

Implementing an Institutional Repository for Leeds Metropolitan University

Final report

Authors: Wendy Luker
Nick Sheppard

Contact: Wendy Luker
w.luker@leedsmet.ac.uk

The Headingley Library, James Graham Building, Leeds Metropolitan University, Beckett Park, LEEDS, LS6 3QS

Tel: 0113 8127468

mobile: 07826876549

Table of Contents

Table of Contents

| | |
|--|----|
| 1. Acknowledgements | 4 |
| 2. Executive Summary | 5 |
| 3. Background | 6 |
| 3.1. Integration with other projects | 7 |
| 3.2. Streamline and PERSoNA | 8 |
| 4. Aims and Objectives | 9 |
| 5. Methodology | 9 |
| 5.1. Advocacy and dissemination | 10 |
| 5.2. Collaborative practice | 12 |
| 6. Implementation | 12 |
| 6.1. Repurposing intraLibrary | 12 |
| 6.1.1. Open Access | 13 |
| 6.1.2. Functionality of the IRISS interface | 13 |
| 6.1.3. Results | 13 |
| 6.2. Additional functionality required | 16 |
| 6.2.1. Search | 16 |
| 6.2.2. Appropriate format for results | 16 |
| 6.3. Developing the SRU interface | 17 |
| 6.3.1. Research material | 17 |
| 6.3.2. Learning objects | 17 |
| 6.3.3. Collections | 17 |
| 6.3.4. Metadata for research material | 17 |
| 6.3.5. SRU and metadata | 18 |
| 6.4. Ensuring research is discoverable | 19 |
| 6.4.1. OAI-PMH | 19 |
| 6.4.2. Facilitating Google search: XML sitemaps | 19 |
| 6.4.3. Cover sheet for full text items | 19 |
| 6.5. Workflows | 20 |
| 7. Outputs and Results | 21 |
| 7.1. Questionnaire delivered at the Technology and Learning day in June | 21 |
| 7.2. Questionnaire delivered by postgraduate student Beth Hall (Helen Elizabeth Hall) as part of her research for MSc in Information Studies | 22 |
| 7.3. The Repository open search SRU interface | 26 |
| 7.4. Content | 27 |
| 7.4. Project blog | 27 |
| 8. Outcomes | 28 |
| 9. Conclusions | 30 |
| 10. Implications | 30 |
| 10.1. Ongoing repository development at Leeds Met | 30 |
| 10.1.1. Development of SRU | 31 |
| 10.1.2. Workflows | 31 |
| 10.2. Development of intraLibrary | 32 |
| 10.3. A stakeholder driven approach | 33 |
| 11. Recommendations | 34 |
| 12. References | 35 |

Note on appendices – There are no appendices submitted as part of this report; all supplementary documentation is available from the project blog:

<http://repositorynews.wordpress.com/documentation/>.

Table of figures

| | |
|---|----|
| Figure 1: A screenshot of the SRU interface developed by IRISS demonstrating auto-suggest functionality from user input: | 1 |
| Figure 2: <i>Results of the search are then returned in an easy to read format:</i> | 1 |
| Figure 3: <i>Full details:</i> | 1 |
| Figure 4: The 56 respondents were asked “Are you aware that Leeds Met is developing an open access institutional repository?” Answers are divided by length-of-service of the respondents. ... | 23 |
| Figure 5: The 56 respondents were asked “Are you aware that Leeds Met is developing an open access institutional repository?” Answers are divided by research discipline of the respondents. | 23 |
| Figure 6: <i>The 56 respondents were asked “Are you aware of the Open Access movement which promotes free, unrestricted access to digital, scholarly material?” Answers are divided by length-of-service of the respondents.</i> | 24 |
| Figure 7: <i>The 56 respondents were asked “Are you aware of the Open Access movement which promotes free, unrestricted access to digital, scholarly material?” Answers are divided by research discipline of the respondents.</i> | 24 |
| Figure 8: <i>A screen shot of the open search SRU interface:</i> | 1 |
| Figure 9: <i>Portion of a screen shot of the Browse for Research page:</i> | 1 |

1. Acknowledgements

Implementing an Institutional Repository for Leeds Metropolitan University was funded by the JISC Repositories and Preservation Programme - Repositories Start-up and Enhancement (Strand D).

During the project, input from several user groups and supporting staff was of great value and these include:

Academic staff at Leeds Metropolitan University
The TEL team at Leeds Metropolitan University
The Streamline project team
JISC Emerge community

The project team would also like to acknowledge the support and enthusiasm of our software provider, Intrallect as well as the Repositories Support Project and Web2Rights for their expert advice throughout the project.

Thanks also to Beth Hall who used our project as a case study for her MSc in information studies; some of her results are presented as a formal element of the project.

2. Executive Summary

Leeds Met has been funded under the Repositories Start Up programme to establish an institutional repository. The project began with an institutional needs analysis, which resulted in the starting point for the population of the repository to be based on research outputs, with a clear mandate that the software platform should be extensible to support outputs of assessment, learning and teaching, as well as a range of other materials. The project team led the procurement of a suitable software platform, IntraLibrary, and this was implemented in June 2008. Up until this point, the project team had concentrated the bulk of their activities on the procurement and also on advocacy activities. The result of the latter has been that the repository already has a high profile within the University.

Since the commissioning of IntraLibrary in the Summer of 2008, the project team has concentrated on working with the project consultancy team to agree appropriate policies and procedures. The team has also worked closely with Intrallect, and adapted open source applications developed by other JISC projects, to configure IntraLibrary to function more effectively as an open access research repository.

Procurement of full text content has followed the pattern exhibited elsewhere in the sector. A number of full text articles are available within the Repository. However, to date the bulk of contributions have been in citation format. The University Research Office is very supportive of the project, and are convinced of the potential of the Repository to raise the profile of research at the University. It is hoped that this commitment, combined with the already high profile of the Repository, will lead to higher levels of full text deposit.

The next development of the Repository will be to store and make accessible learning objects. Again, a number of learning objects are already held in the repository, and in addition the University's Pro Vice Chancellor for Assessment Learning and Teaching and the Dean of Partnerships for Students are both supportive of a drive to populate the Repository with existing content held in the University's VLE.

A number of other uses for the Repository have already been identified and will be implemented in due course.

3. Background

In recent years, Institutional Repositories (IRs) have become an established technology at universities. At Leeds Met it was recognised that a Repository had the potential to meet a number of institutional needs:

- An open access research repository
- An assessment, learning and teaching repository for learning objects, assessment objects
- A showcase for students work
- A repository of digital images of heritage collections
- A managed environment for the deposit of internal documents

Following a successful bid to JISC in the spring of 2006 an institutional needs analysis at Leeds Metropolitan University recommended that the initial focus of an IR for Leeds Met should be an Open Access repository for the university's research output. Discussions with the University Research Office, and with the then Pro-Vice Chancellor for Research, Professor Sheila Scraton, gave a clear steer in the direction of a research repository, particularly as this aligned with the University's stated aim of increasing its research profile and to make an improved submission to the Research Assessment Exercise in 2007. Furthermore, in 2007 the University celebrated its centenary (based on the founding of the original college from which the institution later developed) and as part of the centennial celebrations 100 PhD students were recruited and given University bursaries. At a meeting of the University's Research Sub-Committee in May 2006, Wendy Luker presented the concept of open access publishing and the purpose of the Repository to the committee, and the project received the support of the committee.

It was also important that the repository would have the capacity to fulfil its broader potential; in the words of Clifford Lynch (2003) "[A] mature and fully realised institutional repository will contain the intellectual works of faculty and students - both research and teaching materials - and also documentation of the activities of the institution itself in the form of records of events and performance and of the ongoing intellectual life of the institution."

Though the concept of a central system to manage disparate resources in this way has been implicit within the sector for some years, the technology has tended to focus on Open Access to research, with the two most widely used software platforms being EPrints, developed at the University of Southampton in 2000, and DSpace, developed at MIT in 2002; early versions of both platforms were primarily designed to manage text based resources (though subsequent versions of EPrints and DSpace can manage a wide range of digital file formats.)

There are a number of historical, cultural and technical reasons that the IR has developed primarily as a tool to disseminate research and the inception of the Open Access movement can be traced back to two pre-web, internet based scholarly resources; in 1990 Stevan Harnad introduced *Psycoloquy*, the first peer-reviewed scientific journal on the internet and in 1991 Paul Ginsparg developed arXiv, a repository of 'preprints' in physics at Los Alamos National Laboratory. In addition, in 1994, soon after the web had been invented, Harnad posted a "subversive proposal" that all academics should follow the example of arXiv and "self-archive" their work so that it could be openly accessible, free from increasingly expensive journal subscription costs; in the 1970s journal prices began to rise faster than inflation, becoming extortionate by the mid-1980s and having a negative impact on serials collections in libraries who could afford to subscribe to fewer and fewer of the expensive journals; the so-called "serials pricing crisis" (Guedon, 2001). Moreover, journals do not generally pay authors for their articles and the majority of scholars publish their research in peer-reviewed journals not for financial, but for professional gain (Yiotis, 2005).

The primary purpose of Open Access to research is to remove barriers to access and the benefits of wider dissemination of their research are fairly clear cut to its authors in the form of increased research impact and potential career advancement, so it has been relatively easy to persuade the

academic community of the value of the model¹. However, this is not necessarily true for other types of academic content and authors may be less willing to relinquish control of teaching and learning resources they have developed. Copyright is also fraught with potential problems and the ease with which multi-media can be shared over the internet and the lack of specialised copyright knowledge amongst practitioners coupled with the fact that teaching materials can comprise “nested” multi-media from a number of sources means that the issues are not easily resolvable².

Also key to the repository model of research dissemination is for the content of individual repositories to be discoverable externally and later developments focussed on protocols to facilitate a "global system of distributed, interoperable repositories" (Crow 2002). The approach was to harvest metadata via the Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH) which is based on Dublin Core metadata which comprises 16 fields and can be extended (qualified) to incorporate, for example, the bibliographic information associated with journal material. Learning Objects however, potentially require a number and type of fields for which qualified DC is not practical and alternative schemas have been developed, most notably IEEE UK LOM.

The cultural and historical precedents and resulting technical foci conspired such that the first generation of Institutional Repositories were geared towards textual material and it is Open Access to research where there has been the most rapid development; institutions and organisations that wished to manage learning and teaching materials and other more complex digital objects have generally had to look to alternative systems and there is ongoing debate around whether the issues facing OA research archives and Learning Object repositories are so different that they should not be managed in the same systems. However, LO repositories can benefit from the culture of openness and sharing exemplified by OA archives of research and there is certainly scope for complementary technology as the zeitgeist shifts towards Open Access to a wider range of educational resources. There are also potential administrative benefits for this integrated approach.

3.1. Integration with other projects

Implementing an Institutional Repository for Leeds Met is one of three JISC funded repository projects that have been running in parallel at Leeds Met over the past 2 years:

- Streamline is looking at the work flow associated with the use of learning object repositories and developing a suite of tools and practices that will reduce the administrative impact of this on teaching and research staff and was funded by JISC under the Users and Innovation programme: e-administration strand.
- PERSoNA (Personal Engagement with Repositories through Social Networking Applications) is examining the use of social networking applications around the repository to promote its use and was funded by JISC under the Users and Innovation programme: Personalising Technologies strand.

As part of the background context to *Implementing an Institutional Repository for Leeds Met* it is important to outline how work across these three discrete but related projects has informed development of and impacted upon methodology and implementation (discussed in detail in the relevant sections of this report.)

During the progress of the project, the project team have also benefited from support from and liaison with a number of other JISC funded projects, including the Repositories Support Project, information from the various Sherpa projects (RoMEO and JULIET) and Web2Rights.

¹ Though academics may have been persuaded of the value of Open Access to research in principle, in practice Institutional Repositories of Open Access research continue to be under populated and under utilised. This is due, in part, to the perceived administrative overheads for academic staff and is one of the areas investigated by the Streamline and PERSoNA projects (see below).

² Copyright issues remain complex for research material but there is a far greater understanding of the issues involved including appropriate negotiation with academic publishers who generally own the copyright, due, in large part to the work undertaken by the JISC funded SHERPA/RoMEO project - <http://www.sherpa.ac.uk/romeo/>

3.2. Streamline and PERSoNA

Repositories have the potential to enhance productivity and quality through effective management of assets relating to learning and teaching, research, policy and decision making. However unless their use is integrated into existing work practices there is a danger that they will increase rather than decrease the overall workload of staff and ultimately therefore be under-utilised. Moreover, the relatively low take up of the use of repositories is a result of attitudinal issues as well as technical ones. Streamline consulted potential repository users as part of its activities in supporting workflow and found that even staff who are willing to engage have serious concerns about loss of control and ownership of their outputs. The need to manage their own resources, while still sharing them with others, is therefore important. Building a community of trust around repository use would alleviate some of these issues.

The activities of managing assets and resource discovery, though pivotal to the successful deployment of repositories to support research and learning, are viewed by staff as adding to their administrative overload. The Streamline and PERSoNA projects have investigated ways of integrating these essential functions into the tools and work practices that research and teaching staff already use in order to provide personal e-administration support for the use of repositories.

The aim of Streamline has been to develop and evaluate tools to integrate the necessary repository-related functions of creating metadata for resources, resource discovery. The project has focussed on learning and teaching resources, in order to make the scope of the project manageable, but will develop generic tools which are applicable to other asset management areas.

In addition, one component of Streamline was to develop customised tools to support individualised work flows and it was from here that the PERSoNA project had its genesis; this project was designed to focus on the individual's interaction with the repository at every stage, investigating how social networking tools might facilitate the stakeholders' connection between the necessary institutional functions of the repository and the individual's own use and exploitation of it and to facilitate intuitive, communal interaction with the repository both for placing material into and for retrieving material from the repository from a variety of appropriate locations on the web.

The full project reports to JISC for Streamline and PERSoNA are available separately.

4. Aims and Objectives

In the project plan, the aims and objectives were outlined as follows:

The broad aim of the project is to establish a repository for Leeds Metropolitan University, based on open standards, which not only meets the needs of the key stakeholders at Leeds Met, but also makes a valuable addition to the open access community. Leeds Met will be making materials available which would otherwise not be promoted to the open access community, and at the same time the awareness of staff and students at Leeds Met and across the wider Regional University Network will be raised in respect of the value of open access publishing.

The objectives are:

- to raise the profile of open access publishing, and access to open access materials, through presentations to key stakeholder groups and training events
- to conduct an institutional needs analysis to establish the most appropriate (initial) focus for a Leeds Met institutional repository
- to conduct a market analysis of the most appropriate software solutions and, based on the institutional analysis, usability testing, software and hardware needs and support issues, and stakeholder input, choose a software / hardware solution for the Leeds Met repository
- to establish workflows for the ingest of content to the repository, including metadata standards
- to commission a representative body of initial content, which will promote the use and benefits of the repository, and in its turn generate new content
- to establish the Leeds Met repository as a standard element of the workflow of those generating research outputs and or learning / assessment objects, or any other content which is identified at the outset as being the primary initial focus of the Leeds Met repository

These aims and objectives remained constant throughout the project. In some cases, decisions had to be made on the best way of achieving the objective longer term. An example is the objective to achieve a representative body of initial content, combined with the overall aim to make a valuable contribution to the open access community. In the context of research this could be seen to comprise full text material only or include citation of material that we do not have copyright permission to make available as full text (i.e. bibliographic reference only). After appropriate discussion with stakeholders, including the University Research Office, it became apparent that the inclusion of citations in the repository, as well as full text wherever possible, would potentially provide a centralised database for the URO that could be exploited in a variety of ways; particularly considering the increasing importance of citation data/bibliometrics mooted for the Research Excellence Framework. The judgement of the project team in this instance is that, by meeting the needs of the research community and establishing strong collaborative working in this way is more likely to lead to successful attainment of the greater objective to amass a body of full text content.

5. Methodology

The initial stage of the project focussed on an institutional needs analysis to determine what type(s) of resources the repository would most usefully manage and, throughout the early part of the project, the Project Manager met with a wide range of staff and attended key meetings at which the concept of an institutional repository was discussed, and the merits of focussing on research vis a vis learning objects and other types of material were debated. One of the outcomes of this early engagement was the formation of the Repository Consultancy Group which comprised representatives from all of the major stakeholder groups including the University Research Office; Assessment, Learning and Teaching (ALT) and the Streamline project. The disparate needs of stakeholders meant that it was extremely important from the outset that the repository solution was extensible to other types of repository content, in particular, learning objects and a rigorous market analysis was undertaken with the Repository Development Officer responsible for undertaking preliminary research into potential technology, consulting with the Streamline project team and engaging with the wider repository

community for input in order to draw up an appropriate system specification - available from <http://repositorynews.wordpress.com/documentation/>. The RDO visited numerous digital repositories on the World Wide Web and performed a preliminary assessment against the specification before drawing up a shortlist of six commercial providers of repository solutions and contacting their representatives to arrange on-site demonstrations. The possibility of using Open Source repository software and undertaking all development work in-house was also investigated though this approach was discounted due to inadequate in-house technical expertise, additional project-management overheads and the competitive pricing of the available commercial solutions³.

The six platforms short-listed were Digital Commons; EPrints; Open Repository; Digitool; IntraLibrary and Harvest Road. On site product demonstrations were held in January 2008 and were attended by members of the Repository Consultancy Group as well as additional staff from the Streamline project and Learning and Libraries Innovation as appropriate. When all products had been seen, attendees were asked to provide preliminary feedback. In addition, the research officer for the Streamline project produced a detailed evaluation document of all 6 products – available from <http://repositorynews.wordpress.com/documentation/>. Representatives of each potential solution were also asked to provide preliminary costing information to inform the decision making process.

In order to provide a meaningful comparison of the 6 platforms, the system specification was developed into a spreadsheet that could be used to score each system against a standard set of requirements - available from <http://repositorynews.wordpress.com/documentation/>. When all the relevant information had been gathered it was presented at a meeting of the Repository Consultancy group on 11th March 2008, with the aim of arriving at a decision with respect to the platform that should be used to implement the Leeds Met Repository.

At the meeting, a consensus was reached that the two strongest systems were EPrints and intraLibrary and that each system had different strengths and weaknesses with EPrints identified as better suited for research outputs and intraLibrary better suited for Learning Objects. Several specific issues were raised that required further investigation; given that the Leeds Met Repository was to be used for both research outputs and Learning Objects it was deemed necessary to postpone the decision and the meeting was adjourned pending further investigation. Information was collated that directly compared the two platforms according to agreed criteria – available from <http://repositorynews.wordpress.com/documentation/> - and the Repository Consultancy Group reconvened on the 16th April 2008.

When all the evidence had been considered the consensus reached was that intraLibrary offered the best long term solution for the disparate needs of the Leeds Met Repository though it was recognised that as a purpose built learning object repository it would require considerable developmental work to also function as an Open Access research archive.

5.1. Advocacy and dissemination

Open Access to research is an evolving paradigm and represents a considerable shift in the established academic publishing process; Open Access to a broader range of educational resources still more so.

The invention of the internet and the web has made viable Open Access to research, which had been “physically and economically impossible in the age of print, even if the copyright holder wanted it” (Suber). However any paradigm shift is likely to take time to evolve and Open Access, to research and other materials, is no exception, especially given that academia, perhaps, tends to subscribe rather strongly to established tradition. Taking this into account and learning from the experience of the repository and Open Access communities, a key aspect of the original bid necessarily focussed on the need to “raise the profile of open access publishing, and access to open access materials”. However, due to the development and customisation required, a functioning Open Access archive

³ Two of the solutions investigated were in fact 'hybrid' solutions using Open Source software but underpinned by a third-party commercial business model; BioMed Central's Open Repository (DSpace) and EPrints Services (EPrints)

was not available to demonstrate to stakeholders until relatively late in the project. In any case, the project's dual focus on OA research and learning objects has meant that advocacy has necessarily focussed on the wider remit of the repository and a pragmatic approach was adopted whereby the RDO provided general information before responding flexibly to the interests and requirements of specific stakeholders. This stakeholder driven approach is one of the main reasons that has ensured a high profile for the repository though there have been both benefits and drawbacks to the approach which are reported on more fully in the *Implications* section of this report.

Early activity included liaison with the research administrators of specific faculties with the RDO attending numerous faculty based research committee meetings to advocate the use of the repository and the benefits of Open Access publishing, as well as engaging with Library staff to promote their role in supporting the project. A formal presentation was delivered as part of the Carnegie Research Institute seminar series (slides available from <http://repositorynews.files.wordpress.com/2008/04/draft-presentation.ppt>).

The Leeds Met Staff Development Festival in September was another key forum that generated interest and awareness around the repository and Open Access. A high profile presence throughout the festival fortnight meant that the RDO was able to engage face to face with academic and teaching staff in a less formal context and this was invaluable in engaging with people on a personal level. Publicity material was widely disseminated both in paper format and by pointing people at the project blog - <http://repositorynews.wordpress.com/>:

- An introduction to the project and to IRs specifying our dual remit for the Leeds Met repository. (Available at <http://repositorynews.files.wordpress.com/2008/08/the-repository.doc>)
- "Open Access: What's in it for you?" emphasising the evidence that OA increases citation. (Available at <http://repositorynews.files.wordpress.com/2008/08/open-access.doc>)
- A simplified flowchart of the (self)-archiving process that aimed to clarify issues of copyright. (Available at <http://repositorynews.files.wordpress.com/2008/08/copyright.doc>)

The project blog, maintained by the RDO, became an important mechanism, both for project management activity and as a point of contact with the institutional and wider communities. The profile of the blog has continued to grow throughout the project and regularly exceeds 250 hits per month. Any contact with staff was also taken as an opportunity to generate Frequently Asked Questions which were initially posted on the blog and the project web page before eventually being included on the repository interface itself.

At Leeds Met there is a strong culture of reflective practice with staff encouraged to submit 200 word reflections for daily update on the institutional website. Categories for reflections include *Research, Assessment, Learning and Teaching* and *Ethical*; the repository team took full advantage of this established practice to engage with the university community, publishing several reflections throughout the project. In addition, the RDO submitted a paper on Open Access to the ALT journal which is published internally at Leeds Met and subject to formal peer review. The article has been accepted for publication. Reflections and a preprint of the paper are available from <http://repositorynews.wordpress.com/documentation/>.

The original project plan specified that a key factor to evaluate would be "Raised profile of open access publishing within the University" and that this would be assessed by a "Questionnaire to be circulated amongst the [research] community". In the first instance a questionnaire, comprising input from Streamline and PERSoNA was distributed at the Technology and Learning day in June – available from <http://repositorynews.wordpress.com/documentation/>. Many researchers will recognise the difficulty in persuading busy people to complete a questionnaire and we were pleased to generate 20 respondents at this event – the results for this, admittedly small, sample are discussed in the *Outputs and Results* section of this report. An opportunity then arose for collaboration with a postgraduate student undertaking an MSc in Information Management and who wished to use the Leeds Met repository-in-development as a case study for her dissertation "Analysis of the opinions and use of open access repositories by researchers in different disciplines; with specific focus on the development of a new institutional repository at Leeds Metropolitan University". Data collection included a questionnaire, adapted and extended from that developed by the repository team, and

follow up interviews with research staff; 56 respondents to the questionnaire and 11 follow up interviews represents a more valid sample and has the additional benefit of greater independence from the project team – analysis is presented as a formal element of project – see *Outputs and Results*. The full dissertation is available from <http://repositorynews.wordpress.com/documentation/>.

5.2. Collaborative practice

A recurring discussion within the community is that repositories are often under resourced, especially by institutions themselves who potentially stand to benefit most from the technology. There is also ongoing debate around how and where a repository should most effectively be managed within an institution with the library being the usual locus, which, though appropriate for Open Access to research, *may* be less so for more complex multi-media and reusable learning objects. There is certainly scope to explore how repositories can be resourced and managed on a sustainable basis and effective collaborative practice has the potential to more effectively integrate repositories both into institutions (and/or their libraries) and the national infrastructure. Leeds Met has a very strong tradition in information science with a highly regarded CILIP accredited Masters programme in Information Management; the library itself has an excellent reputation both within the institution and within the Higher Education sector and we have been able to build on these institutional foundations throughout the project. In addition, as the beneficiary of JISC funding for three separate but related repository projects, Leeds Met has been in the fortunate position of having considerable resources to develop and promote our repository and the impact on awareness and the broader institutional culture should not be underestimated.

Streamline was chronologically the first JISC project to commence with a project team comprising a range of academic and technical staff including teaching fellows and members of the Technology Enhanced Learning team. Implementing an Institutional Repository for Leeds Met and PERSoNA were funded quite separately from Streamline, however, the Repository Development Officer is also the project officer for the PERSoNA project (contracted 0.5 and 0.4 FTE respectively) which has maximised development for both projects. Moreover, PERSoNA was formally integrated into the monthly Streamline project meeting in June 2008 to facilitate effective collaboration amongst people working on discrete but related projects and to become consolidated as a team. In addition, the expertise and contacts within Streamline have been invaluable in raising the profile of the repository amongst the wider academic community both in the context of Open Access to research and reusable learning objects.

Other areas of collaboration include the close involvement of a postgraduate student as part of her research for an MSc in Information Management – as discussed in the previous sub-section.

6. Implementation

As part of the initial evaluation, intraLibrary was costed both on the basis of hosting the software in-house, or having it hosted externally by Intrallect; when all costs were taken into account – which included, for example, server hardware and a database administrator - it emerged that the most economic approach was for Intrallect to host the system for us

Given the delays in selecting the software, this was also expedient to ensure that intraLibrary was implemented and usable as quickly as possible; if appropriate, there is also the option to bring the software in house in the future.

6.1. Repurposing intraLibrary

intraLibrary is designed as a Learning Object repository and initial development work necessarily focussed on ensuring the software was also fit for purpose as an Open Access research archive. The process proved to be a steep learning curve.

Two over-riding issues were identified as follows:

- Open Access: intraLibrary is designed to be accessible by authenticated users whereas Open Access to research, by definition, requires unauthenticated access on the public internet.
- Metadata: intraLibrary uses the IEEE LOM standard rather than Dublin Core which, based on current practice, is the most appropriate metadata schema for Open Access research material.

6.1.1. Open Access

Statistics indicate that the vast majority of traffic will arrive at an OA repository via a search engine; intraLibrary already facilitates Open Access in this context via a public URL which can be linked to directly - on the open web and with no need for authentication - however, it is still considered important to integrate an accessible search interface with extensive functionality to ensure the repository has a clear identity / web-presence and for demonstration purposes to stakeholders.

The solution to the authentication issue and providing Open Access to browse the research collection in this way was to use a separate, web based interface to query intraLibrary using SRU (Search and retrieve by URL), a standard search protocol utilizing CQL (Common Query Language).

Intrallect developed an SRU client as part of the CD-LOR project⁴, which was subsequently developed further into a more sophisticated interface by the Institute for Research and Innovation in Social Services (IRISS). The code for this version is available under an Open Source licence at <http://code.google.com/p/sruopensearch/>; it was decided that this version would be installed on a Leeds Met server and adapted to our requirements.

Technical development work would be undertaken by an in-house web developer who was seconded to the project in June 2008.

6.1.2. Functionality of the IRISS interface

The IRISS interface comprises a search box into which a user can input a search term. Search terms are restricted to only those terms recognised by the database and the interface auto-suggests according to which specific terms may be extrapolated from a partial user input; for example, inputting the letters HE will auto-suggest any terms beginning with or that contain those letters (helplines, fathers, family therapy etc); suggested terms are refined as the user inputs more letters (e.g. HEL will refine suggested terms to helplines, sheltered housing, shelters etc.). It does not support Boolean operators (AND, OR, NOT) and there is no facility for a more refined search i.e. by specific metadata fields.

6.1.3. Results

The default configuration defined by the software developers IRISS comprised a sub-set of metadata suited to their specific requirements; results are returned as a list of hyperlinked titles that point to the resource, a brief description of the resource and the option to expand "full details" which comprises Creator(s); Publisher; Type; Subject(s) and Copyright.

⁴ See <http://www.academy.gcal.ac.uk/cd-lor/> for more information

Figure 1: A screenshot of the SRU interface developed by IRISS demonstrating auto-suggest functionality from user input:

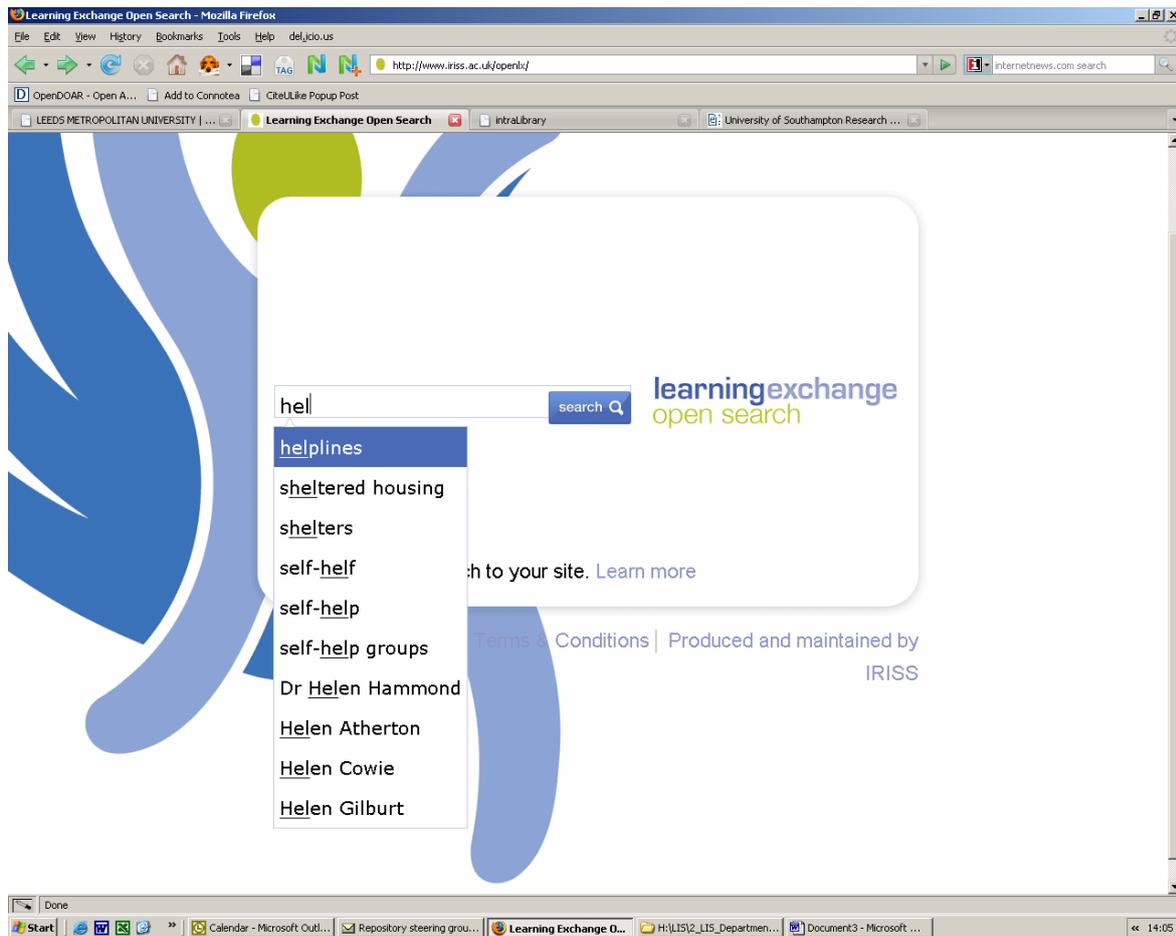


Figure 2: Results of the search are then returned in an easy to read format:

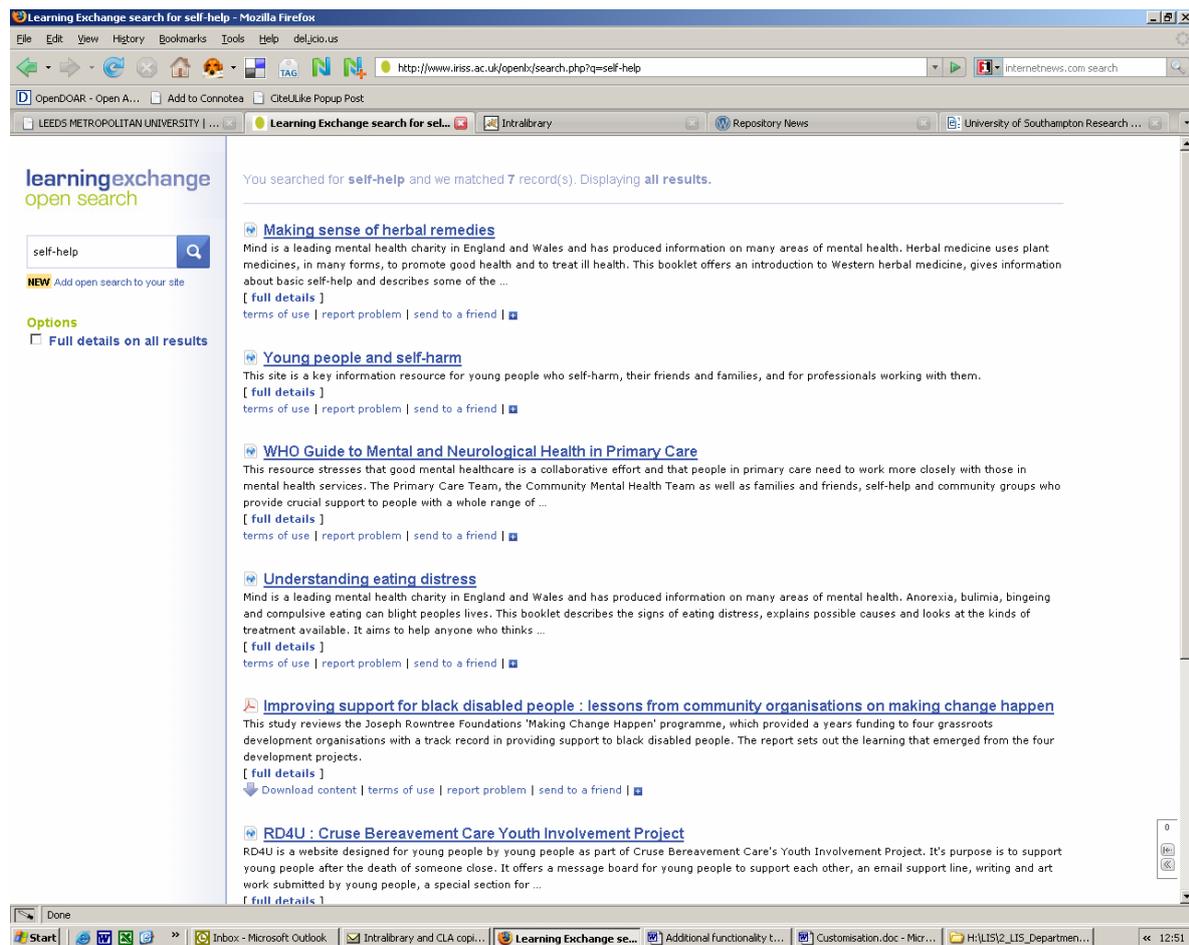


Figure 3: Full details:

[WHO Guide to Mental and Neurological Health in Primary Care](#)

This resource stresses that good mental healthcare is a collaborative effort and that people in primary care need to work more closely with those in mental health services. The Primary Care Team, the Community Mental Health Team as well as families and friends, self-help and community groups who provide crucial support to people with a whole range of mental disorders need to talk to one another, respect each other's contribution and jointly agree who will provide which service to whom. Mental health provision has been dogged, perhaps more than any other area of healthcare, by differences in how we think about mental health and the words we use. This makes it hard for different professional groups and non-professionals to talk to each other. This web site aims to improve this problem and is a resource that can be used in a number of ways. It can be used by an individual practitioner in the care of his or her patients; it can also be used by a primary care team or a primary care organization (or local health group) to review, jointly with mental health teams, the service they provide, identifying gaps and training needs or developing locally appropriate, shared criteria for referral to specialist services. It is hoped that this web site improves communication and collaboration between all who have a stake in the provision of good primary mental healthcare.

| About this resource | |
|---------------------|--|
| Creator(s) | WHO Collaborating Centre for Research and Training for Mental Health, World Health Organisation |
| Publisher(s) | Royal Society of Medicine Press Ltd |
| Type(s) | Narrative Text, website, information resource |
| Subject(s) | primary care, mental health care, interprofessional relations, collaboration, health professionals, people with mental health problems, mental health problems, neurology, learning disabilities, diagnosis, psychotherapy |
| Copyright(s) | Copyright: http://www.mentalneurologicalprimarycare.org/copyright_information.asp |

[partial details]

[terms of use](#) | [report problem](#) | [send to a friend](#) |

6.2. Additional functionality required

There are several dedicated software platforms that are used to implement Open Access repositories. The most widely used in the UK is EPrints (67 installations in UK; source <http://www.opendoar.org/>; March 2009); globally it is DSpace (392 installations worldwide; source <http://www.opendoar.org/>; March 2009.)

A range of Open Access repositories using both EPrints and DSpace as well as a number of other platforms were examined in order to identify:

- the type of functionality that should be developed into the SRU search interface
- the appropriate format for search results

6.2.1. Search

The following elements were identified to be incorporated into the search interface:

- I. Search box that supports Boolean operators and returns results in an appropriate format.
- II. Browse functionality that corresponds to internal organisational structure and that returns results in an appropriate format.
- III. Advanced search that allows users to search by cross-referencing multiple metadata fields and that allows results to be differentially ordered by appropriate criteria and that returns results in an appropriate format.
- IV. Integration of appropriate Help and Information
- V. Link to most recent additions to the repository.
- VI. RSS functionality that allows users to easily subscribe to content updates.

6.2.2. Appropriate format for results

The following metadata was identified to be displayed in search results:

- Title
- Abstract
- Journal ISSN
- Digital Object Identifier (DOI)
- Journal reference
- Item type
- Copyright information
- Published status
- Published URL
- Refereed status
- Classification
- Date of deposit
- Deposited by
- Author(s)
- Subject(s)

6.3. Developing the SRU interface

It soon became apparent that developing the required functionality was contingent upon appropriate configuration of intraLibrary itself. For example, it was clear that research material and learning objects would require different organisational structures and metadata and that, in turn, this would impact on how the SRU interface would search and display results

NB. The priority in terms of functionality is to be able to browse the research collection; external public discovery of learning objects is contingent on ongoing developments in institutional policy.

6.3.1. Research material

Following review of other repositories of research content and appropriate consultation with colleagues from the library it was proposed that Library of Congress Classification would be appropriate for our needs. However, the full system is complex with many more levels of specificity than we require and it was proposed that we use the top two levels of the classification only; we will also have the flexibility to extend the classification if necessary.

6.3.2. Learning objects

The Joint Academic Coding System (JACS) subject hierarchy is generally used by UK institutions to identify the subject matter of programmes and modules which potentially makes it more suitable for Learning Objects - it is the organisational taxonomy used by the national Learning Object repository JORUM and it was also considered suitable for our requirements.

N.B. Resources in intraLibrary may be categorised against multiple classifications and we may wish to add other classification systems in the future e.g. Medical Subject Headings (MESH)

6.3.3. Collections

Collections in intraLibrary serve a variety of functions; the main one in the context of Open Access to research is that it is the means by which the administrator defines whether or not resources are discoverable by external systems or whether they can only be discovered by an authenticated user of the Leeds Met Repository. The first scenario is necessary for Open Access research material; the second scenario is likely to be necessary for *some* learning objects (though we may also want some learning objects to be searchable and discoverable externally.)

In addition, collections provide an expedient facility to organise research by faculty; although it was agreed that the main organisational structure should not be based on Leeds Met faculties, it is still useful to be able to present content in this way which can easily be achieved by the use of collections and which can easily be renamed in the event of faculty name change. Moreover, resources can be stored in multiple collections and easily moved between collections.

6.3.4. Metadata for research material

intraLibrary uses IEEE LOM (Learning Object Metadata); the administrator can define multiple *application profiles* (metadata schema) which can incorporate subsets of LOM that may be differentially applied to collections based on content type.

Based on current practice, the most appropriate metadata schema for Open Access research material is Dublin Core; intraLibrary uses the IEEE LOM standard which can be mapped onto simple DC in a fairly straightforward manner (Title=Title; Description=Description; Subject=Keyword etc). For a full Crosswalk Between IEEE LOM and Simple Dublin Core

Metadata see <http://repositorynews.wordpress.com/documentation/>. Issues arise, however, when research metadata needs to be extended which would normally be achieved by using qualified Dublin Core - to incorporate bibliographic citation metadata, for example. Version 3.0 of the intraLibrary software does incorporate additional fields for this research specific metadata which partially resolves the issue (though there is still no field for journal ISSN); it is also possible to utilise additional instances of LOM fields to accommodate qualified DC fields, for example an additional instance of the Description field is being used for ISSN.

The RDO liaised with Intrallect to define an appropriate metadata template for research material that could be differentially applied to collections within intraLibrary based on content type. Metadata templates in intraLibrary are defined in XML – the XML file for the research metadata template is available from <http://repositorynews.wordpress.com/documentation/>.

6.3.5. SRU and metadata

The SRU provided as part of IntraLibrary receives a request via a URL query and returns a set of results formatted as XML. This XML contains the metadata for the resulting set of records. By default the record metadata is presented using the Dublin Core metadata format, however, this omits various elements of the metadata that we ideally require to display in the search results (e.g. metadata detailing bibliographic citation information). It is possible to force the SRU to return metadata results using the LOM format by adding an extra parameter to the URL query sent to the SRU. This exposes additional metadata fields that can be displayed by the SRU interface. For most metadata fields in the XML, the data are given in a plain text format. Fields that describe a person, authors for example, are represented as a vCard, a standardised file format for electronic business cards; each vCard is itself represented in plain text.

In the context of the Leeds Met repository it would be necessary to configure the IRISS Open Search interface to return results in a format appropriate to research material (and, potentially, learning objects). It would also need to reflect the research metadata template (and also potentially other metadata templates defined for learning objects).

A further complication is that in order to use the SRU interface to search both the research collection and learning objects, it would need to differentiate types of content and apply an appropriate template. Moreover, the user would need a sophisticated advanced search facility to appropriately search for material in such a context. For these reasons, development work focussed on research; a more practical solution may be to implement a second SRU interface to browse learning objects - though this would still require the respective interfaces to differentiate types of material. These issues are explored in more detail in the *Implications* section of this report.

As the IRISS open search interface was developed to work with Dublin Core, it required significant modification in order to handle data represented by LOM. Identifying how to extract the appropriate data from within the XML was achieved by analysing the XML results for various different records in order to determine the precise location of metadata fields in the XML structure. This process involved a certain amount of trial and error; certain specific issues were identified only when a search result exhibited unusual behaviour.

6.4. Ensuring research is discoverable

Key to the repository model of research dissemination is for the content of individual repositories to be discoverable externally. There are two main ways for this to occur, they can either be indexed by internet search engines like Google or via the Open Access Initiative Protocol for Metadata Harvesting (OAI-PMH)

6.4.1. OAI-PMH

The OAI-PMH provides an application-independent interoperability framework based on metadata harvesting⁵. It was important that the chosen repository platform supported the OAI-PMH and intraLibrary does indeed support the protocol which exposes XML formatted Dublin Core metadata over HTTP such that it can be harvested by third party service providers. For example, OAster -<http://www.oaister.org/> - is "a union catalog (sic) of digital resources" (OAster, 2009) that provides access to digital resources by harvesting descriptive metadata using OAI-PMH.

It is necessary to register with the Open Archives Initiative as a data provider at <http://www.openarchives.org/data/registerasprovider.html> in order to:

- Provide a publicly accessible list of OAI conformant repositories, making it easy for service providers to discover repositories from which metadata can be harvested
- Provide a mechanism for data providers to ensure their conformance with the OAI-PMH specification.
- Provide a means for the OAI to monitor use of the protocol and plan future activities and strategies.

6.4.2. Facilitating Google search: XML sitemaps

Google is a ubiquitous service on the modern web and the first port of call for the majority of users searching the web, though it does have recognised drawbacks for targeted resource discovery. The details of Google's search algorithm are a closely guarded secret, however, in general terms and in common with other search engines (e.g. Yahoo), it uses a computer program known as a web crawler that automatically indexes the content of HTML pages. Though web crawlers can index automatically, it is not always particularly efficient and, for example, can be hampered by an authentication gateway. Sitemaps are one method of informing search engines about pages that are available for crawling so that web crawlers can index a site more effectively.

Until 2008, Google did support sitemaps using OAI-PMH but have since withdrawn this and now support only the standard XML format. Intrallect have therefore developed a software tool that converts OAI-PMH output to an appropriate XML format. A preliminary sitemap has been generated and registered using Google's webmaster tools. However, work is ongoing and when developmental issues have been resolved, Intrallect will implement a program to automatically update the sitemap on a periodic basis.

6.4.3. Cover sheet for full text items

A further issue is that a query-return from a search engine, in linking to the intraLibrary public URL, will link directly to the resource itself – i.e. a PDF of a research article will open immediately in the browser window. When facilitating Open Access to research this is undesirable for several reasons; it is important to provide context and basic information (abstract, copyright info, whether the paper has been refereed); indeed, there will often be a legal requirement to provide copyright

⁵ See <http://www.openarchives.org/pmh/> for more information

information with many publishers also stipulating that there must be a link to the published version of the paper.

One possibility would be to incorporate a “landing screen” and it might be possible to embed a link to the PDF into a HTML template and have this template returned at the public URL – this is discussed in more detail in the *Implications* section of this report. This is how the majority of Open Access repositories of research work, including both EPrints and DSpace.

Evidence suggests, however, that, in any case, users will often circumvent a HTML “landing screen” and link directly to a PDF – for example, this scenario will arise when a search engine returns results for both the HTML landing screen and the PDF with users generally sophisticated enough to recognise the resource itself (the PDF) and follow the appropriate link.

For these reasons, and in line with consensus amongst the repository community, a cover sheet that comprises contextual and statutory information is appended to full text articles. See <http://repositorynews.wordpress.com/documentation/> for an example cover sheet.

6.5. Workflows

One of the biggest challenges in promoting Open Access to research is establishing an efficient workflow and the model promoted by Stevan Harnard in his “subversive proposal” was that academic authors “self-archive” and undertake administrative duties themselves to upload a copy of their own research paper to an appropriate server and apply suitable metadata. In the case of arXiv, developed by Paul Ginsparg in 1991, this practice had arisen spontaneously as physicists sought to rapidly disseminate pre-prints of their work and proponents of Open Access like Harnard expected the benefits of Open Access would be self-evident to academics and that self-archiving would naturally be adopted throughout academia. However, as OA and self-archiving were promoted throughout the 1990s and as the necessary technology became standardised in the new millennium, it became apparent that this was not the case and that, generally speaking, academics were reluctant to change their working practices and continued to disseminate, primarily post-prints, of their research through the established publishing infrastructure which often meant they waived their copyright and placed their work behind a subscription barrier.

In the context of learning object repositories, the relevant issues have not been explored to the same extent but, arguably, they are directly comparable; academic staff, for example, do not relish the administrative burden of applying extensive metadata to their learning resources to ensure they are discoverable in a repository. A crucial element of repository development, therefore, must be to make the process of deposit in an IR as quick and easy as possible, both for research and learning objects. As previously discussed in this report, the Streamline and PERSoNA projects at Leeds Met have been exploring these issues in detail. However, at the end of this repository start-up project, though a great deal has been learned about workflow (see separate project reports for Streamline and PERSoNA), we are not yet in a position to fully integrate the three discrete project outcomes into coherent workflows for both research and learning objects; in the context of assembling a representative body of content and due, in part, to the technical development work required to repurpose intraLibrary to also function as an OA research archive/citation database, it has been necessary to implement a fully mediated workflow whereby the RDO has added citation information for research outputs and, wherever possible, uploaded the corresponding full text items. This approach has enabled ongoing refinement of a library mediated workflow and the RDO has begun to work with colleagues from the Bibliographic Services Unit whose involvement will contribute to ensuring the sustainability of the project. Moreover, assembling a body of research material in this way has, in the first instance provided a collection that can be searched by stakeholders, thereby raising awareness and building inertia which is essential to long term sustainability.

7. Outputs and Results

In the context of the Leeds Met repository and Open Access to research it is useful to consider the experience across the sector as a whole, which has found that, in spite of the acceptance of Open Access as a desirable goal by academic institutions themselves, lack of commitment by their respective research communities has resulted in Institutional Repositories of Open Access research remaining under populated and under utilised (Swan, 2006). It is accepted that coordinated and sustained advocacy to an academic community is essential to raising and maintaining awareness with both SHERPA and the Repositories Support Project supporting the community in this regard; however, in the context of evaluation for the SHERPA project, Markland and Brophy (2005) suggest that “raising awareness is only part of the process” and that “the difficult part is turning awareness into action” with several commentators describing the inertia amongst researchers that perpetuates the status quo (Harnad, 2006).

At Leeds Met, in common with other Institutional Repository projects, we have found that, while the academic community are supportive of Open Access and self-archiving on a conceptual level, this does not readily translate into action. An additional issue has been that technical development work has meant that there was not a functioning system until relatively late in the project which has made it difficult to demonstrate the putative benefits of an Open Access to research. Moreover, this was a start-up project and there is still a considerable amount of development work yet to be done to perfect the system. Advocacy, therefore, has focussed on building awareness at a grass-roots level particularly with the University Research Office and also within faculty; it is only now, towards the end of the project, that networks developed with the research community are coming to fruition with several major initiatives due to be implemented over the coming months.

We can also report some interesting results in terms of assessing awareness of Open Access in an institution that did not have a functioning repository for its research staff – potentially this could inform ongoing work for JISC projects in the future, other repository start-ups, for example, or the Depot, which accepts deposit of e-prints from researchers at institutions that do not currently have an Institutional Repository⁶. These results were obtained by questionnaire delivered by the repository development team and by an MSc student in the course of research for her dissertation.

7.1. Questionnaire delivered at the Technology and Learning day in June

Of the 20 respondents to our initial questionnaire delivered in June 2008, only 4 of the 20 respondents professed ignorance about the project and 14 said they have “some” (12) or “good” (2) awareness of OA; almost half (9) of people were familiar with publisher self-archiving policies (NB. This very small sample comprised attendees who may well be better informed about new technological initiatives within the university than the academic population at large.)

Just one of our respondents had actually submitted an article to an Open Access repository.

The penultimate question in the OA section of the questionnaire focussed on 6 potential benefits of Open Access and asked people to rank them each from 1 (not important) to 5 (important). For the purposes of summary here I am regarding ranks 1 and 2 (not important); rank 3 (of medium importance); 4 and 5 (important). The full spreadsheet is available from <http://repositorynews.wordpress.com/documentation/>.

a. Public have access to research they have helped fund through taxation

15 respondents considered this important; 4 respondents considered it of medium importance; 1 did not respond

b. Teachers/students have access to key resources without subscription barriers

⁶ See <http://depot.edina.ac.uk/> for more information

18 respondents considered this important; 1 respondent considered it of medium importance; 1 respondent did not consider it important

c. Maximise research impact/increase citation of your work

12 respondents considered this important; 5 respondents considered it of medium importance; 3 respondents did not consider it important

d. Increased return on investment for funding bodies

10 respondents considered this important; 8 respondents considered it of medium importance; 2 respondents did not consider it important

e. Scholars in economically disadvantaged areas of the world (eg. developing countries) have greater access to published research

17 respondents considered this important; 2 respondents considered it of medium importance; 1 respondent did not consider it important

f. Reduced economic constraints on institutional libraries that can currently afford to subscribe to a relatively small sub-set of published research

17 respondents considered this important; 2 respondents considered it of medium importance; 1 respondent did not consider it important

The final question in the OA section asked:

“In the course of your online research, how frequently do you encounter resources that you are unable to access (eg. LeedsMet does not subscribe to the resource)?”

For half of respondents (10) this is a problem “occasionally” with 7 encountering it more frequently; only 3 respondents said this was “hardly ever” a problem for them.

7.2. Questionnaire delivered by postgraduate student Beth Hall (Helen Elizabeth Hall) as part of her research for MSc in Information Studies

The original questionnaire was adapted, extended and disseminated (digitally) as part of an MSc project, 56 responses were collected. The scope of the extended questionnaire was much greater and, in addition to general awareness about Open Access, sought to elucidate disciplinary differences that may exist. Full dissertation available from <http://repositorynews.wordpress.com/documentation/>; the results and Ms Hall’s discussion are reproduced here by permission:

Of the 56 respondents to the questionnaire, 50% said they were aware that Leeds Met is developing an open access institutional repository. When you divide the respondents up by length of service (Table 3), only 22% of postgraduate students were aware that Leeds Met is developing an open access institutional repository. When you divide the respondents up by discipline (Table 4), you see that none of the three researchers in the history/anthropology/listed buildings group were aware of the development of an IR at Leeds Met. The repository development team at Leeds Met have been round the departments discussing the development of the repository and the information on the Leeds Met website including the project blog has been advertised to staff. Perhaps the Leeds Met repository development team need to focus advertising the development of the IR to postgraduate students. They definitely ought to carry out further advocacy work. I asked for further comments at the end of the questionnaire one researcher commented that they “think it’s a good idea but would like more info on how to use the system”, another said they would “like to know more about open access research” and a third said that they “had no idea that Leeds Met were doing this until you sent me this questionnaire!”

Figure 4: The 56 respondents were asked “Are you aware that Leeds Met is developing an open access institutional repository?” Answers are divided by length-of-service of the respondents.

| | Postgraduate student (n=18) | 5 years or fewer (n=10) | 6 to 10 years (n=10) | 11-15 years (n=6) | More than 15 years (n=12) |
|-----------|-----------------------------|-------------------------|----------------------|-------------------|---------------------------|
| Yes | 22 | 60 | 50 | 100 | 58 |
| No | 61 | 30 | 30 | | 25 |
| Not Sure | 6 | | | | 8 |
| No answer | 11 | 10 | 20 | | 8 |

Figure 5: The 56 respondents were asked “Are you aware that Leeds Met is developing an open access institutional repository?” Answers are divided by research discipline of the respondents.

| Answer options | Information/ linguistics/ musicology (N=19) | Social science/ politics/ business (n=7) | Health (n=7) | Ethics/ tourism/ international (n=6) | Education (n=5) | Art (n=4) | Sport (n=3) | History/ anthropology/ listed buildings (n=3) | Writing/ culture/ English/ literature (n=2) |
|----------------|---|--|--------------|--------------------------------------|-----------------|-----------|-------------|---|---|
| Yes | 47 | 57 | 57 | 33 | 80 | 50 | 33 | | 100 |
| No | 37 | 43 | 43 | 50 | 20 | | 33 | 66 | |
| Not Sure | 5 | | | 17 | | | | | |
| No Answer | 11 | | | | | 50 | 33 | 33 | |

When asked whether they were aware of the Open Access movement “which promotes free, unrestricted access to digital, scholarly material” 34% of the respondents said they were not aware and 62% answered that they had some knowledge of the open access movement. Only 4% (2 respondents) said they had good knowledge of the open access movement. When we look at these proportions according to length-of-service in research (Figure 1), we see again that it is the postgraduate students who have the largest number of respondents who say they are not aware of the open access movement. It may be that this is the case because the postgraduate students have yet to publish any papers but I would have expected that with all the literature searching they are doing for their theses that they would have accessed literature from IRs and through search engines that look for free online versions (like Google Scholar). I look at searching behaviour some more in section 4.2.4. There was also little difference according to discipline (Figure 2) with very few researchers stating that they had good knowledge of the open access movement. My results compare with what Swan and Brown (2004) found from a survey of 160 researchers who were listed as “non open-access authors” (had not published work in open access journals) that 62% were aware of OA as a general concept. This is a good proportion of researchers who are aware of OA and the Leeds Met repository team should be encouraged by the level of awareness at the University (with the caveat that my results may be somewhat biased in that those volunteering to answer my questionnaire, may be the researchers who already have a general interest in this area).

Figure 6: The 56 respondents were asked “Are you aware of the Open Access movement which promotes free, unrestricted access to digital, scholarly material?” Answers are divided by length-of-service of the respondents.

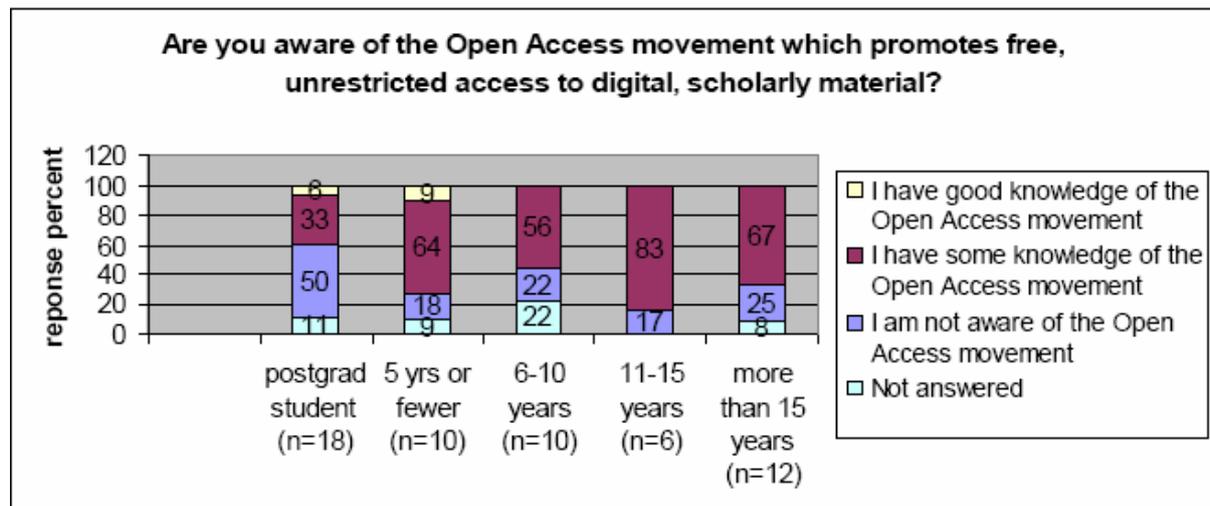


Figure 7: The 56 respondents were asked “Are you aware of the Open Access movement which promotes free, unrestricted access to digital, scholarly material?” Answers are divided by research discipline of the respondents.

| | Information/ linguistics/ musicology (N=19) | Social science/ politics/ business (n=7) | Health (n=7) | Ethics/ tourism/ international (n=6) | Education (n=5) | Art (n=4) | Sport (n=3) | History/ anthropology/ listed buildings (n=3) | Writing/ culture/ English/ literature (n=2) |
|-----------------------|---|--|--------------|--------------------------------------|-----------------|-----------|-------------|---|---|
| I am not aware | 21 | 43 | 43 | 50 | 40 | 25 | 33 | | |
| I have some knowledge | 63 | 57 | 57 | 50 | 60 | 25 | 33 | 33 | 100 |
| I have good knowledge | 5 | | | | | | | 33 | |
| Not answered | 11 | | | | | 50 | 33 | 33 | |

The 56 respondents were asked how many of the 17 following names, services or terms they were aware of: Author-pays publishing, BioMed Central, Copyright Assignment Form, Directory of Open Access Journals (DOAJ), e-prints, Institutional repository, Post-print, Pre-print, Public Library of Science (PLoS), PubMed Central, Repository, Self-archiving, SHERPA project, Subject specific repositories (e.g. ArXiv for physics, maths, computer science), The open access movement, The serials crisis, White Rose Research Online.

25% of all the respondents were aware of 0 terms, and 79% were aware of less than 1/3 of the terms, this suggests that although 62% of the respondents said they had some knowledge they may not have looked into any depth at the issue. Figure 3 shows that the members of staff who had been in research longer (over 6 years) knew somewhat more of the terms than the postgraduate students and newer staff (5 years or fewer) suggesting that the longer the length of service the more likely the researchers have of encountering the ideas of open access and institutional repositories. This may also have to do with the fact that the members of staff who are at professorial level have larger networks of contacts (in their own and different disciplines) and because they are more like to sit on policy boards, grant decision boards and act as editors for journals and may have heard more about open access through these means. This results conflicts with Nicholas et al (2005) who found that older researchers knew less about open access. As can be seen in Figure 4, more of the researchers in the Information Science/Linguistics/Musicology discipline grouping knew more of the terms than

researchers in the other discipline groupings; this is perhaps because the information science researchers are hearing about open access in literature related to their profession.

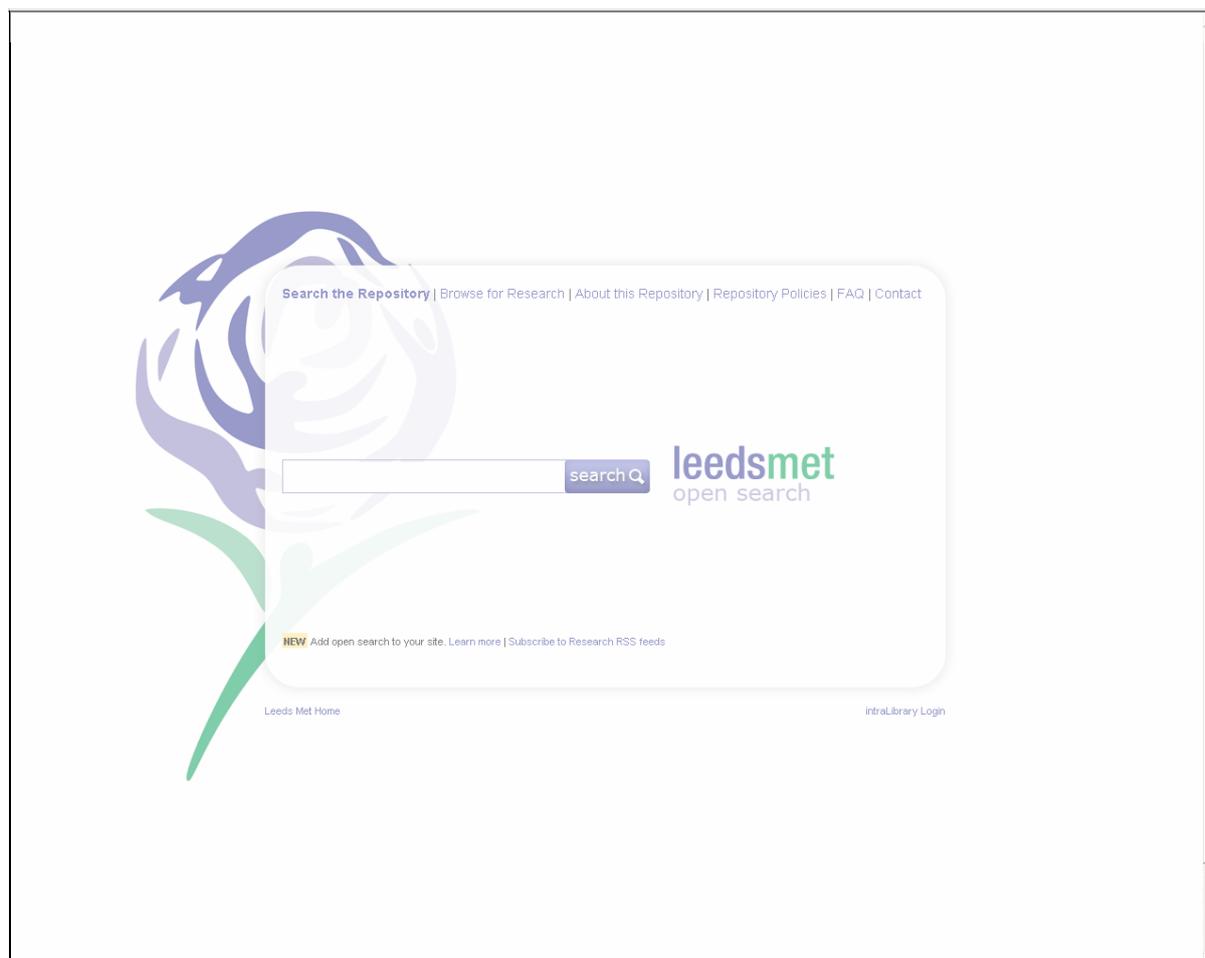
When asked whether they were aware “that a large proportion of academic publishers will allow you to deposit your published research in an Open Access repository where it can be accessed free of charge”, 49% of respondents replied that they were not aware and 20% were not sure (leaving only 31% who could say they were aware). There was little difference between researchers with different lengths-of-service: 22% of postgraduate students, 20% of new researchers (<5 yrs), 40% of researchers who had been in service 6-10 years, 33% of researchers who had been in service 11-15 years and 25% of senior researchers (>15 yrs) were aware that this was true, the others were unaware or unsure. Figure 5 shows that in some disciplines (ethics/tourism/international studies grouping, art, sport, history/anthropology/listed buildings grouping and writing/culture/English/literature grouping) no researchers answered that they were aware of this fact.

It is clear from the answers to the questionnaire that although the general awareness of open access and self-archiving amongst researchers in Leeds Met is quite high, few researchers have further knowledge of the finer details and implications of the open access movement.

7.3. The Repository open search SRU interface

The IRISS open search interface⁷, from which our open search has been developed, is a web based interface written using PHP 5 for the core server-side functionality. The IRISS interface is released under the GNU General Public License v3 and as such, our modified interface is released under the same license. This allows either interface to be adopted/adapted by other JISC projects or indeed anyone should they wish to. This could be as simple as re-branding the front end of the interface or could involve large scale modifications, enhancements or additions.

Figure 8: A screen shot of the open search SRU interface:



Links to navigate the site are **Search the Repository** (this page); **Browse for Research** (screen shot overlaid); **About this Repository**; **Repository policies**; **FAQ**; **Contact**.

Links at the bottom of the page are **Leeds Met Home** and **intraLibrary Login**

There are also links to **Add open search to your site** which provides code that can be cut and pasted to generate a search box for the repository on any html web page; **Subscribe to Research RSS feeds** (currently in development – see *Implications* section of this report).

⁷ The IRISS project aimed produced a customisable search interface for displaying SRU formatted XML; the code can be downloaded from <http://code.google.com/p/sruopensearch/>

Figure 9: Portion of a screen shot of the Browse for Research page:



The interface allows the repository to be browsed by faculty or by LCC. Development work is ongoing – see *Implications* section of this report.

7.4. Project blog

The project blog is a valuable project outcome that has been both an invaluable project management tool and actively used as discussion forum by the repository and Open Access communities. For example, posts on **Adapting intraLibrary** generated insightful comments from colleagues at other institutions that informed the development of the SRU interface. Issues that have generated, sometimes heated discussion, include copyright and the putative citation advantage of Open Access to research.

At the end of the project the blog will be officially archived and remain available at <http://repositorynews.wordpress.com/>

8. Outcomes

The over-arching aim of the project was described in the project plan as follows:

The broad aim of the project is to establish a repository for Leeds Metropolitan University, based on open standards, which not only meets the needs of the key stakeholders at Leeds Met, but also makes a valuable addition to the open access community. Leeds Met will be making materials available which would otherwise not be promoted to the open access community, and at the same time the awareness of staff and students at Leeds Met and across the wider Regional University Network will be raised in respect of the value of open access publishing.

This aim has been achieved. A repository has been established, and it supports open standards. Stakeholders have been widely consulted throughout the process of selection and implementation, and the repository is being set up and configured to meet their needs – this is an on-going development. Considerable advocacy activities have taken place, and the concept and value of open access publishing has been raised across the University. If anything, the aim to promote this across the Regional University Network has not yet been achieved, but this remains a goal, and there will be opportunities to promote this. A body of content is being assembled in the Repository, and so the aim of making materials available to the community on the open access model is also being achieved.

The objectives were described as follows:

- ***to raise the profile of open access publishing, and access to open access materials, through presentations to key stakeholder groups and training events***

This objective has been met. Both the Project Manager and the Repository Development officer have engaged with key stakeholder groups around the University. Both have taken any opportunity to promote the advantages of open access publishing, including at the University's Research Sub-Committee, and at various Faculty-based fora. Both have published Reflections on the University's website. The outcomes of the Masters project associated with the Repository indicate a degree of awareness amongst stakeholders. The University's Academic Committee has received and approved a paper advocating that the University make its learning objects accessible through an open access environment (using the Repository). The University's Research Office regularly refers staff to the RDO to encourage use of the Repository as a means of facilitating access to research outputs. Both the former and the current Pro Vice Chancellors for Research are supportive of the project and

- ***to conduct an institutional needs analysis to establish the most appropriate (initial) focus for a Leeds Met institutional repository***

The institutional needs analysis was carried out in the early stages of the project through the Project Manager meeting with a wide range of prospective stakeholders, both in one to one meetings and more formal surroundings, and sounding out what potential uses of the Repository they may have. The timing of this stage of the project, coinciding as it did with preparation for the 2007 RAE, alongside the University's clearly articulated agenda to increase its research profile, led to a distinct steer towards the starting point for content of the Repository being open access research outputs. However, this process also raised the profile of the Repository as a solution for a number of other outputs, which include: learning, teaching and assessment objects; digitised readings; digitised archival materials; electronic exam papers; electronic theses; student project work.

- ***to conduct a market analysis of the most appropriate software solutions and, based on the institutional analysis, usability testing, software and hardware needs and support issues, and stakeholder input, choose a software / hardware solution for the Leeds Met repository***

This objective was met, and the choice of software platform was very clearly based on the outcomes of the institutional needs analysis, which pointed to the need for a flexible product which would suit the diverse needs of Leeds Met. Stakeholder input was key to the process, and the project team was given a very clear steer by the Repository Consultancy group. The choice of software platform was made in tandem with a decision around internal or external hosting of

the software, and it was decided to outsource this to the software provider, Intrallect.

- ***to establish workflows for the ingest of content to the repository, including metadata standards***

The Repository Development Officer has worked with Intrallect to draw up metadata templates for research outputs. Some work now needs to be done to address templates for learning objects. He has established workflows for the ingest of materials to the Repository. When this aim was originally expressed, it was envisaged that the workflows would be to facilitate self-deposit by researchers/ academic staff – i.e. the content providers. This was possibly too ambitious within the timescale of the project, but is still seen as a goal for the project in the medium term.

- ***to commission a representative body of initial content, which will promote the use and benefits of the repository, and in its turn generate new content***

This objective has been met, to an extent, although undoubtedly the delay in launching the repository, and the work that has been done to tailor the software platform, has had an effect in the latter stages of the project on the time available to encourage deposit. However, the content that is in the repository is representative of the types of materials we envisage collecting and making available over the short to medium term, i.e. full text research outputs and learning objects.

- ***to establish the Leeds Met repository as a standard element of the workflow of those generating research outputs and or learning / assessment objects, or any other content which is identified at the outset as being the primary initial focus of the Leeds Met repository***

This aim has been achieved, in as much as the repository is regarded as a key element in the University's technical infrastructure. One key element of the choice of software platform was that it would support the range of needs expressed by key stakeholders, and in conferring with these groups and eliciting their needs, awareness of the potential of the Repository to meet a number of business needs has been raised. The University is ambitious in wanting to amass a collection of research outputs and learning objects to support both research and teaching and learning at the University, and, as covered elsewhere in this report, it is becoming clear that the Repository is central to making this a reality.

The methodology for this project was driven by the need to establish the business needs of the University with regard to an open access repository, and to gain the buy-in of key stakeholders in ensuring that these needs were met. It has been process driven, with advocacy, procurement, implantation and embedding activities being carried out in a planned way which has been clearly documented. The project has been steered by a Project Consultancy Group, made up of key (and some very senior) stakeholders from across the University. This has given the project buy-in from an early stage, and has also helped in raising the profile of the project and ensuring that it has support in the longer term. The usefulness of such a group both politically and practically is a potential learning point for other like projects.

9. Conclusions

We are able to conclude from this project that Intrallect's intraLibrary software is extensible to a wide range of content and, in particular, adaptable to serve as an effective Open Access research repository. However, to achieve this has been a steep learning curve and the system still requires some development to be fully effective for this specific purpose. The main areas for further development are:

- Continued development and refinement of the SRU search interface
- Continued development work to ensure OA content is discoverable on the public web; by implementing XML site-maps and, ideally, working with Intrallect to facilitate full text indexing
- Continued development work on self-archiving and/or mediated work flows – possibly utilising SWORD technology

We can also conclude that there are real issues in engaging with the academic community to promote the model of Open Access to research in its current form, at least in the short term. Procurement of full text content has followed the pattern exhibited elsewhere in the sector. A number of full text articles are available within the Repository. However, to date the bulk of contributions have been in citation format. The University Research Office is very supportive of the project, and are convinced of the potential of the Repository to raise the profile of research at the University. It is hope that this commitment, combined with the already high profile of the Repository, will lead to higher levels of full text deposit.

We are confident that, with sufficient institutional support, the Leeds Met repository can develop to fulfil the diverse requirements of its stakeholders to effectively manage the range of institutional, digital outputs.

10. Implications

With intraLibrary and the SRU interface we now have an incipient infrastructure to manage both research material and learning objects; the discrete types of material can be managed entirely separately, however, there is also potential for the ongoing development of a holistic approach to the management of the full range of digital resources produced by a modern university.

The project has utilised appropriate web-technology around a central management system (intraLibrary) to achieve decentralised resource discovery – currently research only. This approach does not preclude intraLibrary being used as originally designed, as an authenticated repository system for the federated management of digital resources.

There are implications for ongoing repository development at Leeds Met, for the community as a whole and for the software provider, Intrallect.

10.1. Ongoing repository development at Leeds Met

Implementing an Institutional Repository at Leeds Metropolitan University was funded as a start-up project. As such, it should be recognised that there is still a considerable amount of work to be done to both develop and refine the infrastructure and to promote its use to the university community.

10.1.1. Development of SRU

- **Differentiating content by type**

If the SRU interface is to be used to facilitate Open Access to learning objects as well as research, it is essential that there is a mechanism to differentiate content by type.

intraLibrary can be configured (by collection) to allow or disallow published content to be searched by external systems; currently all learning objects are in a collection that cannot be searched by external systems, meaning they are only discoverable to authenticated users – i.e. they will not be returned by the SRU interface. However, any resource that is discoverable externally *will* be returned by the SRU interface which, in its current incarnation, has been specifically configured to reflect the research metadata template defined in intraLibrary and which will display an alternative metadata template, such as that used for learning objects, in an inappropriate manner.

It should be possible to differentiate type of content by using an additional metadata field, for example, and either apply a separate template from a common SRU interface or implement a separate SRU installation to search for the different type of content.

N.B. Recent information from Intrallect suggests that it should be possible to use ‘collection tokens’ to achieve this. Further research is required and falls outside the scope of the start-up project due to lack of time.

- **Advanced search**

Initial development of the SRU interface has focussed on simple search and browse functionality; ongoing development will emulate intraLibrary’s sophisticated search facility that allows cross-referencing of multiple metadata fields.

Other issues will be around whether to use a common interface for both research and learning objects and how an advanced search might allow a user to filter by resource type if this approach is taken.

- **Number of resources by faculty/subject**

The SRU interface allows users to browse the collection by faculty or by subject but currently gives no indication of the number of resources available. There have been discussions around how this may be achieved on a technical level and implementation is a priority for the next phase of the project.

- **RSS**

RSS feeds have been implemented (by faculty). However the research metadata template comprises 4 description field and it is the bottom-most field (which holds refereed/not refereed) that is being exposed in an RSS feed; ideally we would want the article abstract exposed which is held in the second description field. There have been preliminary discussions with Intrallect around how this issue may be resolved but a solution is yet to be found. It would then be straightforward to set up additional RSS feeds, by subject or by author for example.

10.1.2. Workflows

- **Research**

In the context of research, we are currently working closely with the University Research Office to draft recommendations to the university’s Research Sub-Committee to ensure that the repository is efficiently populated with citation information and, wherever possible, full text; the process we

are considering would require an individual researcher or their faculty to provide information to the repository team / URO at the point of publication / acceptance for publication. As a minimum this might be a full Harvard reference and abstract but ideally would also include an author produced version of the full text.

In the first instance, upload and cataloguing will be carried out by the repository team / library though, as the repository becomes an established element of the research infrastructure, we also intend to promote self-archiving to academic staff.

- **Learning Objects**

The infrastructure for research (both OA and citation) has now been finalised, ongoing refinements notwithstanding, and this material can therefore be managed effectively and independently, allowing us to focus on developing and implementing workflows and use-cases for learning objects. The Streamline project (report available separately) has developed a number of such use-cases and assembled several user-group cohorts that will inform ongoing development, particularly in the context of the JISC funded PC3 project⁸.

It is also anticipated that outputs and theoretical perspectives from the PERSoNA project (report available separately) will be refined and incorporated into workflows for both research material and learning objects. For example, appropriate development and integration of widgets based on SWORD (Simple Web-service Offering Repository Development); SHERPA RoMEO / JULIET; SRU interface.

10.2. Development of intraLibrary

There are several functional requirements that are contingent on further development of the intraLibrary software, particularly in the context of research; Intrallect regularly issue updated versions of the software and their policy is to prioritise according to the needs of their user base. These needs inform the companies “road-map” of planned future developments to the software.

The majority of Intrallect's customers use intraLibrary exclusively to manage learning objects – however, others have explored using it as an OA research repository which was the stimulus to including bibliographic citation metadata in v.3.0.

Several users recently posted to a discussion on the intraLibrary users forum⁹ outlining a “wishlist” of features they would like to see incorporated in future releases.

From Leeds Met's perspective, given the issues that we have encountered in ensuring research content is fully discoverable on the public web, a primary requirement is for intraLibrary to facilitate full text indexing by Googlebot and other search engine crawlers. Intrallect have indicated that such a development is on their road-map. However, the company is unable to specify when this is planned.

Other features requested include:

- Reporting and statistics; number of views / resource downloads (OA research full text download.)
- More control of collections exposed to external search – metadata discoverable anywhere; authentication gateway for full access.
- Ability to display thumbnail images by SRU
- Ability to search within collections by SRU

⁸ See <http://pc3project.wordpress.com/about-pc3/> for more information

⁹ <http://intraLibraryuserforum.ning.com/>

- Ability to manipulate RSS feeds

10.3. A stakeholder driven approach

From the outset the project has been driven by a disparate group of stakeholders with different requirements and this has had both benefits and drawbacks.

When discussing this issue, it is important, first of all, to consider the institutional context at Leeds Metropolitan University which, historically, is a polytechnic that gained chartered university status in 1992. As such, its heritage is very much in teaching and learning rather than research with, arguably, a more vocational than academic flavour. However, since 1992, the research profile has steadily increased culminating in unprecedented success in the 2008 Research Assessment Exercise. The university is naturally keen to capitalise on this success and enhance its research profile further while also continuing to emphasise its student focussed teaching and learning credentials. The implementation of an integrated repository to support both research outputs and learning objects reflects this dual focus.

The benefit of this approach has been a high level of stakeholder engagement without disenfranchisement amongst disparate members of the university community. Moreover, teaching and research, the two central areas of academic business, are complementary rather than exclusive and any distinction tends to be somewhat artificial in any case.

The Leeds Met repository has been an ambitious project, in large part due to the disparate needs of its stakeholders. It is possible that a more focussed project towards either Open Access research or reusable learning objects would have made more rapid progress in either of these respective areas; for example, if we had implemented EPrints or DSpace software then there would have been fewer issues in implementing a functioning Open Access repository of research, however, such an approach might have disenfranchised other important stakeholders at Leeds Met.

11. Recommendations

Specific recommendations derived from this project include:

- Sector emphasis is still focussed on Open Access to research; JISC should invest specifically in projects examining multi-purpose institutional repositories
- There needs to be more research into reusable learning objects/open educational resources; both with respect to infrastructure and the prevailing culture
- Institutions and the wider community should consider developing reward mechanisms to encourage sharing of Open Educational Resources.
- Cross project collaboration is valuable and should be encouraged and facilitated wherever possible
- Advocacy and dissemination activity should begin as early as possible, be sustained and be reactive to the specific needs of stakeholders

12. References

Crow, Raym "The Case for Institutional Repositories: A SPARC Position Paper" *ARL Bimonthly Report* 223 (2002).

Lynch, Clifford. A "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" *ARL Bimonthly Report* 226 (2003).

Guédon, Jean-Claude, 2001. In Oldenburg's Long Shadow: Librarians, Research Scientists, Publishers, and the Control of Scientific Publishing,

<http://www.arl.org/resources/pubs/mmproceedings/138guedon.shtml>

(Accessed 1st May 2008)

Yiotis, Kristin, 2005. The Open Access Initiative: A New Paradigm for Scholarly Communications, *Information Technology and Libraries*; 24, 4

Swan, A. (2006) The culture of Open Access: researchers' views and responses. IN JACOBS, N. (Ed.) *Open Access: Key Strategic, Technical and Economic Aspects*. Chandos Publishing.

Harnad, S. (2006) Opening Access by Overcoming Zeno's Paralysis. IN JACOBS, N. (Ed.) *Open Access: Key Strategic, Technical and Economic Aspects*. Chandos Publishing.