in the future. Long-term funding is necessary to ensure that solutions are long-lasting and Boulderstone reiterated the need for partnerships, and for organizations to work together. The British Library is currently developing its Digital Library and the system will start ingesting content in the Summer of 2007 under the auspices of the e-journal voluntary deposit scheme. The Digital Rights Management layer will be implemented in 2007/08 to enable access beyond the British Library.

The afternoon session saw presentations from the publisher and university library perspectives. Steven Hall, Commercial Director at Wiley-Blackwell, highlighted statistics which show that 50% of all serials publications would be online-only by 2016, while Blackwell's forecast suggested that in the sciences, 39% of journals would be online-only by the end of this year. Publishers take the issue of archiving seriously but there needs to be more clarification when it comes to access. Preservation is not the same as access and access is not the same as open access. Publishers will support a library's legitimate perpetual access rights, but it is important not to conflate these with the very different questions of access and of open access. Wiley-Blackwell is taking part in many different initiatives, like Portico, e-Depot and LOCKSS/CLOCKSS to ensure that their subscribers will retain their entitled access to content.

Paul Ayris, Director of Library Services at University College London (UCL), said that UCL is moving towards the e-only delivery of journals and is therefore reviewing, through its information strategy, what solutions need to be in place to ensure continued access. A working group which develops procedures and practices for digital curation has identified key questions concerning digital preservation and life cycle questions. The aim of moving towards e-only delivery of journals is so that STM researchers will not need to be physically present in a library and to deliver materials remotely and 24 hours a day, 7 days a week. A pilot project has seen the cancellation of a large number of print copies, noting those requests for print copies which are not able to be fulfilled. No such requests, Ayris reported, have so far been made. With libraries increasingly moving to the e-only delivery of journal literature, digital curation needs to be embedded in institutional strategies, said Dr Ayris. This depends, however, on robust digital curation arrangements to underpin service delivery.

UCL has also been involved in the LIFE³ project, a joint British Library/UCL project funded by JISC to develop a generic preservation model for costing digital curation at an item level. Additional funding by JISC is enabling the project to explore further the economic modelling and to develop more case-studies to test the models. The results will establish benchmarks for local digital curation services in a university or national library.

The day ended with a question and answer session where attendees were able to quiz an expert panel of publishers, librarians and archive service representatives. Topics varied from whether copyright hinders or helps preservation to what is being done to archive multimedia formats and supplementary data. Attendees were concerned that solutions will purely archive the textual element of a journal and not the additional data to which it can be linked. The issue of standards was also raised as something that needs to be looked into further. Standards can potentially help to make archiving and preservation techniques easier as stakeholders would be working to the same principles.

Hazel Woodward concluded the day by summarizing the key themes which had been raised, namely: that terminology needs to be specified so that there is a common understanding about actions required for archiving and preservation;

that more work needs to be done to gather information on what content is being archived; and finally, that collaboration and co-operation are central to all archiving and preservation activities.

Ongoing activities

In pursuit of the second of these goals – the sharing of information about what is being archived – JISC is funding a scoping study into the feasibility of a registry of scholarly publications to indicate which archiving programmes are preserving them. Such a registry would not only identify gaps and overlaps in content coverage, but also encourage collaboration and shared responsibility between publishers and archiving services, and among archiving services themselves. This in turn would enable research libraries to make an informed choice when selecting providers for e-journal archiving and preservation service.

JISC will also be reviewing the implementation of LOCKSS in the UK prior to the end of the pilot programme in February 2008 as well as funding the second phase of the LIFE project, mentioned above.

Further collaborative initiatives continue through the work of the Digital Preservation Coalition (DPC) and specifically, in this area, its work to explore the potential of the JPEG2000 image compression technique, a possible new archiving format for digital images. JPEG2000 allows a wide range of uses and can support a wide range of formats and multiple resolutions. It is also a flexible file format which allows metadata to be built into the file, a vital element of the digital preservation process.

A workshop was held in June on JPEG2000, following which a report is being prepared that will provide recommendations for further action.

Among international initiatives in this area is the Planets project, which brings together European National Libraries and Archives, leading research institutions and technology companies to address the challenge of preserving access to digital cultural and scientific knowledge. Other digital preservation projects funded under the Framework 6 'Information Society Technologies' programme⁴ include CASPAR, which is researching, implementing and disseminating innovative solutions for digital preservation based on the OAIS reference model⁵ and Digital Preservation Europe (DPE), which is supporting

collaboration between many existing national initiatives in the European Research Area⁶.

The British Library is currently calling for UK publisher volunteers to take part in its e-Journal Voluntary Deposit Scheme and will start ingesting content in Autumn of this year. Currently, requirements are that content is formatted according to the NLM A&I DTD v2.3 but specifications will be broadened as the scheme progresses. The Library will also build on the work of the Voluntary Deposit Scheme to offer a Digital Archive service to non-UK publishers.

Bounds: A Survey of the Landscape, Sept. 2006, Washington, DC, Council on Library and Information Resources

<http://www.clir.org/pubs/reports/pub138/pub138.pdf>

3. <http://www.life.ac.uk/>
4. <http://cordis.europa.eu/ist/>
5. <http://www.casparpreserves.eu/>
6. <http://www.digitalpreservationeurope.eu/>

Article © Emma Cass, Helen Hockx-Yu, Carol Jackson, Philip Pothen and Samantha Tillet

References

1. <http://www.dpconline.org/graphics/events/0703ejournalwkshop.html>
2. Kenney, A R, Entlich, R, Hirtle, P B, McGovern, N Y and Buckley, E L, *E-Journal Archiving Metes and*

■ Emma Cass
Business Manager
The British Library
E-mail: Emma.Cass@bl.uk

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=3&page=221>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=3>