

DAEDALUS PROJECT
UNIVERSITY OF GLASGOW

Setting up a repository: practical advice

Morag Greig and William J Nixon
Project Managers

27 June 2005

Summary

This QA presentation is based on our experiences over the last three years of the DAEDALUS Project. They are not necessarily the only answers, or indeed an exhaustive list of the questions! We hope, however that you will find our experiences helpful.

It should also be noted that this work was done under ‘project conditions’ and this has had an impact on what, how and why we have done things in line with our project plan, targets and range of deliverables.

The workshop and slides are available at: <http://www.lib.gla.ac.uk/daedalus/workshop/>

Correspondence

Correspondence about this presentation should be sent to daedalus@lib.gla.ac.uk

Key Areas

- Policies
- Service Development
- Advocacy and Copyright

Policies

General Issues

- **What needs is your repository being set up to address?**
Open access, wider dissemination of research, RAE, digital preservation, a combination of these? Will determine how you approach things – how you persuade academics to deposit. NB – different stakeholders may have different agendas.
- **What sort of content are you seeking for your repository?**
Published RAE level material, other material, non published material? Theses? Some material will be easier to deal with. Do you want to put different content types into a single repository.
- **Will you be expecting authors to self-archive? Will you be depositing content for users? Or a combination of both? Will you provide training and support for self-archiving?**
May seem the ideal, but may find that academics are not keen and say they don't have time. Mediated deposit is resource intensive. However, self-archiving is likely to involve checking is accuracy of the data is important, and this is also time consuming.
- **Will the encouragement to deposit be coming from the top down, bottom up or both?**
This will depend on where the impetus has come from. Probably a combination of both is necessary. Senior management backing is important but uncertain whether a mandate to deposit is helpful or not – strong encouragement may be better. 'Bottom up' either requires winning hearts and minds and/or making a strong case for why it is in academics' interests to deposit.
- **Do you want to develop policies before you start or as you encounter issues?**
Ideally you would develop all necessary policies before you start. However, bear in mind that many issues may not occur to you until you have actually started to dealing with content, and so you are likely to have to change policies as you go along.
- **Funding body policies on open access – how are you going to incorporate these into the policies for your repository?**
Mentioned in earlier presentation. In particular, if RCUK implement the policy currently under consideration a great deal of thought will need to be given to the policies and procedures that will have to be developed to ensure that authors can fulfil these obligations.
- **What staff resource can you devote to setting up and managing your repository?**
We were extremely fortunate to receive JISC funding. We would strongly suggest that this is not an activity that can take place with minimal staff resources, e.g. one part time person. It is also the case that you will need staff with a range of skills – both technical and advocacy. We have found that a team based approach works well.
- **What will be the relationship of the repository to other University repositories/databases?**
Worth considering if you should try and work with/consider cross searching with any existing repositories in the University (without stepping on toes).

Content and Submission

- **Content – who will you accept content from?**
How about papers from staff who were at another institution when they wrote them but are at your institution now?
How about retired members of staff?
What about staff who have now left your institution?
What about students?
- **What will your policy on withdrawal of material from the service be?**
Ideally material should not be withdrawn from the service. However, there may be situations where it is necessary or appropriate.
- **Will you only accept full text or will you also accept bibliographic details plus links to versions on publishers' sites?**
We started out expecting only to include full text. However, soon became clear that publisher copyright policies were going to make it difficult to get large numbers of full text papers. Decided where this was the case to include bib details and a link to the version on the publishers' site. Not ideal, but those with a subscription will be able to access. Decision helped by academics pointing out that if we only included full text it looked like they hadn't published very much.
- **What is the minimum level of metadata you will accept?**
This applies whether authors are self depositing or you are offering mediated deposit. If self deposit will you insist that authors must include e.g. abstract and ISBN? If they don't will you add them? If e.g. staff/departments provide lists of publications for you to add will you simply use the details provided or try and track down missing details?
- **If authors are self-depositing what level of metadata quality checking will you undertake?**
Will you check each record before it is made publicly available? Can be time consuming. Will you add additional fields, e.g. subject headings (we add LCSH)?
- **Will you scan papers to add to the repository if electronic versions are not available?**
May be given permission to add published version of papers, but no electronic copy may be available. However, scanning will be resource intensive (unless author is willing to do it themselves).
- **Will you add metadata to indicate the version of a paper that is being deposited?**
Many publishers only permit the 'author final version' of papers to be added. Good idea to consider adding information to the record on which version of the full text is being made available. We have recently cover sheets.

Service Development

Selection, installation and maintenance

- **How will you select your IR?**
Selection for us was based on the software available, initially just ePrints and then broadened out to DSpace. Open source or vendor? Budapest OA Guide to Repository Software is now up to version 3. This links back to the use for the IR and your stakeholders.
- **Are you going to implement the IR in-house or outsource it?**
There is a range of Open Source and vendor solutions now available beyond just Eprints.org and DSpace. Biomed Central has recently announced its Open Repository service, Library vendors such as Innovative Interfaces Inc have also been developing a repository service which integrates with the catalogue.

- **What software skillsets are available? Or do you need to acquire?**
For us this has been Perl for Eprints.org and Java for DSpace
- **What hardware and operating systems are available? Or do you need to acquire?**
EPrints.org and DSpace were developed on Linux, we have ported them to a Sun Solaris box but that was not without installation and upgrade challenges. We will go into some of those challenges in the afternoon session on service development.
- **Will your service be hosted centrally or by the Library?**
Our service is hosted on a server based in Computing Services, all of the software is on a single box. Issues here include the backups and patching of the service as well as the ongoing upgrade path. Eprints is now at version 2.3.12 and DSpace 1.2.2.

Configuration and development

- **How much configuration do you want to do?**
The software will work “out of the box” but there may be a range of additional changes which you want to implement. In the ePrints service we added a “full text” flag and a range of additional © fields. Further details on our configuration work will be covered in the afternoon session.
- **Will you set-up a test service for development?**
It can be useful to have a development service particularly once the service is live, we have a test ePrints service which we use to trial changes to fields and also to refine the bulk imports
- **Will the IR integrate with existing institutional authentication systems such as LDAP?**
This is a longer term aim for us but it is important to note that without that kind of integration it is another layer of usernames and passwords for users.
- **What file formats will your repository accept?**
We use PDF’s in DSpace and Eprints. There may also be broader issues of resources available to create PDF’s on campus. We currently convert postprints provided by staff if they are not supplied as PDF’s.. We have avoided proprietary formats such as Microsoft Word.
- **How will your IR be administered?**
Centralised model or, perhaps with DSpace a devolved model, this could also be tied into choices about self-depositing in your service.
- **What will be the relationship between the IR and other institutional services such as the catalogue, VLE, publications database or administrative records?**
The IR is part of the wider knowledge network of the institution and it is important that it does not develop or exist in isolation.

Value added services

- **Will you provide a bulk import service for pre-existing data?**
For example to import bibliographic data from databases such as Reference Manager and EndNote, or for MARC records from your catalogue. Will you also want to import the fulltext? The ePrints Service is populated this way but it should be noted it is only the bibliographic details we import and not the full text objects themselves.. DSpace will also enable the bulk importing of the items.
- **What usage statistics will be available? Will they be posted publicly?**
Usage statistics options range from from hits to the site to the tracking of the downloads of papers. We use local PHP scripts to analyse the logs from ePrints; Do you want Top 100 listings made publicly available – we currently don’t do that. There may metrics/citation concerns amongst your academics? Publicly available statistics are

available from Australian Universities such as the University of Queensland and the Queensland University of Technology.

- **Do you need a local harvesting service?**
A harvesting service was a critical component for our us to ensure that there was one point of access for the range of content made available – and this will be demonstrated later.
- **Do you want to offer an open access journal platform?**
Range of options available here, we used EPrints.org for the Journal of eLiteracy to address a publishing need and to gain experience in this area. Interest and support in open access e-journals was not an area we anticipated interest from academics but which is one for which there is a demand.
- **What training in administration and deposit will you offer to your community?**
Depending on the model of deployment from highly centralised to devolved what support and training will you offer to departments. This can range from the development of publications databases for them to guidance on administration of, say a DSpace community. We will be running training courses for staff later in the year about content for ePrints and depositing content into Dspace.

Advocacy and copyright

Advocacy

- **Will you be targeting all departments simultaneously? Or starting with a small number (e.g. a pilot)?**
Tackling the whole University at once is probably not to be recommended. Probably best to start with a small number of departments. Can lose good will if too much is promised and then not delivered.
- **How are you going to get some initial content to populate the repository with?**
Demonstration of the repository is critical in gathering content, but you need some content in it first in order to do this! Classic ‘chicken and egg’ problem. Also, not ideal to put in Library material, as this isn’t likely to be perceived as quality content – won’t give the right impression. Need to try and find some keen academics to be ‘early adopters’. Ariadne article on content gathering strategies. Don’t underestimate the difficulty of getting content – no magic tactic.
- **How are you going to ‘sell’ the idea of a repository? How will you persuade staff to deposit?**
What is your ‘sales pitch’ going to be? This may differ depending on who you are trying to persuade. Are you going to push the open access agenda/funding body requirements or the benefits of repositories? May need to develop this as you go along.
- **Do the academics in your institution know much about open access? How much informing on the background issues will you need to do?**
Talking about repositories and open access to academics can get quite complicated – lots of strands involved and different agenda. People can get confused and end up with misleading ideas. Need to establish what sort of levels of awareness of open access there are in your institution and use this to establish how much background information you will need to give to staff in presentations etc. Be clear that repositories are not an alternative to publication – message cannot be repeated enough.
- **What publicity activities are you going to undertake?**
What activities will you undertake? Presentations – to which committees? Events? Publicity material? Articles in University publications, e.g. newsletters? Who will deliver sessions and how will you get people to come? List on DAEDALUS of some of the promotional activities we have undertaken.

- **Are you going to use other Library staff to undertake advocacy activities?**
Good if you can, as will add extra resource. However, will need to make sure everyone is fully briefed and consistent.

Copyright

- **Are you going to advise academics on copyright matters?**
Many academics find copyright issues confusing and worry that they may be contravening copyright agreements by depositing content in repositories. It is likely you will need to offer advice to academics on copyright issues.
- **Will you expect authors to check copyright agreements or will you do this for them?**
It would be ideal if authors did do this, and you should definitely encourage them to look more carefully at the agreements they sign than they probably do currently. However, authors may be unwilling to check copyright agreements themselves. Also they may not be fully aware of all the parameters involved, e.g. author final version, personal web sites only etc. SHERPA/RoMEO list of publisher policies available, but not infallible.
- **Are you going to approach publishers for permission to add articles if copyright publisher policies are not clear?**
A lot of policies are quite vague may not specify if deposit in a repository is permitted. If you want to be certain that deposit is permitted you may need to approach publishers directly and ask permission. We have done so with quite high rates of success.
- **If so, what will you do if publishers don't respond?**
Would be ideal to adopt the approach of 'if no response within a reasonable period of time we'll go ahead anyway and take content out if necessary'. However, JISC Legal have advised that this is a dubious position to take. All relates to risk and how much you are prepared to take.
- **What will you do about advising staff on keeping/creating 'author final versions' of their papers?**
This is critical, as many of the larger publishers (Elsevier, Springer) have adopted the position that 'author final versions' of papers can be deposited but not the publishers official pdf. Authors need to be advised to start keeping suitable versions of their papers. Tricky issue, as no guidelines available on exactly what constitutes a suitable copy. Authors may not have suitable versions of older papers.
- **What will you do if a publisher asks you to remove full text from the repository?**
Obvious answer is to take it out, assuming it is because you have put in full text that you shouldn't have. Best to try to be careful that you have done reasonable checking. Rather a 'grey' area.
- **If authors deposit versions of their papers that are not acceptable to publishers in respect of copyright agreements will you remove them?**
Difficult, as you don't want to discourage authors from depositing. However, don't want to breach copyright agreements. May need to contact author and advise why content cannot be held in the repository (or ask for author final version rather than pdf).

Prepared by: William J. Nixon

Last modified on: Monday, 04 July 2005