Moving beyond e-journals

Paul Ayris explains to **Elspeth Hyams** why scholarly communication has moved beyond the debate on e-journals pricing and open access.

The open access movement represents 'a tremendous revolution', according to Dr Paul Ayris, Director of Library Services, UCL (University College London), and it is as important as the internet. 'If the internet is about connectivity and the network and technical side, open access is about the content you access and use once you have the connectivity – the two go hand in hand.'

For that reason, it's not just a research concern, it's a social issue, he says. It's about 'the cost of citizenship, including the cost of the information you need to get hold of to exercise your rights as a citizen'. And it is about freedom of information. For that reason, he is 'puzzled' why more isn't made of it, as the key that unlocks access for everyone.

A revolution in service delivery

Meanwhile, the library community is at a critical stage in the transformation of its own service delivery models. According to recent research by OCLC,[1] only two per cent of students in colleges find material that the library purchases using the library's catalogue. And that represents a real challenge: 'If you don't mobilise this revolution – which is here, happening now – the information that is available in libraries and information centres is going to be invisible to the general population. It's not something you can stop. So how to harness it and make it available is the question.'

Paul is among the most active campaigners for the OA movement in Britain. A lot of the debate in the media has focused on journal prices and 'distortions' of the market through publishers' 'big deal' packages. Undoubtedly, rising prices precipitated a crisis for librarians 10-15 years ago, but now, he thinks, the focus has shifted. It is more about the democracy of access.

Take the subject of health. Along with many of the biosciences research community, he believes that patients should have access to high-quality research, because quite a lot of what is available for free on the internet 'might kill you if you used it'. It's an important example of why 'access' has become a social, not a technological, issue, and he is surprised that patient groups in the UK have not fol-

lowed their counterparts in the US and lobbied for patients' rights in this area.

But while publishers are still discussing the best way to fund the publication of journals that are free at the point of use, thinking in the academic community has moved on. Nowadays the talk is all of scholarly communication. This term used to refer only to the communication of research results to fellow researchers. Until recently, even Sconul (Society of College, National and University Libraries) and Curl (Consortium of Research Libraries) defined it as 'published dissemination and rating of information produced by academics, either for teaching or research, irrespective of format'. The trouble is, that definition is now too narrow.

Scholarly communication: a new definition

Paul is proposing a new definition because, he says, 'Scholarly communication encompasses everything that researchers, teachers and learners need in order to be effective.' According to this vision, the definition should be: 'The authoring, publishing, dissemination and reading of information produced for teaching, learning or research, in whatever format, with the tools, measures and systems needed to provide access to and store these materials in perpetuity.'

Institutional repositories

So, where are we now? In recent years, part of the universities' response to high journal subscription prices has been to set up institutional repositories. These are storehouses where researchers can deposit not just a copy of the article ('pre-print' or 'post-print') published in a journal elsewhere, but the research datasets on which some of the researchers' conclusions are based. (In fact, the whole e-science concept assumes that it will be possible, using 'middleware', to access someone else's dataset, either to replicate the research, or analyse it with a different agenda in mind.)

So the institutional repository is 'a system, hosted by the university or subject grouping, with a software platform whereby research debates or primary data can be stored and made freely available over the internet to anyone with an internet connection'. It is free at the point of use. There are no subscriptions – in the open access environment, subscriptions constitute a barrier to use. Individuals and indeed libraries 'cannot afford all the subscriptions they need to do their work', and that is bad for science and research.

The trouble with this 'brave new world' is that, although librarians are quite positive, it has proved difficult to get academics to deposit their material. UCL has managed to attract about 1,000 papers over three years. The picture for other partner institutions in the Sherpa project[2] 'varies wildly' but the flow of material is increasing. And, generally, awareness levels are rising. In a recent poll,[3] 81 per cent of academics said they would archive completely willingly, and 13 per cent 'reluctantly'. Only five per cent said they would not comply. In a talk to academic librarians hosted by academic book supplier Dawson in May, Paul described this as 'a good figure', which marked a shift in thinking. 'The tide may be changing.'

Some of the UK's research funding councils have announced that they will make research funding conditional on publication of the results in open access journals.[4] The Joint Information Systems Committee (Jisc) has meanwhile published a report on its experimental support of an author-pays model for OA publishing.[5] This follows on from a slightly inconclusive evaluation of the various OA business models conducted for Jisc in late 2004 by Rightscom,[6] and published in April 2005. The shift to a more mixed publishing regime for academic journals slowly gathers pace.

Preservation the biggest challenge

But the shift in many universities towards electronic provision still involves a big act of faith. Preservation and long-term curation pose significant technical problems. 'You need the material to be accessible and readable in 100 years' time.'

This is the biggest challenge in the digital publishing model. 'In the world of paper, libraries took on this role. Publishers have no financial gain in maintaining a paper archive. But academics still need access to the information. Students in chemistry, the sciences and maths need to go back 100 years. In the arts and humanities you need to go back centuries, to see the history of academic interpretation and endeavour.'

In the digital environment, who ultimately should be responsible is still unclear. Some work was done on behalf of Curl at the University of Leeds 10 years ago. Paul describes it, the so-called 'Cedars project', as 'exemplary', but it was not followed up.

> ... the BL and UCL have developed tools to help determine how much it would cost to store an e-journal in a secure e-archive for a period of up to 100 years...

Accurate costing of preservation

But the British Library and UCL stepped into the breach. One of the biggest obstacles to organising digital archiving on a grand scale was the absence of accurate information about long-term costs. In the Life project[7] (Lifecycle Information for e-Literature) the BL and UCL have developed tools to help determine how much it would cost to store an e-journal in a secure e-archive for a period of up to 100 years (or to archive a website for 10 years, or store an individual gif or tif file).

The generic matrix for costing preservation 'at the item level' is in the form of a spreadsheet, which enables institutions to cost technology watch, preservation frequency, and other preservation costs, as well as the cost of 'ingesting' the item, and adding metadata. In the past it was not possible to undertake this kind of exercise because no one had all the costs.

Between them, the two institutions produced indicative figures for such services as acquiring and making an e-journal accessible for UCL users (£3,250 a year). This included cataloguing, providing metadata and making it available over the network, as well as maintaining the results and teaching people how to use it.

The BL found that the cost of archiving a website was about £3,500 for each website. The project is now moving into a second phase.

Libraries do not expect any publisher to undertake 'true digital preservation'. Different countries have adopted different models for digital preservation. Paul thinks that the UK is likely to copy from the US, perhaps inspired by JStor,[8] the not-for-profit archive of important scholarly journals, which acts as a trusted digital archive for scanned e-content. Local e-archives, hosted by universities or by the British Library, are also likely to be a part of the brave new world.

As for the costs of running institutional repositories, these are 'quite low – from £10,000 to £15,000'. The software is open source, and UCL covered the staff costs by using existing staff from its services and built repository support into existing workflows. 'We re-designated existing money to new purposes.'

The existence of repositories is opening up new possibilities. It is now possible to evaluate alternative models of delivery in the scholarly information chain. Publishers rarely share detailed information on usage that fits the management needs of academic library services. In the past they have monitored levels of use and used this information to develop new subscription models, charging most commonly for volume of use (amount of downloading) for certain types of information, number of simultaneous users, or licensing access by size of user community. What they have not done is provide information on the impact on downloading of research results published in open access repositories as well as in a subscription-based journal.

In the highly charged atmosphere caused by the introduction of new open access publishing models, the academic community wanted alternative ways of evaluating quality. It may get them by looking at the levels of use of individual articles, thanks to a partnership between Ex Libris and the US's Los Alamos National Laboratory (LANL) where Johan Bollen and partner institution the State University of California used SFX services to track linking from bibliographic citations to particular articles.[9] These figures are in the public domain, as the aggregated user statistics are anonymous. The survey involved 67,000 individuals, 3.5m accesses and 2m articles.

The collated data produced some challenging results, most notably a significant divergence between the current ranking of top journals by impact factor (citation indices), and actual use of individual articles. 'Six out of 10 of the top 10 journals have significant differences if you use different types of matrix.' Such research covers only subscription journals, but European research libraries are hoping to replicate the North American experiment in a European context, and to include data from open access journals and repositories.

Impact on the RAE

These findings, which will be explored further at a forthcoming conference on European repositories,* are producing an exciting new tool which, however, 'introduces a new instability into the market'. That is because the mechanism by which universities in England compete for research funding is the Research Assessment Exercise. Academics wish to publish their research findings in high-impact academic journals and choose journals with high impact factors so as to maximise the number of readers for their research, and thereby the visibility of their home institution.

The universities and their Vice-Chancellors are watching with keen interest. They want effective measures of quality, that is, to obtain a true measure of impact by monitoring the use of articles in OA repositories and journals as well as subscription journals. But they are unlikely to change to a new method of determining research income (a recent proposal by the Chancellor of the Exchequer) until they have assessed 'exactly what the implications are'. As Paul says, 'It's exciting, but full of danger. Meanwhile research libraries across Europe are planning to collaborate in further research (see our news story on p.13). Scholarly communication is a global issue.

> ... the academic community wanted alternative ways of evaluating quality. It may get them by looking at the levels of use of individual articles...

Managing IPR – the foundation for open access

Paul is the university's Copyright Officer. A researcher himself, he has strong views. 'Research is intellectual property and researchers have rights.' At the moment there is no such thing as a UK approach to copyright in research output. Every university and institution has a different view of how it should be managed. UCL has spent four years compiling copyright policies throughout the institution. 'We have now developed separate policies for staff and students. They have the rights in their output, not the institution.'

Winning trust is key. 'Copyright management makes it easier to work in an open access environment. It is the author who is going to deposit their output. Copyright and IP rights management form one of the building blocks of the new intellectual environment.' UCL has started providing IPR classes to graduate students. They are always oversubscribed. 'We cannot hold enough classes to cover the demand. We not only teach them about their own rights but also third-party rights – how not to misuse other people's rights. They are responsible for how they use other people's IPR. Over-subscribed classes are a good sign – people are taking issues more seriously.'

The future

Is open access yesterday's debate? Probably not for libraries, and certainly not for publishers. But the needs of scholarly communication have outstripped the ability of publishers to service academic researchers appropriately and at the same time remain comfortably solvent. For a long time, publishers developed services that they thought users would want. And they do. The problem was affordability. 'The fact remains that the cost of subscriptions continued to outstrip the ability of libraries to pay. There's now a mismatch between the amount of money in a library budget, and the total cost of the journals package if you bought everything that academics want. The two are irreconcilable drivers. The gap is too wide. There is no more money in the public purse,' says Paul.

Because 'the real cost of teaching is not met by current resources', higher education institutions have had to look at disseminating in a different way, 'getting more value out of the investment you are putting in'. In the OA context that means adding value by providing more people with access.

Paul has been talking to publishers about OA for the last four or five years. Academic librarians have been talking to publishers about cost rises disproportionate to the size of their budgets for about the last 15. Repositories are comparatively new, but they are enabling academic institutions to make informed choices, as well as forcing publishers, sometimes reluctantly, to the negotiating table.

At the same time, the huge amount of information it is now possible to collect means that, for the first time, it will be possible to construct a comprehensive picture of user behaviour. What are users looking at? How do they get from A to B? Is it planned? Is it searching, or inter-acting? What are they looking at? How long do they spend on particular activities? The digital library is split up into journals, books, internet searches, and so on. If you put all this accumulated data into a matrix, you would build up a picture of how European students and researchers behave. 'This would be bigger than OA.' Although it would only track digital resources, not paper-based materials, it would still build a very helpful picture of user behaviour.

The debate, then, is not about business models for publishing academic journals. It is much bigger than that, something 'so vibrant that it changes every six months'. E-journals were 'the basis of an interesting experiment'. Now they are 'absolutely the way people want to receive information'. Quite how scholarly communication will evolve in future is uncertain. 'All revolutions are chaotic.' Paul Ayris is rising to the challenge. 'We enjoy reinventing the world. We hope we don't go on to destroy it in the process!'

References

1 *College Students' Perceptions of Libraries and Information Resources: a report to the OCLC membership.* OCLC, 2006. Especially pp. 1-7.

2 Sherpa is a partnership between research-led HE institutions which collaborate on the construction of institutional open access repositories (www.sherpa.ac.uk).

3 Alma Swan and Sheridan Brown. *Open Access Self-archiving: an author study*. May 2005. Especially Table 29 (www.keyperspectives.com/).

4 'Research bodies say yes.' *Library & Information Update*, July-August 2006, p.3.

5 'Jisc's open access programme provides boost to publishers and authors.' 16 May 2006 (www.jisc.ac.uk).

6 *Business Models for Journal Content. Final report submitted to Jisc by Rightscom* (www.jisc.ac.uk/coll_jiscejournals_jwgs.html).

7 *The Life Project: bringing digital preservation to life* (http://eprints.uc.ac.uk/archive/00001855).

8 www.jstor.org

9 Johan Bollen, Herbert Van de Sompel, *et al.* 'Toward alternative metrics of journal impact: a comparison of download and citation data', paper submitted to *Information Processing and Management*, special issue on Informetrics, 3 March 2005, and 'Alternative metrics of journal impact based on usage data: the BX project,' paper by Johan Bollen, Oren Beit-Arie and Herbert Van de Sompel at a meeting on alternative metrics of publication impact, 20-21 February 2006, Berlin, Humboldt University.

Dr Paul Ayris is Director of Library Services, UCL (University College London), and UCL Copyright Officer (p.ayris@ucl.ac.uk). You can read his papers and presentations on the topic of open access at http://eprints.ucl.ac.uk

** 'New challenges for open access repositories' takes place at Glasgow University, 18-20 October (www.lib.gla.ac.uk/openscholarship/programme.html). Dr Ayris will be chairing sessions, and other notable participants include Stephen Pinfold, Bill Hubbard and David Prosser.*