PALS Conference: Institutional repositories and their impact on scholarly publishing

Royal College of Obstetricians and Gynaecologists, Regent's Park, London
24 June 2004

Report by Anne Partridge, Systems Librarian, Cranfield University

The PALS (Publisher and Library/Learning Solutions) conference provided a forum for interested parties to engage with the concept of and issues surrounding the current 'hot topic' of institutional repositories. The audience included publishers, library and information workers, academics and funding agency representatives.

Throughout the day each speaker gave their own definition and slant on the topic, and whilst there was not universal agreement on the precise definition of an institutional repository or their scope, it is fair to say all saw them as web-based, institution-focused archives of scholarly communication. They are seen by some advocates of self-archiving as a more promising route to open access than subject-based archives, although the latter have been very successful in a few disciplines. Their contents, which are generally freely available, can include journal article e-prints (both preprints and postprints [author postprints and publisher PDFs]), theses and dissertations, technical reports, working papers and other grey literature, conference proceedings, books or chapters of books, teaching materials, datasets and other digital material. They include both material contributed by the institutional community and material documenting the lifeline of that community. Their role in the preservation of such material and their potential impact on scholarly publishing were subject to discussion.

The day kicked off with a keynote speech from *Clifford Lynch* of the Coalition for Networked Information. Clifford provided a strategic overview of the forces that are driving changes in scholarly communication and the development of institutional repositories. He questioned the archive role of institutional repositories by arguing that not everything should stay there for ever. He suggested that material that is embargoed or limited to the institutional community shouldn't be precluded on open access grounds. Clifford also challenged whether current scholarly publishing was the best mechanism for scholarly output, given the heavy reliance on datasets, software and sensors, particularly in the sciences.

Mark Ware (Mark Ware Consulting Ltd) reported on the current situation in the development and uptake of institutional repositories, based on his recent pathfinder research for PALS. He argued that there was now a compelling case to use institutional repositories in the infrastructure for scholarship in the digital age. This was due to a number of reasons including reduction in technology costs (especially storage), standards (e.g. OAI-PMH, metadata), growing awareness of need for long-term preservation, and developments in web publishing (Open Archives Initiative, open access journals, disciplinary repositories). There are also increasing numbers of institutional repository software packages, many of which are freely available. The research showed that whilst the number of repositories is growing steadily (over 129 e-print repositories were identified), the quantity of documents deposited was relatively small, on average 350 e-prints per repository. The type of documents generally fell into three categories: e-prints, theses and dissertations, and grey literature. The impact on publishing was unclear. Due to the low current deposit rates, there is little to indicate that institutional repositories are reforming scholarly publishing or beginning to tackle long-term preservation issues.

Chris Awre (JISC) reported on the growing interest in repositories amongst the HE and FE communities. He commented on the reasons why an institution may want a repository (to

disseminate and raise the profile of the institution's research output, a complementary form of publishing, institutional management of its assets and long-term digital preservation) and then gave an overview of the FAIR (Focus on Access to Institutional Resources) programme. This consists of 14 projects whose objectives are to investigate the deposit and disclosure of institutional assets and increase the understanding of the related technical, organizational and cultural processes.

Raym Crow (SPARC Consulting Group) provided an overview of current repository software systems, the functionality offered and implementation issues. He covered ten systems (all complying with OAI metadata harvesting protocols and available publicly via an open source licence). They support the expected types of content – preprints, postprints, theses and dissertations, teaching materials and conference proceedings. Whilst content submission and management processes are available, it was interesting to note that none yet has a fully implemented peer review mechanism – a function still retained by publishers. Few barriers to system implementation were identified, although multiple hurdles for content recruitment were recognized.

The afternoon session consisted of four case studies, beginning with a description of the eScholarship Repository launched by the California Digital Library. This was presented by *Greg Tananbaum* (Berkeley Electronic Press) who outlined some of the reasons why Berkeley Electronic Press (a commercial technology provider) was formed. These included addressing the 'double dip' (where intellectual output funded by universities is captured by peer-reviewed journals and sold back to university libraries at increasing cost) and providing outlets for monographs and other speciality publications that would not be commercially viable to publish using traditional routes. The commercial system developed has an impressive set of features including peer review functionality, currently absent in all the non-commercial systems. The presentation highlighted the importance of selling participation to academics and ended with a list of open-ended questions on the future position of institutional repositories and their relationship to current forms of scholarly communication.

Steve Probets (Loughborough University) spoke next on the RoMEO (Rights of MEtadata for Open Access) Project. This JISC-funded project was set up to investigate the issue of intellectual property

rights in the self-archiving process. The project undertook an analysis of existing copyright transfer agreements, investigated what authors can legally do with their work and carried out a survey of academic authors to determine their views on what permissions, restrictions and conditions they though appropriate to assign to their work in an institutional repository. Of particular interest was the issue of who owns the copyright, whether institutional repositories ever double-check copyright status of submissions and the fact that an institutional repository provider making available non-copyright cleared work could be liable to second-hand infringement. Steve also described a follow-on project which is developing ways that will allow authors and repositories to describe the ownership and allowed uses of copyright materials in a machine-readable way.

Leo Waaijers (DARE Project) focused on the Dutch universities' Digital Academic Repositories initiative. The DARE programme aims to give cheaper and quicker access to the results of academic research. Its premise is a community co-ordinated approach using harvesters to extract metadata from institutional repositories for use in educational, research and other applications. The recent entry of major search engines such as Yahoo and Google into the area provides an interesting parallel with traditional publishing. Google, for example, impact-ranks sites by placing those with the highest number of links at the top. This is comparable to the ranking mechanism applied in the journal world based on citation.

The day ended with *Richard O'Beirne* (Oxford University Press) and *Johanneke Sytsema* (Oxford University Library Services) discussing their collaboration during the SHERPA (Securing a Hybrid Environment for Research Preservation and Access) project. In SHERPA, Oxford University is setting up an open access institutional e-print repository. Oxford University Press is contributing by providing online access to a range of articles published in OUP journals by Oxford-based authors. Richard described the benefits to a publisher (including increased discoverability, increased usage and readership), and spin-off related activities (including e-offprints [and links to full text] for authors).

Further details on the papers presented at the conference can be found at:

http://www.alpsp.org/events/previous/PALS04 prog.htm