The JISC's FAIR Programme: disclosing and sharing institutional assets

Chris Awre
Joint Information Systems Committee

© Chris Awre 2004

ABSTRACT: An account is given of the work of the FAIR programme in developing and supporting projects in the sharing of digitally stored institutional assets within the higher and further education sectors in the United Kingdom.

Introduction

Higher (HE) and further (FE) education institutions all produce a range of outputs from their activities, and nowadays many of these are digital. These outputs mainly originate from core activities, such as materials used for learning and teaching, or a variety of research outputs, but they may also cover outputs from more generic institutional processes. Many of these outputs are not known about outside the institution where they were produced. While this may be perfectly reasonable, the Joint Information Systems Committee (JISC) highlighted the need to investigate how technology might assist institutions in depositing, disclosing and sharing these digital assets so that they might be more widely known and more widely used across the HE and FE communities. It would also be possible to investigate how the disclosure of such assets might be enabled through channels additional to those in use for non-digital assets; in particular, how disclosure might take place through the sharing of metadata about the assets themselves. Such sharing of information would, it was hoped, facilitate co-operation between individuals and institutions and assist with their learning, teaching and research activities.

The FAIR (Focus on Access to Institutional Resources) Programme[1] was launched in August 2002 to investigate these issues. The Programme will run until October 2005 and is a collection of 14 projects which, in the words of the original call for proposals,

. . . involve members of the HE/FE community in projects to deposit and disclose institutional assets and to gather intelligence about and increase our understanding of the technical, organisational and cultural challenges of these processes. The FAIR Programme will also contribute to developing the mechanisms and supporting services to allow the submission

Chris Awre

and sharing of content generated by the HE/FE community.

In particular, the FAIR Programme is 'inspired by the vision of the Open Archives Initiative (OAI)'.[2] This standard allows the sharing of digital assets based on a simple mechanism allowing metadata about those assets to be harvested into services. Most of the projects are thus investigating the use of it for disclosure and sharing.

The projects themselves are in three main areas, reflected in the clusters of projects that have been formed to facilitate the exchange of experiences. The clusters are:

- *E-prints and e-theses.* Eight projects testing the deposit and disclosure, using OAI, of e-prints and e-theses, and digital outputs from institutional research and from teaching and learning.
- *Museums and images.* Four projects examining the use of OAI in the sharing of museum objects and images and how OAI disclosure might be combined with digital preservation.
- *Institutional portals.* Two projects investigating the presentation of external information sources alongside disclosed assets within an institution.

Details of the individual projects and their areas of work are given in the Appendix.

This article reports on key issues that have arisen within the clusters and the FAIR Programme so far in the 18 months since it started. Work during this period has identified issues to be addressed to enable the full sharing of assets across the community and to decide if OAI can be realized as a mechanism to enable that. These are:

- Culture
- IPR
- Metadata
- Service provision
- Software choice

Culture

The FAIR call recognized the need to explore the cultural issues surrounding the disclosure of institutional assets as well as the technology. All the projects have found that this aspect has been and will be one of the largest parts of their work.

issues to be addressed to enable the full sharing of assets across the community

E-prints and e-theses

Disclosing digital institutional assets using OAI is facilitated by depositing them in a common place. Repositories are being considered and used by institutions as the means for this deposit. These repositories can be used to store many different material types, but two that are of key interest within FAIR are e-prints and e-theses. A repository for the latter has been identified as a means for making the content of theses and dissertations more widely known, as much research is otherwise locked away in paper copies. E-prints, commonly regarded as electronic versions of research articles, are also generating much discussion. Institutionally based e-print repositories (IBERs), which are being developed within FAIR projects, offer the following advantages:

- They allow research material to be disseminated rapidly through immediate availability once deposited on an open access basis. The use of OAI allows wide and efficient search and retrieval of repositories across the world.
- They allow the dissemination of both pre-prints and post-prints, with the peer-review status of the article clearly tagged. This allows the research path to be identified and quality to be ascertained.
- They provide a companion form of publication alongside traditional or open access journals. Note that these routes of publication are *not* mutually exclusive, but can play different roles in ensuring the widest possible dissemination of the work.
- They allow institutions to be clearly identified as the origin and location of the research. They also allow the institution to better manage its digital assets, including the maintenance and management of copyright inherent in these.
- They allow long-term digital preservation handling procedures to be built in at the point of deposit.

When collecting pre-prints, the main issue found by the projects is the question of priority for the various outputs. Partly because of the requirements of the Research Assessment Exercise (RAE) there is often pressure for publication in quality peer-reviewed journals only. Deposition, or self-archiving,

of pre-prints in an institutional repository as well is not yet seen as a desired or accepted activity. But pre-prints are an established entity in communities that traditionally make use of subject-based repositories. These repositories have been in existence for some time and they occur in a number of subject areas, notably physics (the ArXiv[3] repository). FAIR projects are examining possible conflicts or synergies that might arise between institutional and subject repositories. Indeed, the Theses Alive! project has found a direct correlation between staff members who self-archive to their own personal websites and those who deposit in subject repositories. The DAEDALUS project is making a connection of its own by developing a local subject repository based on the contents of the institutional repository.

Projects have found greater enthusiasm for post-prints. Here, many academics are happy to offer their papers for inclusion in the repository. There is a general lack of awareness of whether they are allowed to do this, though, and 30% of publisher policies surveyed by the RoMEO project explicitly ban such self-archiving. The survey, however, found that many publishers are far more generous and the directory of publisher self-archiving policies compiled by RoMEO is currently being maintained by the SHERPA project to ensure both publishers and academics can be fully aware of what is possible. Licence arrangements for academics depositing in repositories are being developed by projects and will act as exemplars.

2003 saw a wider awareness of both institutional repositories and the open access publishing movement in general, and both have contributed to FAIR projects. Many senior institutional managers now recognize the valuable possibilities of an institutional repository. In the projects' institutions this has resulted in welcome top-level support. Developments in the open access arena, and in particular the JISC/BioMed Central agreement,[4] have also raised awareness of open access journals. This has complemented the work of projects in collecting e-prints from academics who have or who may publish in such journals. Indeed, DAEDALUS is assisting with the creation of such a journal at the University of Glasgow.

Institutional portals

Within HE, much valuable evidence has been gathered on how users would like to see external resources presented alongside internal resources within an institutional portal. The PORTAL project has produced a number of very well received reports in this area. The report on Stakeholder Requirements for National Content in Institutional Portals[5] has made it clear that access, preferably of a personalized nature, to external resources, commercial and public, is highly desirable, though mechanisms are required to ease the involvement of the providers of such content.

IPR

E-prints and e-theses

Inherent in producing an e-print is the generation of associated rights. This is especially apparent when academics are asked to assign copyright or agree to licence their work to the publisher of a journal the work may appear in. It is clear from experience within projects so far that the IPR issues involved in using an institutional repository are not yet fully known and that greater awareness is urgently required if repositories are to be used to their greatest extent.

The RoMEO project has found that academics-as-authors are quite happy for their papers to be used outside the published journal under certain, limited, conditions, and indeed would welcome it. To achieve this, the ownership of rights needs to be confirmed so that the relationship between institutions and academics can be clarified and so that both parties are aware and happy with what can be done with publications. The relationship between academics and publishers also needs to be clarified. It is arguable that academics do not have the right to assign copyright on their works as their employer may be the legal owner. Further debate on this is required as well as encouraging academics to be more aware of what is being assigned to whom.

The RoMEO project has also dealt with the IPR of e-prints once they are disclosed using OAI. This has resulted in collaboration with the OAI developers in the USA and the establishment of an OAI-Rights

many academics are happy to offer their papers for inclusion in the repository

Technical Group[6] to incorporate much of the RoMEO work within the OAI standard. There has also been collaboration with the Creative Commons.[7]

Museums and Images

Image copyright is a major issue being tackled by the BioMed Image Archive, particularly due to the medical nature of the images. While the project's work to develop a self-archiving system for the community to use in disclosing medical images is progressing well, the issue of copyright and patient permission will have a major effect on what can be deposited and disclosed.

Institutional portals

Within FE, the FAIR Enough project has also discovered gaps in awareness of who owns the copyright on learning resources produced by staff. Addressing this has been required in order to enable sharing of these resources to take place.

Metadata

E-prints and e-theses

The use of OAI currently mandates a minimum use of Dublin Core metadata[8] to describe the e-prints being deposited in the repository. Dublin Core originated in the bibliographic world, and it suits the needs of e-prints very well. However, establishing a metadata structure is one thing; establishing good practice for how the fields within this structure are filled is another. The ePrints UK project has produced a set of guidelines on this and all FAIR e-print projects are making use of these; a formal Dublin Core Metadata Initiative paper is planned in order to encourage their wider adoption.

   Beyond providing guidelines, the quality of the metadata entered is important to ensure that records can be searched and accessed correctly and efficiently. Poor metadata means e-prints will not be retrieved in a search, negating one of the main purposes for depositing the e-print in a repository – that of increasing awareness of the materials. This issue has been addressed in a paper to the 2003 Dublin Core Conference by members of the TARDis and HaIRST projects,[9] and in an article by the ePrints UK team in *Ariadne*.[10]

the richness of metadata is an issue that is also affecting museum metadata

In collecting metadata when e-prints are deposited in a repository, there is also the issue of what this metadata is needed for. Its primary use is for retrieval through searching. However, the TARDis project has found that academics entering metadata would like to use it for other purposes, such as reading lists or bibliographies, which may require different fields of information. Adaptations in the deposit procedure have resulted from this.

   Parallel to this, the HaIRST project has found that Dublin Core for e-prints is fine, but if you wish to provide a search across different material types, e.g. e-prints and learning materials, Dublin Core on its own is limited. Further investigations are underway to see whether a lowest common denominator such as Dublin Core will be the pragmatic solution, or whether a richer metadata structure could or needs to be used.

Museums and images

The richness of metadata is an issue that is also affecting museum metadata. The Accessing the Virtual Museum project has found that cataloguing Egyptian artefacts using Dublin Core is limited (e.g. who is the author?). Accordingly, they have been keen to establish subject access as a key route to the metadata. The benefits of rich subject classification are clear for those who are aware of the terminology involved. The Harvesting the FitzWilliam project has been involved in the development of a demonstrator service with the Archaeology Data Service (ADS),[11] involving the harvesting of metadata from the collection of coins. While these were very rich in their metadata, the metadata was very much aimed at numismatists. While the technical aspect of harvesting and serving the metadata worked well, the ADS found that its richness was less useful when presenting the data to archaeologists, who wished to search on very different aspects.

Service provision

E-prints and e-theses

Most of the projects developing e-print and e-theses repositories are concentrating on being OAI data providers, relying on others to harvest the metadata and serve it to the

wider community. They will be making local services available, but these will only offer access to a single institution's assets. The ePrints UK project is building an OAI service provider to fulfil the role of serving cross-institution access, and there are a number of international service providers (e.g. OAIster[12]) and subject-based services (e.g. CogPrints,[13] ArXiv) that can already provide this service. HaIRST is also examining the requirements of a service provider for cross-material type access, and DAEDALUS is examining service provision across different repositories within an institution.

Software choice

e-prints and e-theses

Fortunately, the technical side of the FAIR projects dealing with institutional repositories has not proved to be overly complicated, with all institutions involved getting software and repositories up and running relatively smoothly. The first 18 months of FAIR has, however, seen options of which repository software to choose increase. When FAIR started the established software was e-prints.org.[14] This has been adopted in most projects. DSpace[15] from MIT is now being seen as a viable option to e-prints.org and the two are being tested alongside each other in the DAEDALUS project; an article outlining initial experiences was published in *Ariadne* in October 2003.[16] There are other repository software packages now available as well, but these have not been considered in the UK as yet. A *Guide to Institutional Repository Software*,[17] providing an overview comparison, has recently been published by the Open Society Institute and will be updated as developments take place.

Conclusion

The first 18 months of the FAIR Programme has been a period where many issues have emerged. This process has been encouraged so that they can be fully addressed within their projects and their institutions. The cross-project themes identified so far are described here.

Overall, the projects within FAIR are proving very successful in addressing the aims of the programme and will produce valuable information and experience for institutions and others to learn from, both in the use of OAI and in the wider issues of disclosing and sharing institutional assets. All projects have widely disseminated their aims and findings through workshops, conferences and journal articles. The timing of the FAIR Programme will both allow projects to benefit from wider interest and developments and to feed into and influence these developments.

Appendix: the FAIR projects

The summary provided here gives background to the FAIR Programme for those unfamiliar with the projects it includes. Further information and background can be found on the JISC website at http://www.jisc.ac.uk/programme_fair.html and at the individual project links given below.

E-prints and e-theses (the EFAIR cluster)
DAEDALUS – University of Glasgow
http://www.lib.gla.ac.uk/daedalus/

Investigating the use of different repository software to store different types of materials within the University, from which they can then be disclosed to the wider community. These materials include pre-prints, post-prints (deliberately kept separate), electronic theses and dissertations (ETDs), administrative documents and grey literature.

HaIRST – University of Strathclyde
http://hairst.cdlr.strath.ac.uk

HaIRST involves a consortium of both HE and FE institutions, and is investigating the ways in which different materials can be disclosed and shared in different situations. The project is developing a model of how different levels of metadata can be incorporated in an overall search for materials across the range.

SHERPA – University of Nottingham
http://www.sherpa.ac.uk

SHERPA concentrates on the building up of a body of content of self-archived articles in order to increase the range of research outputs available on open access within institutional repositories. It includes both pre-prints and post-prints.

TARDis – University of Southampton
http://tardis.eprints.org

This project concentrates on the technical and cultural aspects of setting up an institutional repository and making this an integral part of the research and publication process within the university. It is also exploring the role of the library in providing such a service.

ePrints UK – UKOLN, University of Bath
http://www.rdn.ac.uk/projects/eprints-uk/

Within the context of OAI, the four projects above are developing data providers, repositories of content and associated metadata. ePrints UK is building an OAI service provider. Although individual repositories will be accessible at institutions, ePrints UK is envisaged as the route to search across all or a selection of them.

RoMEO – University of Loughborough
http://www.lboro.ac.uk/departments/ls/disresearch/romeo/
index.html

This one-year project is now complete. Surveys of authors, journal publishers, OAI data providers and OAI service providers were undertaken and built up a body of data on the rights these various parties are willing to allow and/or would like to see. A directory of publisher self-archive policies has been created. Solutions to allow all e-prints to include an IPR statement in associated metadata have also been created.

Theses Alive! – University of Edinburgh
http://www.thesesalive.ac.uk

Developing a pilot distributed system for the management of e-theses. This will involve the development of an OAI-compliant thesis submission system and the testing of an infrastructure to enable e-theses to be published on the Web by institutions.

Electronic Theses – Robert Gordon University, Aberdeen
http://www.rgu.ac.uk/library/e-theses.htm

To complement the work of the DAEDALUS (storage) and Theses Alive! (submission and dissemination) projects, this project is examining the practice and methods of e-theses production, management and use, including theses that have been digitized and those born digital.

Museums and images

Harvesting the FitzWilliam – FitzWilliam Museum, University of Cambridge
http://www.fitzmuseum.cam.ac.uk/htf/index.html

This project is preparing and digitizing items from the FitzWilliam collections for disclosure using OAI and delivery through the Archaeology Data Service and Arts & Humanities Data Service. Issues being addressed include the metadata requirements for museum objects and the disclosure of images using OAI.

Accessing the Virtual Museum – Petrie Museum, University College London
http://www.petrie.ucl.ac.uk/randd/avm.html

This project is preparing and digitizing items from the Petrie Museum of Egyptian Archaeology. Disclosure will be through the Archaeology Data Service. The major issue being addressed is the creation of metadata for objects which require transliteration from original scripts and which do not naturally fit with the Dublin Core requirements of OAI.

Hybrid Archives – Arts & Humanities Data Service (AHDS), King's College London
http://www.ahds.ac.uk/about/projects/hybrid-archives/index.htm

Hybrid Archives is examining the preservation requirements for institutional collections disclosed for delivery through a service such as the AHDS. The project is developing a model of partial deposit, whereby a preservation copy of the actual content is deposited at the AHDS, while the OAI-disclosed metadata still connects users to the original collections.

BioMed Image Archive – Institute for Learning & Research Technology, University of Bristol
http://www.brisbio.ac.uk/

The BioMed Image Archive itself is an established source of biomedical images that has been available for a number of years. This project is investigating ways that would allow institutions to self-archive images to the archive remotely, in order to share these with the wider community.

Institutional portals

PORTAL – University of Hull
http://www.fair-portal.hull.ac.uk

PORTAL is examining how JISC collections and services can be embedded within an institutional portal. It is taking advantage of Hull's current work in developing an institutional portal and the project will be assessing user needs for such a portal and what information should be delivered through it.

FAIR Enough – Western Colleges Consortium, Keynsham, Somerset
http://www.fairenough.ac.uk

The Western Colleges Consortium is a collection of five further education colleges in the West Country. This project is examining the needs of further education for access to JISC collections and services within a virtual learning environment, and how these sit alongside local assets.

References

1. FAIR Programme, http://www.jisc.ac.uk/programme_fair.html
2. Open Archives Initiative, http://www.openarchives.org
3. ArXiv, http://www.arxiv.org/
4. The JISC/Biomed Central agreement, http://www.jisc.ac.uk/index.cfm?name=coll_biomedcentral&src=alpha
5. PORTAL reports, http://www.fair-portal.hull.ac.uk/deliverables.html
6. OAI-Rights Technical Group, http://www.open-archives.org/documents/OAIRightsWhitePaper.html
8. Dublin Core Metadata Initiative, http://dublincore.org/
9. Barton, J., Currier, S. and Hey, J.M.N. Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. http://www.siderean.com/dc2003/event.jsp?event=http://www.siderean.com/dc2003/Paper60-abstract.pdf
10. Guy, M., Powell, A. and Day, M. Improving the quality of metadata in eprint archives. http://www.ariadne.ac.uk/issue38/guy/
11. Archaeology Data Service, http://ads.ahds.ac.uk/
12. OAIster, http://www.oaister.org/o/oaister/
13. CogPrints, http://cogprints.ecs.soton.ac.uk/
14. eprints.org, http://software.eprints.org
15. DSpace, http://www.dspace.org/
16. Nixon, W. DAEDALUS: initial experiences with EPrints and DSpace at the University of Glasgow. http://www.ariadne.ac.uk/issue37/nixon/
17. Guide to Institutional Repository Software, http://www.soros.org/openaccess/software/

Chris Awre
Programme Manager, JISC
3rd Floor, Bridge House, 138–142 The Strand
London WC2R 1HH, UK
Email: c.awre@jisc.ac.uk