



Open Scholarship 2006

Abstracts

Wednesday 18th October

Tutorials and Workshops

Functionality and Software Issues

William Nixon (University of Glasgow, UK)

Institutional repository software continues to develop both in open source and through commercial vendors. This tutorial will not advocate any specific software for an IR but will provide attendees with an opportunity to consider and review their repository software choices, issues and options.

It will provide a brief overview of software choices made at the University of Glasgow and the generic lessons learned.

As repository functionality continues to develop, new features new issues and new challenges arise. Many which may not have been originally considered when the repository was established. This session will provide an opportunity for participants to share their experiences and to identify areas where functionality could be added or enhanced.

This tutorial will be of interest to repository and library staff involved in the administration, workflow and population of their repositories.

EPrints Application Profile

Julie Allinson (UKOLN, University of Bath, UK) and Andy Powell (Eduserv Foundation, UK)

This tutorial, run by Julie Allinson (UKOLN, University of Bath) and Andy Powell (Eduserv Foundation) offers an introduction to the Eprints Application Profile, a Dublin Core metadata profile developed to facilitate the creation and sharing of a richer set of metadata for describing scholarly works (eprints) than currently offered by simple Dublin Core. The Profile makes use of the Dublin Core Abstract Model and FRBR to group together descriptions of different expressions of a work. The application profile has been developed with funding from JISC and is hoped to form the basis for the UK eprint search engine currently being developed by Intute. The tutorial will be of use to anybody with an interest in metadata, particularly the metadata used to describe scholarly works in institutional, and other, repositories. Those for whom this tutorial may be of special interest include repository managers or administrators; developers or implementers of repository software; librarians; and developers of aggregator services. Those interested in implementing the Eprints Application Profile in their own repository are particularly encouraged to attend.

Advocacy and Policy Issues: Beginners Session

Morag Greig (University of Glasgow, UK)

With open access and repositories assuming a high profile some may question whether advocacy is still necessary. Those involved in the business of setting up and populating repositories are aware that in the majority of institutions there is still a great need for advocacy. This tutorial will give participants an opportunity to discuss different advocacy methods and approaches, including the 'top down' and 'bottom up' approach, publicity methods and the opportunities offered by funding body positions on open access.

Participants will have the opportunity to share experiences of what works and what doesn't.

The advocacy role often encompasses responsibility for advising academics on IPR issues. This is a particularly critical area where repository staff are engaged in depositing content on behalf of academics. The tutorial will offer an opportunity to discuss the IPR issues encountered by those managing repositories.

The tutorial will draw on the experience of participants who have been engaged in advocacy activities for some time. The tutorial should be particularly useful for those new to the area, but it will also present an opportunity for sharing ideas.

Advocacy and Policy Issues: Advanced Session

Joanne Yeomans (CERN, Switzerland)

If setting up a repository is the first goal, increasing the content is the second. Whether you have a mandate for deposit or not, there is still a lot of work to be done in getting papers into the database and persuading people that it is not just a dark store but a living collection. The nature of this work is varied and challenging - building relationships, finding technical solutions, negotiating, innovating, managing, juggling the balls, and keeping up everyone's enthusiasm, not least your own. The workshop will be a chance for participants to share experiences of both success and failure and also to start to work with some new ideas for future workplans.

The session should build on the ideas formed in the beginners session, and also be useful for those who have been working with a repository for a few years and who wish to take a step back and a fresh look at what they are doing.

Thursday 19th October

Session 2: Overview of Repository Development

Repository development: where are we now and where are we going?

Stephen Pinfield (The University Nottingham, UK)

Where have we come from? Where are we now? Where are we going? This presentation will attempt to answer these questions in relation to open scholarship in

general and digital repositories in particular. It will try to make a realistic assessment of the achievements so far and the challenges ahead. Although a great deal has been achieved, major developments are still required in technical, economic, cultural, and policy areas in order to ensure that repositories become embedded in the scholarly communication process. A number of difficult issues relating to repositories are yet to be resolved. These include questions such as version identification, business models and how repositories relate to journal publications (open access or otherwise). Work is ongoing to address these issues but further action is still required by institutions, subject communities, funding agencies and governments. This presentation will outline attempts in various countries to lead and coordinate key developments. It will argue that the signs are promising but advocates of open-access repositories should perhaps bear in mind that the significance of changes tend often to be overestimated in the short-term and underestimated in the long-term.

OpenDOAR

Bill Hubbard (SHERPA, The University Nottingham, UK)

Institutions and subject bodies across the world are moving swiftly to take advantage of open access repositories for a variety of purposes. OpenDOAR maintains a list of live open access repositories and analyses their structure and content. This presentation will review OpenDOAR, the development of the global repository network, the response from Europe and the current state of repositories. It will look at the way in which service providers can take advantage of repositories for innovative services and the need for community building and sharing of best practice in establishing and developing archives.

Session 3: Added-value services

Cross-repository interoperability

Simeon Warner (Cornell University, USA)

Open-access repositories could be the foundation of a global and natively-digital scholarly communication system, yet the current infrastructure falls far short of this goal. One missing element is the interoperability necessary to support the many workflows and value-chains involved in scholarly communication.

Since it is unrealistic to imagine the domination of a single repository architecture, content representation, or set of management policies, one must therefore create an interoperability framework to connect the many heterogeneous systems that will exist. What is the minimum set of features necessary? Can we find a common core content model to support interoperability?

I will describe ideas and experience from the NSF Pathways project between Cornell University and the Los Alamos National Laboratory. In particular, I will use as an example our prototype implementation of an overlay journal over a set heterogeneous repository architectures (aDORE, arXiv, DSpace, Fedora).

RepoMMan: the repository as service-oriented institutional infrastructure

Richard Green and Chris Awre (University of Hull, UK)

The RepoMMan project has been funded as part of the Joint Information Systems Committee's Digital Repositories Programme, and is running until May 2007. The project is closely aligned with the deployment of the Fedora digital repository system within the University of Hull, and seeks to address two areas of functionality that have generated wide interest in recent months: workflow and automated metadata generation. This presentation will describe the work of the RepoMMan project to date, and place this development in the context of wider infrastructural developments within the University that aim to establish the repository as an underpinning part of work and study.

The Fedora digital repository system is a framework upon which many repository services can be built. This flexibility requires a degree of development and configuration to meet the particular requirements at hand, but also provides a powerful basis for enabling technology to meet needs rather than needs adapting to meet technology. In considering how a repository will be used there are likely to be a number of steps involved. Each task will have its own series of steps, and these are followed to achieve the desired aim. Repository software can build in workflow capability that allows these steps to be followed seamlessly and simply, but this is often quite specific to the areas of functionality provided by the software. The RepoMMan project is developing a standards-based, flexible workflow tool through which users can interact with Fedora in a configurable way. The tool is based on a combination of two standards approaches: it is making use of Fedora's Web Service API interface and orchestrating a series of Web Service calls to this interface using BPEL (Business Process Execution Language). Access to the workflow tool is being examined through an institutional portal (based on uPortal) and the Sakai Collaboration & Learning Environment.

The second area of functionality being investigated by RepoMMan is automated metadata generation. The creation of metadata is essential for the proper management and use of a repository, though it is also clearly recognised that it is not viable for this to be entirely human-generated. Many different types of metadata can be sourced automatically. For example, technical metadata about digital objects can be gathered through tools such as JHOVE, whilst administrative metadata can be gathered through institutional profiles: by logging into the institutional portal, the portal already knows who you are and can pass this metadata onto digital objects ingested through the portal into the repository. However, automated generation of descriptive metadata is still a holy grail for the most part. RepoMMan is investigating the different approaches that can be taken to enable automated descriptive metadata generation and will be testing possible alternatives to gather further information on requirements and practice.

Whilst the two areas of functionality being investigated by RepoMMan are of wide interest technically, it is regarded as vital that they are meeting identified user requirements. A survey and a set of interviews with researchers have provided evidence of how research is carried out and how research documentation and information is managed, from research initiation through to publication. The

development of the workflow tool is centred on how the repository can be provided as a working tool on a day-to-day basis throughout the research process, and not simply a store for documents at the end of this. Similarly, metadata generation is best carried out as close to the point of document creation as possible, and automated techniques can enable this at different stages of research.

In designing the repository and the tools around it to fit into everyday practice it is not our intention to force users to make use of these systems. Rather the developments are part of an ongoing aim within the University of Hull to provide infrastructural components that can be used in a flexible fashion to enhance and facilitate business processes. The institutional portal and web content management system are other similar components and a study of the Sakai Collaboration & Learning Environment has suggested that it too offers flexible capability to support work and study within the institution.

The development of Fedora as the University's repository offers another component that can be used alongside and in tandem with the others. All of these can be described as service-oriented in their approach, insofar that they are designed to allow services to be built with them, rather than they dictate the services that will be presented (e.g., workflow and metadata generation on top of Fedora). The service-oriented architectural approach is in many cases still an approach, and it is accepted that many systems do not yet fully adhere to this model. The potential such architecture offers, however, has encouraged ongoing adoption. Fedora provides a service-oriented ability to work with digital content in a coherent and structured manner to support all aspects of storage, management, access, and preservation.

Bielefeld Academic Search Engine : a scientific search service for Institutional Repositories

Friedrich Summann (Bielefeld University Library, Germany)

Bielefeld University Library has followed open access strategies since longer. Since 2004 we are running an etd repository (BieSON) and since 2005 a virtual OAI server (BieTAS) to disseminate metadata for all electronic documents distributed on different platforms. Both are registered in the well-known OAI registries. The official decision of Bielefeld University in June 2005 to support the open access idea led to the development of a platform for all publications of university members which has been prepared by the library. While these efforts are located on the data provider side we are working on the service provider side as well.

We started already in 2002 with the concept of a scientific search environment based on modern search engine technology. Since OAI repositories and their documents became more and more relevant in the scientific community this content forms now the emphasis of BASE, the Bielefeld Academic Search Engine. BASE is a registered OAI service provider since summer 2005.

This lecture summarizes the design, the software architecture and the search environment of BASE. Based on the technology of FAST Search & Transfer it uses both web crawling and OAI PMH to collect metadata information. The work with OAI PMH led to a lot of tools to support the process of harvesting, correcting and filtering metadata to deal with problems of the OAI protocol and its usage.

The search engine is highly flexible and can be integrated as institutional repository search service in external environments as we did for our own repositories based on HTTP. In spring 2006 we added an integration of Google Scholar to add enhanced features as citation links. At present we are working on feeding the local OPAC data to provide a local view which includes the local print holdings. Besides that the BASE experience and technology will provide a basic module of the European Union project DRIVER, 'Digital Repositories Infrastructure Vision for European Research'. The technical integration concept will be implemented based on a SOA interface which will be used for other integration purposes as well.

Session 4: Quality Assessment

Citation analysis in research evaluation

Henk Moed (CWTS, Leiden, Netherlands)

This contribution discusses a number of studies that are currently carried out by the author and his colleagues at the Centre for Science and Technology Studies (CWTS) at Leiden University. A first cluster of studies relates to ranking of world universities, and a second to the effects of evaluation processes, and particularly of the use of bibliometric indicators therein, upon author publication practices and upon journal editorial practices. The core of this contribution presents the outcomes of a recent study examining the effect of "Open Access" upon citation impact and visibility.

Institutional repositories and research assessment

Les Carr (University of Southampton, UK)

Institutional Repositories provide a mechanism for individuals and institutions to collect and curate the published outputs - as well as the data results, intermediate analyses and digitised artefacts - of their research and scholarship. As well as being used as an Open Access literature for reference by other researchers, the repository provides a locus for administrative and managerial functions, including personal CV generation and institutional publication lists. In the UK, the Research Assessment Exercise has prompted many institutions to consider the repository as a useful mechanism for collecting evidence of research quality. This talk will describe the adaptation of repositories for research Assessment purposes (the JISC IRRA project) and report on the processes that have surrounded the use of the IR for RAE 2008 at the University of Southampton. In particular we will examine the conflicting goals of Open Access and Research Assessment that have both helped and hindered the development of the repository.

Session 5: Policies and Implementations

Overview of Repository Developments

Alma Swan (Key Perspectives Ltd, UK)

Repositories - both institutional and centralised - are coming on apace in terms of numbers, though content levels remain low. The remedy to that situation is proper policies for getting the content deposited and in this regard there have been a series of impressive new developments from funders and institutions. I will present a brief overview of the state of the policy art, the effects of a good policy and touch on how this all fits together into a coherent picture for the future of scholarly communication.

Open access a view from the bioscience funders

Robert Terry (Wellcome Trust, UK)

The Wellcome Trust was the first research funder to introduce a condition of its funding that there should be free online access to all the published research papers arising from Trust-supported research. This talk will outline how the Trust is supporting its policy through the funding of open access articles and the establishment of UK PubMed Central at the British Library in partnership with the US National Library of Medicine and a group of UK biomedical life science funders including the Department of Health, research councils and other major medical research charities.

Enlighten: Encouraging deposit at the University of Glasgow

Steve Beaumont (University of Glasgow, UK)

Enlighten, the University of Glasgow's institutional repository service was launched in early 2006. This launch was accompanied with the release of a university statement which "strongly encourages authors at Glasgow University to deposit copies of their published work into the University's Institutional Repository [Enlighten]". This presentation will provide an overview of the development of Enlighten, its transition from the JISC funded DAEDALUS Project to a full university service and the release of the University's statement. It will also consider the impact of the statement at Glasgow and the options available to staff to encourage them to deposit content.

Friday 20th October

Session 6: Sustainability

Permanent access to digital resources

Erik Olthmans (Koninklijke Bibliotheek, National Library of the Netherlands)

Electronic journals dominate the field of academic literature, and it is of great importance to the international scientific community that this electronic intellectual output remains accessible in perpetuity. It is uncertain, however, whether the traditional model, based on national deposits and geographical frontiers, will be able to guarantee the long-term safety of the international academic output in a digital form. Academic literature is produced by multinational publishers, and has often no longer a country of origin that can be easily identified and thus no obvious guardian.

Hence, in the traditional model there is a huge risk of academic records being lost forever. A systematic and more concentrated approach is needed to address this unacceptable risk.

In this presentation we discuss the policy and ambitions of the National Library of the Netherlands (Koninklijke Bibliotheek, KB) regarding digital archiving of and permanent access to electronic publications. We discuss three possible threats against permanent access, and we propose a coordinated and systematic approach to address these risks: the Safe Places Network. This paper also includes a comprehensive overview of the e-Depot system and the KB approaches to digital preservation.

Financial aspects of institutional repositories

John MacColl (University of Edinburgh, UK)

This paper will begin with an overview of the literature on the vexed question of the cost of institutional repositories, presenting the range of options suggested, and looking behind the figures to discern agendas which account for an often wide variation.

The University of Edinburgh has had an open access repository in operation for the past three years, and it will provide a case study to permit an examination of the lifecycle costs of running an open access repository. Comparison will be made between the University of Edinburgh model of in-house development and support of an open source-based repository, with models based on third-party services, including hosted repository solutions. The paper will argue that the coalescence of new operations around an open access repository gives rise to more profound realignments within academic library workflows and operations. The establishment of a successful open access repository can therefore become a catalyst for changing organisational structures and roles for libraries, and the cost implications for these – within a culture of cost-neutrality – will also be examined.

Outsourcing a Repository

Ingeborg Zimmerman (University Library of Zurich, Switzerland)

The University of Zurich has decided, following the signing of the Berlin Declaration and the Berlin 3 Open Access Recommendations, to

- require their researchers to deposit a copy of all their published and refereed articles in the Institutional Repository of the University of Zurich, if there are no legal objections
- encourage and support their authors to publish their research articles in open access journals where a suitable journal exists and provide the support to enable that to happen.

To get the Institutional Repository of the University of Zurich up and running as soon as possible, it is being outsourced for an initial period of two years. The presentation will discuss the advantages of this approach.

Session 7: Legal Issues

E-Repositories: Euro-Legal Issues for Publishers

Laurence Bebbington (The University of Nottingham, UK)

Abstract Publishing on the Internet crosses boundaries. Publishing via e-repositories crosses more than merely European territorial boundaries. This presentation will focus on the legal issues related to publishing online content via e-repositories. Although supposedly harmonised in some areas (e.g. copyright and database right) publishing carries different risks in different countries, even within the EU. This presentation will review the main risks associated with e-repository publishing (e.g. infringement of intellectual property rights; licensing issues; liability for information; defamation; offences peculiar to particular national laws and jurisdictions). An overview of the main risk areas will be presented together with a framework and strategy for legal compliance.

Licence to Publish: Science ajar

Wilma Mossink (SURF Foundation, Netherlands)

To count as an open access publication in accordance with the Berlin Declaration firstly a scholarly work must be placed in an online repository maintained by a proper organisation that seeks to enable open access. Secondly, the work must have a licence attached to it that allows the end user to reuse the work in any digital medium for any responsible purpose.

By signing the Berlin Declaration an academic institution declares it is in favour of the abstract principle of open access. Satisfying the two conditions of the Berlin Declaration however need concrete steps.

To fulfil the first condition of the Berlin Declaration many institutions have set up an institutional repository using suitable standards to enable open access. However, to act in accordance with the other condition is more difficult because the stipulated requirement throws up problematic or complicated legal problems. To deposit a publication that has been published in a journal that is not an open access journal is not at all times possible because the underlying publishing agreement not always permits depositing in a repository. Furthermore, not many authors or institutions add a licence to the work that permits the user to reuse the publication in such a way as described in the Berlin Declaration.

To overcome these legal problems the SURF Foundation has drafted two licences: a Licence to publish that sets new terms and conditions between author and publisher and a deposit licence between author and institution, which covers the requirements necessary to store and archive the publication and to realise the conditions for unrestricted distribution.

The presentation focuses on and explains the two licences that can assist and support authors and institutions to produce an open access publication. More background information is given on the process to draft and get accepted a Licence to publish that can be endorsed by leading European organisations that support the principles of open access.

Session 8: Academic Activities

Digital architects and their new houses

Sylvia van Peteghem (Ghent University, Belgium)

Research libraries tend to focus on scholarly communication making books, journals and articles easy accessible. Full text if possible. We all know that the first step to find information is the internet, also for a researcher. But most researchers are also teachers or lecturers and they need pictures and illustrations for power points, blackboard applications, publications etc. Scientists often keep a huge collection of thousands of photographs, slides and digital pictures on their subjects in cupboards and in one or the other database, on CD's, DVD's, hard disks, memo sticks... And there are even more pictures to find. Libraries with cultural heritage collections often keep original and completely unknown material that can be of interest for their own research community. We all know that during the last decade these parts of the research libraries were often a bit neglected to realise the best digital library in the world.

Versions of academic papers and open access : attitudes and current practice among economics researchers

Frances Shipsey (VERSIONS Project, London School of Economics, UK)

A typical research project conducted by an academic economist will involve the production of several different research outputs, each of which may potentially be made available in open access form. Report for funding body, conference paper, conference presentation, working or discussion paper, journal article, book chapter are among the most common types of output produced. Each of the research outputs will go through an iterative process during its development. When there are co-authors involved in the research, the number of distinct versions produced for each research output as it goes through the process of revision can become very large.

The VERSIONS Project, funded by the UK Joint Information Systems Committee, has conducted a user requirements study to investigate in detail the issues relating to versions in the context of open access digital repositories. The study looked at the experience of academic economists and considered their roles both as authors and as readers.

Issues relevant to open access repositories which were addressed in the study include:

- Authors' experience of managing versions of their own work, given the number of iterations produced – can authors easily locate an appropriate open access copy in order to self-archive their work in an open access repository?·
- Availability of older research outputs for self-archiving – the effect of moving institution, changing computers, or producing papers in the pre-digital era on researchers' ability to make older work available in open access repositories· Authors' attitudes towards making different versions of their own work openly accessible at different stages - issues may include concerns about releasing an 'unfinished product', uncertainty about IPR and permitted use, loss of citations, time needed to locate and deposit the open access copy

- Authors' intentions regarding deposit of their 'final author versions' in open access institutional repositories if requested (that is, in the absence of a mandate to self archive, how likely are authors to deposit their work?)
- Readers' experience of encountering multiple versions of others' work – almost all readers do come across multiple versions of the same work either very frequently, frequently or sometimes; for a sizeable minority it is not always quick and easy to establish which version(s) they wish to read
- Importance attached by researchers to various proposed approaches to describing and presenting different versions of the same work in ways that could aid discovery

Digitising heritage collections can give them a new life and there are wonderful examples of this already but how to make this useful to research and education and easy accessible?

The digital architects of Ghent University set up a sophisticated structure using aDore on top of existing databases, be it pictures or a book catalogue or full text, with one portal to many collections. Researchers are fond of it because they can fill their databases on their own speed, with their own restrictions if copyrighted and they can reach it wherever they are. And guess what: they are often happy to share their collection with the world!

IRIScotland and the Repository Landscape

Philip Hunter (IRIScotland Project, University of Edinburgh, UK)

The IRIScotland project has four main strands: the creation of a pilot hosting repository for individuals and institutions who do not have a repository or access to one; the creation of an aggregation of eprints metadata plus a search and browse service for the Scottish geographical area; a survey of the research community and university administrators in Scotland on attitudes to Open Access; and to effect a degree of cultural change in that community. This presentation discusses the creation of the IRIScotland service, the survey activity (and some conclusions), plus the place of the IRIScotland project in the Scottish and UK repository landscape.

Session 9: Beyond Research Papers

Open Data - Can repositories help?

Peter Murray-Rust (JISC Spectra Project, University of Cambridge, UK)

Most scientific data - even funded - is never published. This is in part due to technical difficulties, but largely due to a mixture of inertia and unwillingness to share. However the value of data re-use is enormous and increasingly required for eScientific applications such as the Grid.

The SPECTRA project - run jointly by the University of Cambridge and Imperial College with contributions from the Libraries and Chemistry departments has looked at the technical difficulties and social problem of making scientific data easily and freely available to the scientific community. We have shown that for certain classes of data, especially crystallographic, computational and spectral/analytical data that the technical aspects can be solved by systems which capture chemical and other

metadata at time of measurement or final analysis. The scientist of service group has the data in a form where it could be published for re-use. However the sponsoring scientist often feels that there should be a delay before publication, perhaps because of concerns that competitors might steal a march or that potential exploitation would be jeopardised.

We have created a repository system with differential access whereby scientists are invited to deposit the data into a "dark archive" or escrow where after an agreed time the data may be published.

Requirements and attitudes of chemists have been collected and a prototype system is being deployed. We expect that these studies will help to address the very large non-publication of scientific data.

Pulling the threads together

Paul Ayris (UCL (University College London), UK)

This paper will attempt to summarize the current state of development in Open Access Repositories. It will suggest likely drivers for change at institutional and subject level, attempt to identify new areas of development, challenges to the embedding of repositories into the information landscape, and the likely impact on conventional publishing models.