JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

Final Report

Project

Project Title	Repository Bridge: Automated Linkage of National and Institutional Repositories		
Start Date	01/07/2005	**End Date**	30/06/2006
Lead Institution	University of Wales Aberystwyth		
Project Director	Bill Hines (whh@aber.ac.uk) 01970 62 2392		
Project Manager & contact details	Stuart Lewis (stuart.lewis@aber.ac.uk) 01970 62 2860 Information Services, UWA, Penglais Campus, Aberystwyth, Ceredigion, SY23 3DB, Wales		
Partner Institutions	University of Wales Swansea National Library of Wales		
Project Web URL	http://www.inf.aber.ac.uk/bridge/		
Programme Name (and number)	Projects in Digital Repositories (3/05)		
Programme Manager	Neil Jacobs		

Document

Document Title	*Final Report*		
Reporting Period	*NA*		
Author(s) & project role	Stuart Lewis – Project Manager		
Date	28/06/2006	**Filename**	Repository Bridge Final Report v1a.doc
URL	http://www.inf.aber.ac.uk/bridge/private/		
Access	Project and JISC internal		General dissemination

Document History

Version	Date	Comments
0b	13/06/2006	First draft sent to JISC
0c	22/06/2006	First draft circulated to project board
1a	28/06/2006	First public release

Repository Bridge Final Report

Table of Contents

Acknowledgements 3
Executive Summary 4
Background 5
Aims and Objectives 5
Methodology 6
Implementation 7
Outputs and Results 9
Outcomes 10
Conclusions 12
Implications 12
References 13

Acknowledgements

The Repository Bridge project has been funded by the Joint Information Systems Committee (JISC) as part of the Projects in Digital Repositories (3/05) programme with contributions from the project partners. The project wishes to acknowledge the funding and additional support received from JISC and the partners. Particular thanks go to:

- JISC:
  - Neil Jacobs: Programme Manager.

- University of Wales Aberystwyth:
  - Project director: Bill Hines.
  - Head of Department: Michael Hopkins
  - Project manager: Stuart Lewis
  - Project officer: Jonathan Bell
  - Project advisor: Elizabeth Kensler
  - Project advisor: Dorothy Hartley
  - Project advisor: (Department of Information Studies): Lucy Tedd

- National Library of Wales:
  - Project manager for NLW: Sally McInnes
  - Project advisor: Arwel Jones
  - Project programmer: Glen Robson

- University of Wales Swansea:
  - Project manager for UWS: Martin Price
  - Project advisor: Andrew Brown

Thank also go to the technical and management staff of the EThOS project for their close collaboration. The project was also assisted by input from Andrew Green (the Librarian of the NLW) and Lucy Burrows (Records Manager, Cardiff University).

Executive Summary

The aim of the Repository Bridge project was to investigate and implement a technical solution that allows electronic theses deposited in Welsh university repositories to be exported automatically to an archival repository hosted by the National Library of Wales.

The idea of the project was conceived on the back of an agreement which dates back nearly 100 years between the National Library of Wales and the Welsh universities. The agreement allows the National Library of Wales (one of the UK copyright libraries) to store a paper copy of all higher-level or Welsh-interest theses produced in Wales. With the move towards electronic storage of theses, the project aimed to produce a system which would allow the National Library of Wales to continue with this agreement, but to collect the theses electronically. Benefits of this include:

- Automatic rather than manual submission of theses to the National Library of Wales
- Reduced duplication of metadata creation and cataloguing
- Reduction in expensive storage space requirements
- Alignment with the principle of LOCKSS (Lots of copies keeps stuff safe)

The approach taken was to make use of the standard harvesting protocol OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) to transfer metadata between the institutional and library repository systems. This metadata includes identifiers for the content files, so these can be imported and ingested by the National Library of Wales.

Work was completed in collaboration with the EThOS[1] project to ensure that standards were drawn up and used (UK Electronic Theses and Dissertation – UKETD), allowing the National Library of Wales to act as a regional hub to provide all of its electronic these to the EThOS service. This collaboration also used the National Library to pilot it consortium-based submission model. The collaboration was beneficial to both projects ensuring consistent use of standards and sharing of procedures and policies.

The final software release comprises of five modules as follows:

1. Export of metadata from repositories running DSpace[2] via OAI-PMH
2. Export of metadata from repositories running GNU eprints[3] via OAI-PMH
3. Ingest of exported metadata and of content into Fedora[4] running at the National Library of Wales
4. Exposure of UKETD metadata from Fedora
5. A web-based system to allow institutions to manage which parts of their repositories are harvested

An additional requirement of the metadata exported from Welsh institutions for the National Library was that they preferred the data to be encoded using METS (Metadata Encoding and Transmission Standard) with metadata encoding using MODS (Metadata Object Description Schema) rather than Dublin Core as used by EThOS. The result was that the metadata exported from the Welsh institutions to the National Library of Wales was transmitted using METS with MODS, but also included a duplicate of the metadata encoded in Dublin Core. This method was highlighted by the *Linking UK Repositories* report [1] as an innovative way of using METS to deliver metadata to multiple systems.

The final version of the software is now completed and installed at the National Library of Wales, the University of Wales Aberystwyth and the University of Wales Swansea. The project has a close working relationship with the remaining higher education institutions within Wales and will continue to work with them to work towards harvesting their electronic theses.

[1] Electronic Thesis Online Service (http://www.ethos.ac.uk/)
[2] DSpace repository software (http://www.dspace.org/)
[3] GNU eprints repository software (http://www.eprints.org/)
[4] Fedora repository software (http://www.fedora.info/)

Background

The National Library of Wales (NLW) acts as the national archival repository for Welsh higher level research theses in traditional print format. As such it receives copies of all doctoral and research masters level theses produced by Welsh universities together with those taught masters dissertations of an especially high standard or of specific Welsh interest. Welsh HE institutions such as the University of Wales Aberystwyth (UWA) and the University of Wales Swansea (UWS) have internal procedures and processes within their Information Services (IS) departments to deposit copies of appropriate theses with the NLW. These existing arrangements are described in the project's *Report on thesis submission procedures* [2], together with discussion of how these procedures might be amended to accommodate deposit of electronic copies of theses.

The project 'Repository Bridge: Automated Linkage of National and Institutional Repositories' came about following trials of the DSpace (http://www.dspace.org/) software as an institutional repository to store electronic copies of academic theses. Many FE and HE institutions are moving towards both electronic submission and storage of theses and documents, so as an institution, UWA started a trial of the e-repository software DSpace to fulfil these functions. At the same time, the NLW started trials of Fedora (http://www.fedora.info/) to archive not only theses, but many types of digital asset.

If both the institutional and national level repositories are to become electronically-based, similar processes and procedures to these existing arrangements are going to be required to copy assets between repositories. The required updating of these procedures also gives the opportunity to streamline and automate aspects of their operation, such as avoiding duplication of effort in cataloguing these shared items. Without the facility to easily copy assets between repositories, the processes will fail to work as effectively as they could.

Aims and Objectives

The aim of the project was to create an electronic bridge from the institutional repositories at UWA and UWS, running DSpace, to the national repository at the NLW, running Fedora. Appropriate theses can then pass along this bridge from the institutional repositories, allowing them to be stored in the national repository with little, if any, manual intervention.

The specific objectives that needed to be fulfilled for the project to complete successfully were:

- Documented understanding of how the current manual processes work
- Investigation into the in-built facilities of DSpace and Fedora to export and import items
- Implementation of software for automated import of metadata and content to the NLW
- Integration of the working model into the UWA and UWS library working processes
- Testing and evaluation of the tool written to ensure it fulfilled its purpose

Additional objectives associated with the project, that the project helped to drive but was not responsible for included:

- Promoting and advocating the benefits of electronic theses and institutional repositories
- Advising on technical aspects of hosting institutional e-thesis repositories

If time allowed during the project, there were some additional related issues that could have been investigated. These issues are all pertinent to the Welsh HE community where similar systems have been adopted. The additional tasks that could have been investigated were:

- Cataloguing workflow with DSpace and Voyager (UWA / UWS library catalogue system), or other OPAC systems
- Internationalization of the DSpace allowing it to be used bilingually with English and Welsh interfaces

The cataloguing workflow was not investigated due to time constraints, while the internationalization was undertaken centrally by the DSpace community.

An additional objective was added to the project and completed. Cardiff University opted to use GNU eprints rather than DSpace to run their institutional repository. It was decided that supporting EPrints was more important than investigating cataloguing issues so this work replaced the cataloguing workflow with DSpace and Voyager as mentioned above.

A final objective added to the project was close collaboration with the EThOS project to define standards and processes associated with the movement of electronic theses between repositories within the UK. This work added an additional overhead to the project, but was welcomed by all as an essential aspect of the project.

Methodology

The project was split into three phases:

1. Research (3 months)
2. Implementation (6 months)
3. Testing, evaluation and dissemination (3 months)

Phase 1: Research (3 months)
During the initial research period, investigation took place into the workflow requirements dictated by the way the institutions and the NLW currently interact, and the support for export and import provided by DSpace and Fedora respectively. These reports allowed the software requirements document to be written.

The very simple approach of documenting the current workflows and investigating the current software functionality was adopted to allow the project to define the requirements of its software very quickly. With the project duration of only twelve months, effort was not available to re-engineer workflows or to make significant changes to repository software. By evaluating the current practices and facilities we were able to finalise the requirements specification by the end of the initial three months of the project. Other than the late addition of the requirements to support export from GNU EPrints, the requirements specification remained static during the following nine months of the project. Completing the research stage of the project early also allowed us to start discussions on adopting changes to the UWA's submission procedure to support the deposit of electronic copies of completed theses, initially alongside the paper copies.

Phase 2: Implementation (6 months)
The first stage of the implementation was to define the system design. Two designs were proposed and evaluated, one based on open standards (OAI-PMH exposing METS/MODS/qDC), and the other on a tightly integrated system based upon web services (the NLW queries and harvests from the DSpace repositories using web services). To minimize development effort and to make use of open standards, the first approach was adopted.

The implementation of the final software system was split into four parts, each of which was developed concurrently:

1. **Software to export data from DSpace**
   - This software was written using the crosswalk mechanism built into DSpace[5] to deliver the required metadata formats from the in-built OAI-PMH provider.

2. **Software to import data into Fedora**
   - This software uses a harvester to programmatically obtain metadata for new records. It does this using OAI-PMH. The metadata includes identifiers for the content files themselves, allowing them to be obtained and ingested into Fedora.

[5] http://www.dspace.org/technology/system-docs/application.html

3. **Software to expose metadata from Fedora for EThOS**
   - Our collaboration with the EThOS project led to a requirement for software to allow the NLW's Fedora archive to expose metadata in the UKETD qualified Dublin Core format through OAI-PMH for the re-export of metadata and content to EThOS.

4. **A web front end to allow repository managers to administer their part of the bridge**
   - As the Repository Bridge is intended to be deployed at the NLW for all Welsh HEIs to export theses, there is a need for the HEIs' staff to administer their aspects of the Bridge, such as which collections theses are to be harvested from. A web front end has been provided to allow such administrators to log on and change their institutions' settings as well as view logs of their institutions' interaction with the bridge.

In addition, so as to support the deployment of the bridge across Wales, a fifth module has been developed to support the exposure of records in the correct metadata formats for institutions using GNU eprints. This duplicates the first of the four modules listed above and allows the NLW to use the bridge to harvest records from institutions that run either of the common choices of institutional repository software.

Phase 3: Testing, evaluation and dissemination (3 months)
The software has been installed in the NLW and is being run regularly to harvest items from institutional repositories at the UWA and UWS. For testing, this used development implementations of the institutional repositories, allowing test theses to be exported. This was done partly to avoid the danger of any errors with the software affecting the integrity of the production repositories and partly to allow a faster throughput of items than would be possible if we had to await deposit of actual theses.

Whilst dissemination of the results of the project has been planned for this final stage of the project's work, in practice it has run in parallel with other work, with advantage being taken of opportunities to present the project and its work as they arose. This has included presentations and discussions at a local level, with the UWA's Senior Management Team and the online user group, at a national (Welsh) level with presentations to the Welsh Higher Education Library Forum (WHELF), at the Association of University Administrators annual conference, and at EThOS's Cardiff advocacy event and at an international level with a presentation at the DSpace User Group meeting in Bergen. In addition, a paper on the project's use of OAI-PMH is to appear in the journal *Program: electronic library and information systems* [3].

Implementation

The implementation can conveniently be discussed in subsections corresponding to the three phases of project work introduced in the preceding section.

Research
The research phase of the project consisted of three principal activities:

- Investigation of the exciting arrangements for theses deposit.
- Investigation of the export facilities in DSpace.
- Investigation of the import and ingest facilities in Fedora.

The results of each of these investigations were written up and presented in reports.

The investigation of existing arrangements centred on UWA and UWS as the university partners in the project. As these are both constituent colleges of the federal University of Wales, their arrangements are similar. Both colleges produce written guides for postgraduate students and their supervisors and these were used as the starting point for our investigations. We also corresponded and interviewed members of staff from the libraries and academic offices of both the colleges and also of the University of Wales central academic office to gain a more complete picture of the existing arrangements for submission and deposit of theses. The findings were written up in [2] which also discusses possible approaches to adding the deposit of electronic copies of theses to the existing arrangement.

The investigations into DSpace and Fedora were undertaken in parallel, the UWA project staff investigated DSpace while the NLW staff investigated Fedora. As both of these systems are open source, investigation could and did include examination of the source code as well as the software documentation. The results of both investigations were written up in reports both of which are available on the project's web site [4,5]. Fedora uses the Library of Congress Metadata Encoding and Transmission Standard, METS (http://www.loc.gov/standards/mets/) as a starting point for ingest. This preferred approach makes use of the Metadata Object Description Schema, MODS (http://www.loc.gov/standards/mods/) for the descriptive metadata. As DSpace was found to have some support for both these standards, they were felt to be a useful starting point for the projects export and import.

Implementation of software

The software was developed at UWA and the NLW, being divided into separate modules as outlined in the section on Methodology. Two alternative approaches to design of the software were considered, either developing our own methods for harvesting and import using web services, or taking advantage of open standard approaches to harvesting. We adopted the second of these approaches, for the following reasons:

- The use of open standards should reduce the future maintenance overhead. This is important when a project is of fixed duration, as it increases the sustainability of the resulting software.
- Support available in DSpace and Fedora. Our investigations into DSpace and Fedora showed that both systems had some measure of support available for suitable open standards for the harvesting and import of metadata, thus reducing the project's workload.
- Consistency with EThOS. EThOS have also adopted similar open standards for their metadata harvesting, and as our collaboration with EThOS had resulted in the idea that the NLW act as the hub for Welsh HEIs, there were advantages in taking a similar approach. This also allowed the projects to benefit from each other's work.
- Increased portability. While this is of limited value to the project itself, the use of open standards means that software developed should be readily adaptable both for other software (as our development of export from GNU eprints demonstrates) and for other types of content.

Inevitably, there are disadvantages with this approach. The most significant is that the use of OAI-PMH limits the ways we can restrict harvesting of metadata. Specifically, the OAI-PMH allows the harvest to be restricted by date and by "set". Restriction by date is obviously useful to us as we can restrict a harvest to those items added to the institutional repository since the previous harvest by the bridge. Harvest by set can be used to ensure that only theses are harvested, but only by ensuring that all theses are placed in the right set, and that set only contains theses. As DSpace equates an OAI-PMH "set" with a collection, this places requirements on the design of an institutional repository using DSpace. These implications are discussed in [3] and also in a project report, *Harvesting Records and Repository Design* [6].

Of the modules listed in the preceding section, the export facilities for both DSpace and eprints and the web front end were developed at UWA and the Fedora export (for EThOS) was written by the NLW staff. The remaining module, the importer was divided between the two institutions. The UWA wrote a program to harvest the metadata using OAI-PMH and store the imported metadata in a temporary file and the NLW wrote the code to perform the ingest of the metadata into Fedora and to obtain and ingest the content files. This division of labour fitted well the with each institutions capacity for undertaking project work as well as making use of the NLW's expertise with Fedora and the expertise the UWA acquired with OAI-PMH and DSpace. The software itself is described in the following section, on Outputs and Results.

Testing, evaluation and dissemination

The import software was installed in the NLW and the necessary DSpace export modules installed at both UWA and UWS. Most testing has been done using UWA's development DSpace repository to avoid any danger of corruption of the operational installation and to allow the use of easily generated test data, so as to increase the volume of testing possible. The most important testing activity simply involves depositing items ion the thesis collection of the test DSpace repository and ensuring they are correctly imported by the Bridge code. In addition, the web front end has been evaluated, to test that

the required operations are supported and can be conveniently carried out. This also tests the logging facilities. The logs can be accessed by the NLW as the importing institution and each exporter's logs can be viewed by that institution.

The most important aspect of evaluation of the project work is to establish the correctness of the export process, which the testing programme is doing. The brevity of the project's lifespan means there has not been time for thorough evaluation of some aspects of the project's work, there is room for more evaluation (user testing) of the web front end and also of examination of the acceptability of the repository design that results from the project's use of OAI-PMH. As the project software runs automatically there is relatively little need for testing of user interface components.

As noted above, dissemination ran in parallel with other aspects of the project work so as to take advantage of opportunities as they presented themselves. The events at which the project work was presented were listed earlier. In addition to these events, the project has given the UWA and UWS the opportunity to launch their institutional repositories and the project staff at UWA have taken part in this work, both setting up a working institutional repository, known as CADAIR (http://cadair.aber.ac.uk/) and liaising with academic departments to help and encourage them to start using the service. So far the Department of Computer Science has started depositing material, that department having volunteered to act as the pilot scheme guinea pig. Other departments are now starting to use the service. Another related aspect of the project's work has been to start the ball rolling with regard to electronic deposit of theses. Discussions on this have taken place with UWA's academic office and with the university's senior management team. This has resulted in a set of regulations for submission of research theses is to support the deposit of an electronic copy of a thesis alongside the paper copy being considered for adoption by UWA.

Outputs and Results

The most important output of the project is the software for automated import of theses into the NLW's Fedora archive. In addition to this, the project's investigations and findings have been documented, in the reports listed in the references.

It has already been observed that as DSpace and Fedora support the use of METS and MODS for export and import respectively, this was used as our starting point for the process for harvesting metadata, using OAI-PMH. Unfortunately shortcomings with the DSpace METS support were identified so some work was needed to overcome these. The DSpace METS support was changed to support the use of multiple descriptive metadata sections. This was necessary as we chose to export the UKETD metadata for re-export to EThOS. As both DSpace and EThOS have standardised on versions of qualified Dublin Core, this was felt easier than converting the metadata to MODS and then having the NLW convert it back to Dublin Core for re-export. In fact the NLW Fedora repository simply keeps the Dublin Core metadata as an additional datastream alongside the content and makes no attempt to use it. The METS export file was written in such a way that the EThOS UKETD export file was used to generate the UKETD Dublin Core metadata. This involved slight changes to the EThOS code, but they adopted these changes, so any changes to the EThOS metadata set will be carried across to Repository Bridge on installation of the updated file.

The importer module sits in the NLW and periodically sends an OAI-PMH query to each contributing institution. This results in the metadata for any new items in the thesis collections being imported and, temporarily, stored in a separate file structure ready for processing by the ingest tool. This tool reads the files, and uses the file identifiers given in the METS to obtain copies of the content files themselves. They are then ingested into Fedora, as is the UKETD Dublin Core metadata. The importer is run periodically, and the date of the previous harvest is recorded so items are only imported once. Where the contributing institutional repository runs DSpace, the items are selected for export by depositing them in a collection from which export is carried out. This is done for all PhD and research masters theses as well as those taught masters dissertations considered to be worth preserving in the NLW. This means that the choice of which of these dissertations to export is in the hands of the contributing institution, not the NLW. Provided the thesis is deposited in the correct collection, export is automatic.

Broadly speaking, the EThOS import works in much the same way, their importer also uses OAI-PMH and this requests metadata for any new items from the NLW which in response simply forwards the UKETD metadata. EThOS can then use this to identify the content files.

These processes are undertaken without human intervention, but a user interface is required so that those responsible for administering the system can view the logs generated on import and can set up the necessary details of the contributing institution. The UI is a Web front end that supports the following tasks

- Set the collections the harvester should import metadata from.
- Change the contact details for an institution, so email alerts can be sent.
- Give access to the log database for that institution.
- Change the institution administrator's username and password.

The only other manual task is the selection of taught masters dissertation for export but this is done as part of the submission process for that dissertation so is done in DSpace not the thesis archive system.

The testing carried out suggests that the software developed fulfils the principal aim of the project by allowing the automatic import of theses and the associated metadata into the NLW's electronic archive.

Outcomes

In this section, we briefly discuss the results of the project in terms of the aims listed in Aims and Objectives above. As many of the result and outcomes have already been described, these will be kept brief.

Documented understanding of how the current manual processes work
An early task was to document how the existing thesis submission and deposit processes work and provide some discussion of how the deposit of electronic copies of theses might be incorporated into the existing process. At present, the preferred approach seems to be to have the electronic copy deposited in the institutional repository at the same time as the corrected thesis is presented for deposit in the university and national libraries. This is similar to the approach adopted at Cranfield, written up in [7]. This means that care must be taken to ensure that the electronic and paper copies deposited are identical in content. The initial approach taken is to have the depositor of the electronic copy sign a declaration that it is identical to the paper copy. The existing processes at UWA and UWS and the incorporation of deposit of the electronic copy are discussed in *Report on thesis submission procedures* [2]. This document is available from the Repository Bridge website, the URL is given in the reference.

Investigation into the in-built facilities of DSpace and Fedora to export and import items
As discussed previously, the export facilities in DSpace and the import facilities in Fedora were both examined and as they were found to have the use of METS and MODS in common, that standard and schema were used as the starting point for metadata transfer. The results of these investigations were described in *Import and Export with DSpace* [4] and *Fedora Import Functions* [5]. Both of these reports are available from the Repository Bridge website.

Implementation of software for automated import of metadata and content to the NLW
Most of the preceding sections have been taken up with description of the software that was developed and implemented as the major output of the project. The testing undertaken shows that the software is capable of fulfilling its objective of automated export of theses to the NLW and the additional objective of exporting thesis to EThOS. The use of established open standards is expected to reduce the maintenance overhead of the software and so promote its continuing usefulness.

Integration of the working model into the UWA and UWS library working processes
As the export of theses is automatic, integration into the library working practices is straightforward, at least from that point of view. The only operation a human user need undertake is to ensure that all theses and dissertations for export are placed in a collection that the importer harvests from. This is taken to be a part of the submission process. The need for such specific collections for export is discussed in *Harvesting Records and Repository Design* [6], available on the project website.

Another area of integration into library working practices that is less concerned with the bridge itself than it is with the introduction of an institutional repository is how the deposit workflow is managed. DSpace allows some configuration of the deposit workflow and so some thought is needed as to how this is combined with the library processes. The preferred alternative at present is to have the initial deposit made by the student (or the department) and have the deposit checked by a member of library staff. This allows the library staff to ensure that the metadata looks correct (and possibly add subject classification information that the initial depositor cannot) and more importantly to ensure that the thesis or dissertation does appear in a collection from which it will be harvested. DSpace has the facility to allow an item to be "mapped" from one collection to another, so it appears to be in both. This supports the idea that the library staff can arrange fro the thesis's export. As both UWA and UWS are adopting DSpace this workflow (or variations of it) is applicable to both. It is worth observing that GNU eprints works differently, and the project has not considered how the use of GNU eprints would affect the deposit and export workflow.

Testing and evaluation of the tool written to ensure it fulfilled its purpose
The software has been installed at the NLW and the DSpace export code installed at both UWA and UWS. Export of test theses has been undertaken, both between UWA and the NLW and between the NLW and EThOS. These tests have been successful. It is worth noting that there is scope for further testing of the GNU eprints export code and also for usability testing of the web front end, although initial experience with the front end suggests it is usable.

Promoting and advocating the benefits of electronic theses and institutional repositories
There have been two sides to the advocacy work, one concerned with the introduction of the facility of electronic deposit of theses and the other concerned with the introduction of an institutional repository in general, promoting its use for other output besides theses.

The electronic deposit of theses has been discussed with the library and with the academic office at UWA. This has resulted in a draft set of regulations for submission of theses going before the examinations committee. These changed regulations support the voluntary deposit of an electronic copy of the thesis alongside the paper copies and seek to ensure that the student who so deposits a thesis is made to declare that the content of both formats of thesis is identical and that he or she is aware of the implication regarding publication of third party materials and has obtained the necessary clearance for the use of such materials. Advantage has also been taken of the opportunity to promote the electronic deposit of theses to the UWA's postgraduate students.

In parallel with the work on electronic theses, the project has installed a working DSpace institutional repository, known as CADAIR, and has started working with academic departments to start populating this repository. This work started with the Department of Computer Science at UWA, after a meeting was held to both promote the idea of an institutional repository and to gain a clearer idea of the requirements an academic department might have for such a repository. This resulted in that department being the first to start using CADAIR. More recently, similar meetings have been held with the Department of Information Studies and they have also started using CADAIR. A meeting was also held with the Senior Management Team of UWA, promoting the idea of an institutional repository and discussing its possible use with relation to the RAE.

Advising on technical aspects of hosting institutional e-thesis repositories
Much of the advocacy work described above has also covered advising on technical aspects of running institutional repositories. This includes consideration of how a repository might be laid out and, given that academic departments have their own collections, how each department might take a role in administering its collections. Another area that has been considered is the legal implications of deposit in an open access repository and the project has provided an opportunity for UWA to advise users on these legal facets, such as the inclusion of third party copyright materials.

The section on Aims and Objectives lists two additional aims that have not been met, besides those discussed above.. The integration of DSpace with Voyager was dropped as we felt the development of export facilities for GNU eprints was more closely related to the overall aims of the project and so should take priority. This work was completed by the production of Perl packages to support the exposing of GNU eprints metadata in formats that match the DSpace metadata, enabling the NLW to import form repositories that run either software. In addition, as noted in Aims and Objectives, internationalisation of DSpace has been undertaken by the DSpace community, such that all that is

required for a given language is the production of a text file with the phrases used by DSpace given in that language. The project staff lack the knowledge of Welsh to undertake this work and it is hoped that it will be carried out as part of the work of the Welsh repository support incorporated into the Repository Support Network Project in the current JISC programme call.

Conclusions

The project has succeeded in its primary aim of developing software to enable the automatic import of theses from institutional to archival repositories. Along with the similar work undertaken by the EThOS project, we have demonstrated that combining OAI-PMH and METS is an appropriate approach to achieving this export, especially because of the support for alternative encodings of descriptive metadata. This approach has the following advantages:

- Reduction of the maintenance overhead of the software, as support for the underlying standards is available in repository software.
- Ease of arranging support for other software platforms. As it is generally the case that repository software supports OAI-PMH, it is relatively straightforward to arrange support for different platforms. We have demonstrated this by providing matching export functionality for GNU eprints although this was not originally part of the project brief.
- Flexibility in extending the approach for other types of material besides theses. While this has not been explored, it should be straightforward to use the Bridge software for export of other types of material. It might be necessary to extend the METS supported by the institutional repository export facility for certain types of material, but this should be straightforward. As it stands, the bridge software could be used for other written materials, such as papers.

Areas that the repository community as a whole might take the work of the project forward are outlined in the following section.

Implications

Because the actual export / import process is automatic, it has few implications for users, whether or not they are library professionals. The most important implication is the need to ensure that all materials that are to be exported are placed in a collection from which the import harvests. This is a result of using OAI-PMH and of the fact that DSpace treats an OAI-PMH “set” as synonymous with a collection. From the point of view of the project itself, this has the advantage that materials that are not susceptible to automatic selection for export can be selected by placing them in an appropriate collection as part of the submission process. In our case, this applies to certain taught masters dissertations.

The need to ensure that items for export are placed in an appropriate collection or set would be handled differently when the exporting repository uses GNU eprints. As GNU eprints handles sets differently from DSpace, this is an opportunity for further investigation. This is one area where the community might be able to take the work forward, as the project’s original brief did not include support for GNU eprints. As the approach taken is applicable to other types of material besides theses, it is not impossible that there is sufficient interest in the community for this to be done. Our development of METS and MODS support for GNU eprints also provides a possible starting point for this work.

Another possible field for future work is examining how the approach can be expanded to cover other types of materials. This need consist of little more than expanding the METS and MODS used in the export modules to enable their use for material that requires different sets of descriptive metadata and possibly a more complete set of structural metadata than has been implemented by the Repository Bridge itself. The flexibility of METS and MODS should make this feasible.

References

[1] A. Swan and C. Awre: Linking UK Repositories, Scoping Study report (http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_report.pdf)

[2] J. Bell: *Report on thesis submission procedures*, Repository Bridge project report (http://www.inf.aber.ac.uk/bridge/documents/Notes on submission procedures.doc)

[3] J. Bell and S. Lewis: Using OAI-PMH and METS for exporting metadata and digital objects between repositories, to appear in *Program: electronic library and information systems*, 2006

[4] J. Bell: *Import and Export with DSpace*, Repository Bridge project report (http://www.inf.aber.ac.uk/bridge/documents/export.doc)

[5] G. Robson: *Fedora Import Functions*, Repository Bridge project report (http://www.inf.aber.ac.uk/bridge/documents/Bridge Project- Fedora import.doc)

[6] J. Bell: *Harvesting Records and Repository Design*, Repository Bridge project report (http://www.inf.aber.ac.uk/bridge/documents/Harvesting and design.doc)

[7] Cranfield University: *Research thesis submission and examination process* available online at http://www2.rgu.ac.uk/library/etds/workflow.html