

The Open Archives Initiative: application and exploitation

A one-day seminar focusing on the application and exploitation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)

Wednesday 14th May 2003



Around 40 delegates representing the major institutions and publishing companies interested in the Open Archive Initiative attended this seminar. Attendance was lower than expected probably due to the 'expected' technical nature of the seminar. However, instead of addressing technical issues relating to the OAI-PMH, the seminar was aimed more for those interested in how the OAI protocol has been applied within specific projects. It provided an insight into the present tensions in scholarly communication as seen from the view of publishers and academic institutions and presented their reactions to the use of OAI as a workable solution to this problem. The following notes presented here are the main points from the talks most relevant to the Theses Alive! and SHERPA projects.

1) Scholarly communication and OAI: what are the issues for journal publishers?

Kathryn Toledano, Director: Business Development, Emerald

The scholarly communication system is currently not perfect, however new technology (OAI-compliant) currently available can address this.

OAI promotes:

- The standardisation of metadata
- The provision of metadata (harvesting/licenced/sub-licensed)
- The use of open URL's
- An easier process for authors
- Rights (e.g. the Zwolle principles argument has intensified/project RoMEO)

With the advent of Institutional Repositories and the removal of access barriers, change is inevitable. Likely changes are:

- The roles of the publisher and library
- Quality of the official record
- Erosion of journal brands and investment
- How excellence is measured
- Demise of the peer review system

The drivers for change are many and varied:

- Rising journal costs and restrictive licences
- Barriers to dissemination
- Advances in digital technology
- Copyright restrictions to authors

So, in conclusion OAI challenges publishers by ensuring change, therefore to survive publishers need to collaborate.

2) SHERPA: institutional repositories

Bill Hubbard, SHERPA project manager, University of Nottingham

SHERPA (<http://www.sherpa.ac.uk/>) aims to provide a hybrid environment via various partners including research universities, colleges, British Library, AHDS. Initial key points regarding SHERPA:

- Institutional rather than subject-based
- A repository rather than an archive
- Deals with e-prints rather than e-resources
- All linked through the OAI-PMH

E-prints definition;

“e-versions of research papers include pre-prints (+/- corrections), post-prints, other material (such as conference papers, reports, book chapters)”

The key is quality control, particularly peer review. Benefits for the researcher include:

- Wide dissemination (more visible-cited more)
- Rapid dissemination
- Ease of access
- Cross-searchable
- Value-added services (hit counts, personalised publishing lists, citation analyses)

But, why institutional based?

- Institutions have centralised resources to help subsidise repository start-up, to support the tech/organisational infrastructure, to deal with preservation issues over the long term (esp. electronically).
- An interest in disseminating content
- Work across subject boundaries to share solutions
- Natural part of an institutions function

Benefits for the institution:

- Raises profile and prestige
- Helps manage institutional information assets
- Accreditation/performance management, e.g. RAE
- Long-term savings
- Encourages institutional sense of identity

These points all need testing in practice, i.e. the need for SHERPA exists.

Issues that SHERPA will encounter;

1. Collection policy
2. Preservation
3. IPR
4. Cultural differences and changes (both publishers and academics)

It is worth noting that these are not technical difficulties and the technology already exists.

1. Collection policy

- Document types (pre-print v post-print)
- Document format (html, PDF etc)
- Submission procedures (mediated/DIY)
- Metadata quality standards (self created)

Who can submit e-prints to SHERPA? Staff members only or graduate students aswell? What happens when staff members leave? Do we take out their papers from the repository?

2. Preservation

- Selection and retention criteria. How long do we keep them online?
- Preservation of metadata. Got to think long term.
- Preferred formats
- Life cycle management
- Costing models

Another view is that these preservation issues should be overlooked at the minute and dealt with retrospectively. The benefit of this viewpoint is that the project will start a lot quicker.

3. IPR

- Author permission and licensing terms
- Copyright and copying (plagiarism?)
- Compliance with publishers copyright terms

4. Cultural differences and changes

- Different disciplines have differing opinions on pre-prints (e.g. Physics and medicine). Suggestion that we work with existing quality control models.
- Changing the status quo (advocacy and support)
- Brand values not threatened by Inst. Repositories (e.g. Nature)

3) The DAEDALUS project: a case study for the creation of institutional repositories

Susan Ashworth, Senior Assistant Librarian, Glasgow University Library

The scholarly communications debate has arisen out of excessive price inflation by journal titles and has been strongly taken up by CURL. The FAIR programme is a response to this issue in the UK, and is part of a larger worldwide movement towards OA.

Local issues include;

- Budgeting constraints
- Devolved budgeting (to academics)
- Serials v monographs
- Other strategies (e.g. collaboration with other institutions)

Will institution repositories resolve these problems?

Institutional repositories in the long term will maximise access and reduce the annual serials expenditure. Short-term gains will involve gaining maximum visibility and impact for research.

DAEDALUS (<http://www.lib.gla.ac.uk/daedalus/>) is part of the FAIR programme and is a partner to SHERPA. It is a core strategic aim for Glasgow University Library and comprises two strands; advocacy and systems development. The DAEDALUS project has encountered three main issues to be overcome; 1) Technical (e.g. standards and formats), 2) Cultural and 3) Organisational (e.g. IPR, plagiarism).

The DAEDALUS project uses e-prints software for;

- Published papers
- Grey matter
- Theses

However, Dspace software is being explored because of digital preservation aspects, and the devolution into communities that Dspace offers.

For e-theses, Virginia Tech software is used. Submission is generally voluntary but with strong faculty support. So, the major issues seem not to be technical but cultural. These can only be overcome by advocacy work. The remit for this is to create an open access culture via;

- Gathering content
- Give advice on policy implications
- Formulate an exit strategy that ensures a fully used service.

Strategies for advocacy

- The formation of a project board, including academics from each of the three subject group areas (including sceptics)
- Contacting self-publishing academics (some good success achieved)
- Exploring taking over departmental publishing databases with quid pro quo full text access where possible
- Raising debate within departments by attending a programme of departmental meetings
- Campus events aimed at academics (especially IPR issues)
- Attempting to get strategy decisions at a high level within the university (especially on submission of theses)
- Providing assistance for staff, including focusing on publisher copyright policy, mediated submission service (file conversion, record enhancement)
- Subject librarian input (get on agendas of faculty committees, arrange for project managers to attend meetings, encourage feedback, promote issues, distribute publicity material).

Academic concerns raised by DAEDALUS

- The integrity of their work will be compromised (but is this a real issue?)
- Prior publication
- Break existing copyright agreements
- Where is the benefit?
- Institutional liability
- Collection policy and quality control (especially of grey area material and pre-prints). Generally the academics don't want library control over this aspect

4) Project RoMEO: addressing the IPR issues relating to self-archiving of e-prints under the Open Archives Initiative

Elizabeth Gadd, Research Associate, Loughborough University

The aims of this project were to investigate the attitudes of the four main stakeholders by surveying academic, journal publishers and service/data providers and the journal copyright transfer agreements. The main findings from this project are briefly summarised thus;

Who owns copyright?

Technically the HEI, although 61% of academics thought that they owned the IPR. Multiple copyright owners complicate this issue (also the question of where to archive it?). Third party ownership is established in 25% of the cases.

How will the work be protected under law and under license?

Under normal law the work is strictly protected (i.e. one copy per article for own use only with no distribution rights). Under license it is even more strictly protected. Academics want their work more accessible IF they are acknowledged and the copy is exact. However, academics are generally not aware of alternative copyright regimes.

The prior publication issue.

Academics think that publication constitutes appearing in a formal journal title only, with self-archiving classified as communication. However, publishers have a much broader definition in that publication constitutes making the information available to the public by any means. Perhaps a resolution to this problem is to distinguish between self-archiving and appearance in journals by using the terms *first publication* and *definitive publication* for the respective appearances.

Copyright issues

90% of publishers ask for a copyright assignment. 15% of cases don't revert copyright back to the author even if the manuscript is rejected. However, if the employer owns the copyright then they probably get a better deal. Where the government owns the copyright then an even better deal is ensured as such works are in the public domain. 57.5% of publishers have a US government clause. This suggests that publishers can work with a parallel publication model. Most academics assign copyright reluctantly (49%) but they do not get much in return. In fact, 28.5% impart no rights to their own work at all.

Will publishers allow self-archiving and under what conditions?

Only half of the publishers allow self-archiving and then there seems to be no consensus or standard condition under which this can occur. This creates lots of difficulties for the authors. However, some authors would be prepared to self-archive regardless of an agreement. Publishers that do allow self-archiving are not keen on the self-archiving of PDF's that they provide.

Even if there are significant differences between pre-prints and post-prints it is unlikely that a separate copyright agreement exists for them, so in effect they are covered by the same agreement.

Metadata rights

Each metadata record is subject to copyright, even though it is essentially just raw information. The data providers and depositors own this copyright even though the data providers do not think that metadata constitutes copyright. Even so the data providers still want metadata protected.

Most service providers do not check if they can harvest the metadata. Most thought this data was implicitly free. Service providers enhance the metadata they receive and wish to protect this data.

Some conclusions

There needs to be a redistribution of rights including;

- Institutional repositories may freely publish employees work
- Authors can publish where they like, can reuse their work and would like to specify how others use their work.
- Publishers need the right to definitive print and funding.

For more definitive conclusions see the RoMEO URL;

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>

Theo Andrew **Theses Alive! Project**
15 May 2003