

Project Acronym: OCRIS
Contact: Kathleen Menzies or Gordon Dunsire
Date: October 2009

JISC



Online Catalogue and Repository Interoperability Study (OCRIS)

Final Report
Duncan Birrell, Gordon Dunsire and Kathleen Menzies
Centre for Digital Library Research, University of Strathclyde
September 2009

Table of Contents

Acknowledgements	4
1. Background	10
2. Aims and Objectives	11
3. Methodology	11
4. Implementation	12
4.1. Questionnaires.....	12
4.2. Sampling and in-depth analysis.....	15
4.3. Case Studies	16
5. Outputs and Results	16
5.1. LMS and IR systems and software.....	17
5.2. Interoperability.....	19
5.2.1. Administrative Systems.....	19
5.2.2. LMSs and IRs	21
5.3. Services currently supported	24
5.3.1. Institutional Repositories.....	25
5.3.2. Library Management Systems.....	28
5.3.3. Key findings regarding services offered by LMSs and IRs.....	30
5.4. Integration services and tools	30
5.5. Scope Overlap	32
5.6. Authority Control.....	33
6. Summary of analysis of in-depth studies.....	34
6.1. Introduction.....	34
6.2. Main findings.....	34
6.3. Descriptions of scope.....	35
6.4. Overlap.....	38
6.5. Journal articles	39
6.7. Subject headings	42
6.8. Name Authority Control and finding publications by specific authors, research groups and departments	43
7. Summary of main findings of case studies.....	47
7.1. Rationale	47
7.2. Introduction.....	47
7.3. Systems and Scopes	47
7.4. Key points of comparison.....	48
7.4.1. The Impact of Research Assessment.....	48
7.4.2. Systems Issues	50
7.4.3. Authority Control and Standards.....	50
7.4.4. External sourcing of bibliographic records.....	51
7.4.5. Links between LMSs/OPACs and IRs.....	52
7.4.6. Links with other institutional systems.....	52
7.4.7. Re-imagining the role of IRs and the LMS.....	53
7.4.8. Resource Discovery Platforms	53
7.4.9. Development of Identity Management Systems within Institutions	54
8. Outcomes.....	55
9. Conclusions.....	56
10. Implications.....	59
10.1. Duplication and scholarly needs.....	59
10.2. Creating links between systems and records.....	60
11. Recommendations.....	62
16. References	65
Technical Glossary.....	67
Key terms used by the project	67
Technologies and systems referred to in the Final Report	69
Appendices	74
Appendix 1 - OCRIS Questionnaires and Guidance	74
Appendix 2 - Case Studies in 2 Higher Education Institution Libraries.....	74
Appendix 3 - In-depth studies of a sample of IRs and OPACs.....	74

Table of Figures

Figure 1 - Questions about item types and scope.....	14
Figure 2 - Questions about services supported by Institutional Repositories.	14
Figure 3 - The OCRIS administrative systems questionnaire.	15
Figure 4 - The various software packages used by UK HEI Institutional Repositories.....	18
Figure 5 - Answers given to OCRIS' questions about interoperability of admin and library systems..	20
Figure 6 - Responses to the question, "Does the LMS interoperate with the IR?"	21
Figure 7 - Responses to the question, "Does the IR interoperate with the LMS/OPAC?"	21
Figure 8 - Interoperability between IRs and LMSs.....	23
Figure 9 - Various institutional systems with which IRs interoperate.	24
Figure 10 - Various institutional systems with which LMSs interoperate.	24
Figure 11 - Services currently supported by IRs.....	26
Figure 12 - Services which will be supported by IRs.	27
Figure 13 - Services offered by LMSs.	29
Figure 14 - Use or planned use of what OCRIS terms "integration services" within LMSs and IRs. ...	31
Figure 15 - Venn diagram showing scope overlap between IRs and LMSs in HEIs.	32
Figure 16 - Classification and subject schemes being used within IRs and LMSs.	33
Figure 17 - Name authority schemes being used within IRs and LMSs.	33
Figure 18 - A pre-print held in Glasgow's DSpace Service.	39
Figure 19 - Partial record for the post-print version held in Enlighten.	40
Figure 20 - UCCA bibliographic records for the magazine Design, published from 1949 - 1999.	41
Figure 21 - The Search by Year interface of Design Online.....	41
Figure 22 - Doctoral Thesis defended at the TFTS department at Glasgow, with embargo period imposed.	42
Figure 23 - Extract from the CogPrints repository's "Browse by Subject" menu.	43
Figure 24 - An edited version of results retrieved from e-Prints Soton, showing duplication of bibliographic records between e-Prints and ECS e-Prints (fig.25).	44
Figure 25 - The corresponding records in ECS e-Prints.....	44
Figure 26 - Extract from the University Structure hierarchy used in e-Prints Soton.	45
Figure 27 - Extract from CogPrints with Research Group structures implemented as a menu.	45

Table of Tables

Table 1 - Interoperability and integration enabling features of LMSs used by in scope HEIs.	17
Table 2 - OAI-PMH and interoperability focussed features of RDPs used by in scope HEIs.	18
Table 3 - Number of responses received to the OCRIS questionnaires, by group.....	19
Table 4 - Number of responses received to the OCRIS admin systems questionnaire, by department.	19
Table 5 - Conflicting questionnaire responses.	22
Table 6 - Interoperability of LMSs and IRs across the group of HEIs - both before and after the introduction of a new category to resolve conflicting data.	23
Table 7 - Services currently supported by IRs.	25
Table 8 - Services currently supported by LMSs.....	28
Table 9 - Integration services in place or pending in IRs.	30
Table 10 - Integration services in place or pending in LMSs..	30

Acknowledgements

The OCRIS Project was funded by the JISC Scholarly Communications Working Group (SCG) as part of the JISC's Information Environment Programme and their contributions are gratefully acknowledged.

The participation of **Cambridge University Library** and the **University of Glasgow Library** were essential to Workpackage 3 of OCRIS and invaluable to our research in more general terms. The Project Team would therefore like to thank the following individuals:

Barbara Bültmann: DSpace@Cambridge Support and Liaison Officer, Cambridge University.

Marie Cairney: Principal Library Assistant, University of Glasgow.

Ed Chamberlain: Systems Librarian, Cambridge University.

Patricia Killiard: Head of Electronic Services and Systems, Cambridge University.

William Nixon: Deputy Head of Library Information Systems, University of Glasgow.

Elin Stangeland: DSpace@Cambridge Repository Manager, Cambridge University.

Karen Stevenson: Library Systems Administrator, University of Glasgow.

Rosemary Stenson: Head of Cataloguing, University of Glasgow.

Hugh Taylor: Head, Collection Development and Description, Cambridge University.

Individuals who pilot tested the three online OCRIS questionnaires, devised and administered under Workpackage 1, provided us with much useful feedback and advice. They are:

Chris Awre: Head of Information Management at the University of Hull.

Alan Dawson: Programmer at the Centre for Digital Library Research, University of Strathclyde.

Emma McCulloch: Research Co-ordinator at the Centre for Digital Library Research, University of Strathclyde.

Alan Slevin: Institutional Repository Manager, University of Strathclyde.

Judith Smart: Systems Librarian, University of Strathclyde.

Hugh Taylor: Head, Collection Development and Description, Cambridge University.

April Woods: Directorate Administrator, Information Resources Directorate, University of Strathclyde.

Ben Wynne: Programme Manager (Information Environment), JISC Executive.

Chris Awre and **Alma Swan** (External Consultant and Director, Key Perspectives Ltd.) are also thanked for granting us permission to use their ontology of Institutional Repository Services within the OCRIS questionnaire.

Others who contributed informally to the project via email should also be thanked for their contributions:

Jenny Delasalle: E-Repositories Manager, Research & Innovation Unit, University of Warwick Library.

Stuart Hunt: Data Services Manager, the Library, University of Warwick.

Dr. Heather V Haskins: Project Officer, MERIT (Making Excellent Research Influential and Transferable).

Stuart Halliday: Technical Services Manager, Information Services, University of Sunderland.

Simon Kerridge: Head of Graduate Research Support & Assistant Director (Research), Academic Services, University of Sunderland.

Linda Noble: Head of Library and Information Services, National Marine Biological Library, Marine Biological Association.

Dave Price: Head of the Systems and e-Research Service, Oxford University Library Services.

Janet Wharton: Cataloguing & Metadata Librarian, the University of Nottingham.

Executive Summary

Context

What is an Institutional Repository (IR) and what should be its role? Library Management Systems (LMSs) and their Online Public Access Catalogues (OPACs) have traditionally been used by HE institutions to provide information about the publications and other bibliographic output of the organisation. Although this role is well-established, the development of IRs potentially conflicts and overlaps with these functions.

Library and repository systems are undergoing intensive development to adapt to the demands of a digital, networked environment, taking into account changing technologies, user behaviours, and research and service agendas of institutions - yet links between these and other HEI systems are often non-existent or inadequate.

There is little evidence that the range of institutional, departmental, subject or format specific IRs and OPACs used in UK HEIs are being developed with shared workflows, re-use of data, or other considerations of efficiency. This inhibits the development of services such as the production of management reports or publications lists for research assessment and planning. It also inhibits internal and external end users wishing to easily access the institution's rich range of documentary output.

Interoperability between systems is a key mechanism for improving this situation. This is underpinned by **accurate metadata**, the **re-use of data** and, where possible, **open standards**. These allow, for example, the derivation of system and function specific information from a single master record. Fragmentation must be avoided. Interoperability should be built into the workflows of relevant institutional systems, including administrative processes dependent on such metadata; *all* institutional stakeholders managing information systems should be involved in discussions about how this can be achieved.

The **aims and objectives** of OCRIS were to:

- Survey the extent to which repository content is in scope for institutional library OPACs, and the extent to which it is already recorded there;
- Examine the interoperability of OPAC and repository software for the exchange of metadata and other information;
- List the various services to institutional managers, researchers, teachers and learners offered respectively by OPACs and repositories;
- Identify the potential for improvements in the links (e.g. using link resolver technology) from repositories and/or OPACs to other institutional services, such as finance or research administration;
- Make recommendations for the development of possible further links between library OPACs and institutional repositories, identifying the benefits to relevant stakeholder groups.

Key Findings

Interoperability and services

1. Interoperability between IRs and LMSs in UK HEIs is currently rare - only 2 percent of questionnaire respondents state that their systems *definitely* interoperate, with a further 14 percent stating that interoperability is *pending*.
2. Interoperability of either of these system types with *some* type of other institutional system is moderately high, and is slightly higher for LMSs than IRs.
3. Interoperability between LMSs or IRs and a *range* of other institutional systems is limited. It cannot be said that interoperability is substantial or that a wide *variety* of administrative systems interoperate with any individual library system.
4. The REF has clearly been a factor in the establishment or consideration of interoperability between Institutional Repositories and other administrative systems.

5. Services stemming from library systems are limited and narrow, excepting the generation of usage statistics and metadata enhancement services.
6. The generation of reports for specific administrative departments is not a common service offered by either IRs or LMSs.
7. The most popular service offered by IRs remains "advice on Open Access" suggesting perhaps that they are still in their infancy, still require explanation thus have yet to spread their wings in terms of widening their range of services.
8. The use of metasearch/linking tools as well as web services and APIs is moderately popular within LMSs and IRs; the data gathered is not sufficient to discern why or what these tools and services are being used for.

Duplication and scope

1. There is significant scope overlap (81 percent) for *all* item types held in IRs and OPACs.
2. The scoping distinctions and boundaries for IRs and OPACs are becoming increasingly blurred, with many IRs containing bibliographic data and OPACs containing links to full text.
3. Duplication at both record and item level is frequent, especially for print/electronic copies of theses or journal articles.
4. For many OPACs and IRs, *any* type of content is in scope, regardless of the items *currently* recorded or held.
5. Links between print (OPAC) and electronic (IR) copies of theses are frequently instantiated within both systems.
6. Linking for content other than theses is not common. However, within some IRs link resolvers are beginning to be used to direct users to related library holdings.
7. Some HEIs are choosing to expose both OPAC and IR data through the use of Resource Discovery Platforms (RDPs) which offer "vertical search" functionality. The popularity of these systems seems likely to increase.

Authority control and description

1. Authority control within LMSs is high; format and content standards are well supported, with MACHine-Readable Cataloguing (MARC) standards and Library of Congress Subject Headings (LCSH) the most commonly used alongside local authority lists.
2. In IRs there is little authority control for subjects, and only a moderate amount of effective classification. In-house lists are predominantly used for the construction and maintenance of name authorities.
3. Within Institutional Repositories standards are not applied adequately and often not at all.
4. There are frequent inconsistencies and a lack of completeness in statements made on both IR and OPAC Web pages about item types and scoping.
5. Modifications allowed to the item and format fields of DSpace or EPrints software, and the presentation of administrative or outdated terms by LMSs within their search/browse lists, undermines consistency, standardisation and clarity for end users, across the sector.
6. The use of vocabularies and standards within library systems in any given HEI is fragmented and disjointed; there is little commonality in resource description.
7. LCC is frequently used within EPrints repositories as the top levels come bundled with the software; however there appears to be some confusion in the IR community about the distinction between Library of Congress Classification (LCC) and Library of Congress Subject Headings (LCSH). This is clearly significant, suggesting limited knowledge and expertise of professional standards within the IR community.
8. There are many benefits in recognising the Author/Creator/Person field as the metadata element common to all internal HEI systems.

9. There is an increasing awareness of the role that could be played within Institutional Repositories by institutional IDs.
10. There will be an increased role for institutional (or even international, if one is to be more ambitious) personnel and group identity management schemes to enable authentication, access to services and the gathering and compilation of data for internal and external purposes.
 - 10.1. Supporting this can become time-consuming as HR-produced codes or the relationship between a department and the University hierarchy may be subject to constant change. Keeping records up-to-date can therefore be an ongoing challenge.
11. Resource Discovery Platforms (RDPs), because of their nature (for example, the use of "tag clouds") reveal inadequate metadata in catalogue records more readily than do other interfaces.
12. RDPs are only a partial solution to making all relevant items visible to users in one place; they may present search results in ways not best suited to the scholarly needs of higher-level end users and do not necessarily reveal the richness of a library's collections.

Flexibility and working practices

13. The fragmentation or disconnectedness of HE information systems puts a strain on the abilities of cataloguers/bibliographic services staff to work effectively across both IRs and LMSs or for shared workflows to be developed across departments.
14. Lack of resourcing puts a strain on the abilities of cataloguers/bibliographic services staff to work effectively across both IRs and LMSs or for shared workflows to be developed across library departments.
15. Administrative staff and the systems with which they work are not sufficiently involved in cross-departmental collaborations with library systems staff.
16. Batch processing is still considered an effective way to share data between non-library and library systems (particularly HR systems and LMSs).

Recommendations

To Higher Education Institutions

1. Expose all LMS and IR records for harvesting and linking (except in cases where legal requirements restrict such data re-use) via distributed/federated/meta search using technical protocols such as OAI-PMH, Z39.50, SRU/SRW or link resolvers, as appropriate to the technical infrastructure.
2. Improve co-ordination between all departments possessing institutional information-gathering systems and their staff, with support at the highest levels of the institution, in order to develop efficient workflows, reduce un-necessary duplication of effort and formalise collaboration.
3. Align the systems of both libraries and administrative departments, and their attendant data-processing practices, more closely.
4. Consider establishing a centralised system and attendant workflows for cross-checking and cleaning metadata that is to be shared between systems, to ensure quality, usability and re-usability by both internal and external service providers.
5. Consider options other than batch processing (such as web services or applications, underpinned by open standards) where administrative departments are sharing data with library systems.
6. Develop clear policies on the scopes and uses of IRs and OPACs.
7. Present clearly, comprehensively and comprehensibly, to both staff and end users, the scopes of IRs and OPACs.
8. Develop a single scheme for describing item types/formats and scope within OPACs and IRs, with interoperability requirements and local needs fully accounted for.
 - 8.1. This scheme should be tested with a variety of users (teaching staff, researchers and undergraduates) to ensure it speaks to their needs and is understood by them.
 - 8.2. If JISC act on recommendation A above, library staff should be allowed to attend meetings and contribute to the formulation of a cross-institutional scheme, with local needs discussed as part of the activities of the group.
9. Ensure the use of format and content standards within IRs to avoid the need for future "retro-conversion" or "re-keying".
10. Support the interoperability of subject authorities across institutional systems if common ones are not appropriate – this might build on the work of existing mapping and switching projects such as OCLC's Terminology Service Pilot and the High-Level Thesaurus Project (HILT).
11. Use interchange formats and cross-walks based on open standards more widely and extensively, to assist in the sharing and exchange of records conforming to different format and content standards.
12. Reassess the use of Library of Congress Classification (LCC) within IRs; staff should be familiar with the distinction between LCC as a classification system and LCSH as a subject heading system in order to determine whether these schemes meet the needs of their users and whether they accurately reflect their repository collections/items.
13. Establish controlled name authority lists for staff throughout the institution using agreed, recognised standards, to be made available to all relevant departments.
14. Develop (or if already in place, make consistent use of) persistent, institutional or departmental IDs, making these available internally and to other institutions.
 - 14.1. Relevant institutional systems could possibly hold these IDs in the form of a flat file. This would allow data relating to specific individuals to be "pushed" and "pulled"

between various systems. The IDs would become the "glue" allowing information to be disambiguated.

- 14.2. These IDs should be built into metadata workflows and be usable by both staff and end users. Person and role information from various institutional systems should be "warehoused" and made available as "a rich source of contextual metadata" (Green, 2007).
15. Recognise that LMSs are a rich source of bibliographic information on books, book items, monographs, conference proceedings and other items authored by institutional staff (which may not be recorded in an IR) hence an interoperable LMS could be leveraged for research assessment data gathering activities or for use within Research Management Systems.

To JISC

- A. Formulate and promote a standardised, controlled vocabulary describing item types, formats and scope, for use across HEI libraries, created in consultation with staff from those libraries and their institutions.
- B. Promote the development of, and collect, *unique* numeric, institutional identifiers derived from HEI HR systems to be shared amongst institutions and used as a field to which author/creator names can be mapped, within and across institutions. If desirable these might be mapped against RAE person identifiers or HEFCE institutional identifiers.
- C. Support the interoperability of subject authorities within and between institutions – this might build on the work of existing projects such as OCLC's Terminology Service Pilot or the High-Level Thesaurus Project (HILT).
- D. Raise awareness of the impact of current work-practices on future needs (in particular, the need to both utilise open standards to support data linkage within and across institutional systems, as the Semantic Web continues to develop).
- E. Create recommendations on where, or consider developing a service through which, libraries can source high-quality, openly available, "clean" bibliographic records with guarantees that these can be freely re-purposed and edited. This would allow HEI libraries greater certainty about the utility and cost-effectiveness of record-sharing.
- F. Investigate further, and in depth, the interoperability of administrative systems and the opinions of staff working with them, without placing undue emphasis on library systems.
- G. Investigate the implications for interoperability of tools/services becoming more common within IRs and LMSs such as Web Services and APIs.

1. Background

OCRIS builds on the work of many previous projects and reports including: *Linking institutional repositories* (Swan and Awre, 2006), the *JISC & SCONUL Library Management Systems Study* (JISC and SCONUL, 2008), IRI-Scotland (Institutional Repository Infrastructure for Scotland), and ongoing investigations into the Research Excellence Framework (REF) such as the JISC support project for HEFCE's REF bibliometrics pilot exercise (Higher Education Funding Council for England, 2009). OCRIS also considered the objectives of local projects such as the Welsh Repository Network (WRN) (Wales Library Education Forum, 2009), 'Metadata Enrichment for Repositories in a London Institutional Network' (MERLIN) (JISC, 2009a) and 'Enlighten and Research Integration' (ENRICH) (JISC, 2009b), all of which were informative to our analyses.

The context within which the project sits is broad, the needs and priorities of the full range of stakeholders associated with HEIs (from funding bodies and top-level University management to library systems and services staff, researchers, undergraduate students and the wider community) being of direct relevance.

85 UK HEIs have at least one Institutional Repository (IR) (with others planned for many, if not most, of the remainder) and all have an Online Public Access Catalogue (OPAC) - thus clear, solid recommendations on how to marry the functionality of the two systems are urgently needed.

There is debate about what exactly an IR is and what its role should be; although the OPAC (traditionally used for the dissemination of publications associated with the library's parent institution) is well-established, its scope and role are increasingly also being re-imagined. Logic dictates (and evidence shows) that Library Management Systems (LMSs) already hold much metadata pertaining to IR items - thus IRs should *theoretically* supplement the OPAC.

Developments are intensive - both LMSs and OPACs are changing and adapting to the demands of a networked world, the popularity of Google, "Web 2", "Web 3", and to the research and service agendas of their institutions. All are informed by stark economic and other competitive pressures. Yet, reviewing the literature, it is hard not to conclude that there are many divergent discourses around the enhancement of LMSs and OPACs, and that opinion is divided as to the direction and pace of future development, with little clear co-ordination or collaboration - both at institutional level and beyond.

The House of Commons Science and Technology Committee (2004) has stated that a "secure central archive for digital publications [is] essential to ensuring that today's research findings are not lost to the future"; in terms of HEIs, LMSs and IRs *must* begin to interoperate if this is to be achieved and if useful services are to stem from such an archive.

These services might include: an internal auditor or external assessment agency gathering data about an institution's research output; a member of staff compiling RAE/REF returns by interrogating the various catalogues; the generation of comprehensive publications lists for a particular author or subject area, for use within a taught course. If items are essentially siloed in separate systems that do not communicate with one another, such outcomes cannot be readily achieved. Application and information integration must be guaranteed, with the consistent application of metadata standards playing an essential role.

It is clear that there will be a place in any future "interoperable ecosystem" for emergent tools and technologies such as Resource Discovery Platforms (RDPs), "add-ons" which help users enter metadata consistently and correctly, and Service Oriented Architectures (SOAs).

The OCRIS Final Report, based on evidence, analysis and direct consultation with professionals, addresses in clear, practical terms the role of catalogues and repositories within HEIs and the ways in which links between them can be improved, taking into account the "range of different motivations" (SHERPA, 2006) amongst institutional stakeholders.

2. Aims and Objectives

The aims and objectives of OCRIS were to:

- Survey the extent to which repository content is in scope for institutional library OPACs, and the extent to which it is already recorded there;
- Examine the interoperability of OPAC and repository software for the exchange of metadata and other information;
- List the various services to institutional managers, researchers, teachers and learners offered respectively by OPACs and repositories;
- Identify the potential for improvements in the links (e.g. using link resolver technology) from repositories and/or OPACs to other institutional services, such as finance or research administration;
- Make recommendations for the development of possible further links between library OPACs and institutional repositories, identifying the benefits to various stakeholder groups.

Each of these aims was fulfilled satisfactorily.

3. Methodology

Using a combination of qualitative and quantitative research methods the project utilised desk research, purposive sampling and analysis, case studies within the libraries of 2 research-intensive HEIs and the deployment of 3 discrete but inter-linked online questionnaires to conduct a survey of all 85 UK HEIs possessing at least one IR. Many relevant practitioners within the IR and LMS communities commented on the project informally, via email or telephone conversations.

Using this mixed approach we were able to gather a large amount of numerical, factual and discursive or attitudinal information. Taken together, this constitutes a detailed description of the LMSs/OPACs and IRs in operation within in-scope UK HEIs and (to a much lesser though still illustrative extent, associated administrative systems), the services currently offered in or being developed for the 3 types of systems, and the views of those professionals working with them.

Although the questionnaire was one major component of the project (and the basis of Workpackage 1), combining it with other instruments meant that the project was able to gain a far broader picture with which to complement its data; the methodology also provided an in-built mechanism which could compensate for any low response rate or inconclusive data. Speaking directly to those working on the "front line" (through the case studies and informal discussions) and assessing HEI systems independent of any mediation (the purposive sampling) allowed for the aggregation of far more data than would have been possible using a questionnaire alone. This combination of methods was critical to the formulation of robust conclusions and recommendations.

There were 5 distinct workpackages, each flowing logically into the next with a degree of overlap where appropriate. These were as follows:

Workpackages

1. Questionnaire

This entailed the preparation and execution of 3 online web questionnaires. The first asked systems librarians and cataloguers about their LMSs. The second asked repository managers and staff about their IRs.

Both groups were asked questions about how their systems are implemented, which standards are in use, whether they interoperate with other institutional systems, and a range of related issues including which services are currently supported. They were asked whether or not they perceive a need for systems to interoperate and were given space for additional comments and corrections.

The third questionnaire asked a wide range of institutional administrative staff (including, research, development and innovation, internal auditors, finance, HR and information systems and services

staff) to provide details of the systems and standards they utilise and whether or not these interoperate with other institutional systems.

2. Desk research, sampling and analysis

Having identified HEIs with both OPACs and IRs, a purposive sample of these (10 in total) was selected. Establishing which item types were in scope for the OPACs and IRs, each sub-sample of institutional systems was interrogated at User Interface (UI) level in order to identify and retrieve records for which examples of overlap or duplication might be found (for example, a record for the same thesis, held by both the IR and the LMS). Using these records, many issues and problems stemming from duplication, arrangement and description were explored.

3. Case Studies

Two short case studies carried out over the course of several days at two research intensive institutions provided a chance to discuss, examine and expand on the questions posed by the research. What issues, barriers and challenges face professionals working with OPAC and IR development in both established workflows, and in the wider context? How do these relate to the wider agendas of institutions? How are these institutions establishing links between their LMSs, IRs and other systems and what are their visions for the future?

4. Conclusions, Recommendations, Final Report and Action Plan

A set of Conclusions and Recommendations of use to a wide range of stakeholders, detailing the technical, organisational and perceived issues relating to OPACs and IRs in UK HEIs were developed within Workpackage 4. These suggest how problems might be solved at technical, semantic, organisational, and cultural levels and aim to reflect the unique and diverse situations of UK HEIs as well as acknowledging commonalities. They also suggest future work which might be done in this area.

5. Dissemination

Sharing findings and recommendations with the wider community will be achieved through publications, presentations, and the use of mailing lists, attendance at relevant conferences and, potentially, the feeding of outputs into ongoing and future projects.

4. Implementation

4.1. Questionnaires

The OCRIS questionnaires comprised web-forms (at User Interface level) and a Microsoft Access Database (as the "back-end"), designed and implemented by the Principal Investigator, hosted internally at the CDLR and administered by the project team.

Using internet-based desk research, drawing on existing knowledge of and contacts within UK HEIs and using the Registry of Open Access Repositories (ROAR), the Directory of Open Access Repositories (OpenDOAR), the *JISC & SCONUL Library Management Systems Study* (JISC and SCONUL, 2008) and The University of Illinois OAI-PMH Data Provider Registry, information was compiled to help construct questions. Where helpful this data was pre-loaded into the questionnaire to make the process of answering less time-consuming for respondents and to demonstrate engagement with the community.

In the IR and LMS questionnaires, the OCRIS team asked recipients about the content types held by each identified IR/OPAC; the standards and authority controls in place within each (where this could not be discovered through desk research); plans to extend scope; OAI-PMH compliance; whether or not the addition of metadata was mediated by staff; which resource discovery tools and integration services were used or being considered and the other institutional systems – if any - with which their systems interoperated. Respondents were also given space to provide comments or corrections.

Project Acronym: OCRIS
Contact: Kathleen Menzies or Gordon Dunsire
Date: October 2009

The bottom half of the form asked about services offered by (or planned for) library systems using a list based largely on the ontology proposed in *Linking UK Repositories* (Swan and Awre, 2006).

While this list was not entirely applicable to LMSs, devising a distinct list for LMSs of equivalent granularity would have required a separate scoping exercise and utilised time and resources beyond those the brief 3 month project could supply. Furthermore, it would have introduced a level of incompatibility at the data analysis stage which could have compromised project findings. The guidance notes were used to explain why the list had been chosen.

While the questionnaires for IRs and LMSs were largely similar, the one designed for administrative staff had, by necessity, to be quite distinct given the wide ranging, non-library nature of the systems coming under the umbrella of "administrative". Further, the unfamiliarity of the OCRIS team with technical details of administrative systems and the likelihood that that stakeholder group would have less time (or inclination) to participate, deterred production of an overly-long or complex form.

In this we asked only 8 (rather than 15) questions, included additional space for comments, focussed on key points such as whether the systems supported Open Standards or interoperated with any library systems.

The following screenshots illustrate portions of all 3 questionnaires. (See Appendix 1 for full transcripts).

Collaboration			
(?) Does the catalogue contribute metadata to any of these Union catalogues?	<input type="checkbox"/> Archives Hub	<input type="checkbox"/> CAIRNS	<input type="checkbox"/> COPAC
	<input type="checkbox"/> WorldCat	<input type="checkbox"/> Other	<input type="checkbox"/> InformM25
			<input type="checkbox"/> RLUK Database
			<input type="checkbox"/> Talis Base
Scope			
(?) Which of these scholarly materials, often found in Institutional Repositories, might also be present in the OPAC?	<input type="checkbox"/> Abstract	<input type="checkbox"/> Journal article	<input type="checkbox"/> Scholarly text
	<input type="checkbox"/> Archival material	<input type="checkbox"/> Journal item	<input type="checkbox"/> Student coursework
	<input type="checkbox"/> Book	<input type="checkbox"/> Learning object	<input type="checkbox"/> Submitted journal article
	<input type="checkbox"/> Book item	<input type="checkbox"/> News item	<input type="checkbox"/> Thesis or dissertation
	<input type="checkbox"/> Book review	<input type="checkbox"/> Other	<input type="checkbox"/> Website content
	<input type="checkbox"/> Conference item	<input type="checkbox"/> Patent	<input type="checkbox"/> Working or discussion paper
	<input type="checkbox"/> Conference paper	<input type="checkbox"/> Report	
	<input type="checkbox"/> Conference poster	<input type="checkbox"/> Research dataset	
(?) Do you plan to extend the scope?	<input type="radio"/> Yes	<input type="radio"/> No	<input checked="" type="radio"/> Not known
(?) Do you plan to implement additional OPACs with different scope?	<input type="radio"/> Yes	<input type="radio"/> No	<input checked="" type="radio"/> Not known

Figure 1 - Questions about item types and scope.

Functions and services			
(?) What resource discovery improvement tools are in place or pending?	<input type="checkbox"/> Metasearch/Linking	<input type="checkbox"/> Software as a Service (SOAS)	<input type="checkbox"/> Vertical searching
	<input type="checkbox"/> Other		<input type="checkbox"/> Web service/API
(?) Services supported by the IR	Service	Current	Pending
	Aggregator Level		
	Automated metadata creation and enhancement	<input type="checkbox"/>	<input type="checkbox"/>
	Manual metadata creation and enhancement	<input type="checkbox"/>	<input type="checkbox"/>
	Other	<input type="checkbox"/>	<input type="checkbox"/>
	Verification of content (e.g. document format/layout/other standards-compliance)	<input type="checkbox"/>	<input type="checkbox"/>
	Verification of metadata	<input type="checkbox"/>	<input type="checkbox"/>
	Verification of the digital object itself	<input type="checkbox"/>	<input type="checkbox"/>
	Data Source Level		
	Collaborative or cross-Institutional Repositories at international level	<input type="checkbox"/>	<input type="checkbox"/>
	Collaborative or cross-Institutional Repositories at national level	<input type="checkbox"/>	<input type="checkbox"/>
	Format specific repositories gathering primary content	<input type="checkbox"/>	<input type="checkbox"/>
	Interim repositories for authors with no institutional repository	<input type="checkbox"/>	<input type="checkbox"/>
	Media-specific repositories gathering primary content	<input type="checkbox"/>	<input type="checkbox"/>
	Other	<input type="checkbox"/>	<input type="checkbox"/>
	Subject specific repositories gathering primary content	<input type="checkbox"/>	<input type="checkbox"/>
	Ingest Level		
	Digitisation	<input type="checkbox"/>	<input type="checkbox"/>
	Other	<input type="checkbox"/>	<input type="checkbox"/>
	Providing advice on IPR	<input type="checkbox"/>	<input type="checkbox"/>
	Providing advice on Open Access	<input type="checkbox"/>	<input type="checkbox"/>
	Providing technical help	<input type="checkbox"/>	<input type="checkbox"/>
	Repository construction or hosting	<input type="checkbox"/>	<input type="checkbox"/>
	Output Level		
	Access and authentication	<input type="checkbox"/>	<input type="checkbox"/>
	Bridging and mapping	<input type="checkbox"/>	<input type="checkbox"/>
	Citation	<input type="checkbox"/>	<input type="checkbox"/>
	Meta-analysis	<input type="checkbox"/>	<input type="checkbox"/>
	Other	<input type="checkbox"/>	<input type="checkbox"/>
	Other statistics services (e.g. subject coverage; number of authors)	<input type="checkbox"/>	<input type="checkbox"/>
	Overlay journal	<input type="checkbox"/>	<input type="checkbox"/>
	Preservation	<input type="checkbox"/>	<input type="checkbox"/>
	Publishing	<input type="checkbox"/>	<input type="checkbox"/>
	Reporting for specific administrative department (e.g. R&I)	<input type="checkbox"/>	<input type="checkbox"/>
	Reporting for use by IR Manager and associated staff	<input type="checkbox"/>	<input type="checkbox"/>
	Research and quality assessment (other than REF)	<input type="checkbox"/>	<input type="checkbox"/>
	Research Excellence Framework (REF)	<input type="checkbox"/>	<input type="checkbox"/>
	Seeking submissions from staff/students not currently included in scope	<input type="checkbox"/>	<input type="checkbox"/>
	Technology transfer/business advice	<input type="checkbox"/>	<input type="checkbox"/>
	Usage statistics	<input type="checkbox"/>	<input type="checkbox"/>

Figure 2 - Questions about services supported by Institutional Repositories.

Administration			
(?) Administrative department/function	<input type="radio"/> Senior management <input type="radio"/> Finance <input type="radio"/> Information systems and services <input type="radio"/> Internal audit	<input type="radio"/> Personnel and human resources <input type="radio"/> Quality assurance <input type="radio"/> Research assessment <input type="radio"/> Research support	<input type="radio"/> Technical support <input checked="" type="radio"/> Other
(?) Is this function supported by an automated information system?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not known		
(?) What type of software is used?	<input type="checkbox"/> In-house <input type="checkbox"/> Open-source <input type="checkbox"/> Proprietary <input type="checkbox"/> Other <input type="checkbox"/> Don't know		
Metadata			
(?) Does the metadata use open standards?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not known		
Interoperability (sharing data at machine to machine level)			
(?) Does the system interoperate with the Library Management System/OPAC?	<input type="radio"/> Yes <input type="radio"/> Pending <input type="radio"/> No <input checked="" type="radio"/> Not known		
(?) Does the system interoperate with the Institutional Repository?	<input type="radio"/> Yes <input type="radio"/> Pending <input type="radio"/> No <input checked="" type="radio"/> Not known		
(?) Does the system interoperate (current or pending) with any of the following internal systems?	<input type="checkbox"/> Course Administration <input type="checkbox"/> Electronic Resources Management <input type="checkbox"/> Enterprise <input type="checkbox"/> Finance	<input type="checkbox"/> Human Resources <input type="checkbox"/> Research Administration <input type="checkbox"/> Departments <input type="checkbox"/> Research Assessment Systems <input type="checkbox"/> Serials Management System	<input type="checkbox"/> Virtual Learning Environment <input type="checkbox"/> Virtual Research Environment <input type="checkbox"/> Other
(?) Could interoperability between internal systems be usefully improved?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Not known		

Figure 3 - The OCRIS administrative systems questionnaire.

The answers recorded in the database allowed us to address the aims of the project from a solid evidential basis. Additionally, comparing these responses to the findings of the in-depth sampling revealed inconsistencies between answers supplied via the questionnaire (for certain HEIs) and the data uncovered by desk research.

A mailing list was constructed with the email addresses of 1149 Systems Librarians, Repository Managers, Bibliographic Services staff, Finance, Research Co-ordination, HR and Information Systems and Services departments. 262 individuals were contacted regarding the IR questionnaire; 280 regarding the LMS/OPAC questionnaire; 599 the Administrative Systems questionnaire and 8 additional contacts from AHDS and SHERPA LEAP.

A major risk of the project was that there would be a low response rate to the questionnaire, especially given the time of year (summer vacation) during which it was administered. Although the response from the IR community was satisfactory (41 percent), responses coming from the other two groups were less so (LMS, 18.2 percent; administrative systems, 8.2).

4.2. Sampling and in-depth analysis

OCRIS devised a workflow by which to investigate LMSs/OPACs and IRs within 10 HEIs deemed (both individually and as a group) to be representative of a particular facet of the project and its context. This purposive sampling provided a snapshot of the current situation regarding duplication of both *records* and *scope* between bibliographic/publication systems, the way item types are described and curated, the implementation and customisation of the software being used and the extent to which links are being made between the two systems.

The 10 institutions chosen were:

The Universities of Aberystwyth, Cambridge, Cardiff, Glamorgan, Glasgow, Hull, Southampton, Strathclyde, the University College of the Creative Arts and the National Marine Biological Library (which serves 3 multidisciplinary marine research centre organisations).

The results of this analysis are detailed fully in Appendix 4 and summarised below (see page 26). In the cases of both Cambridge and Glasgow, results were directly informative of the two case studies.

4.3. Case Studies

Two case studies were carried out within large research dominated HE libraries. By discussing the aims of the OCRIS project with staff at the University of Cambridge and the University of Glasgow, (see Appendix 3) specific issues pertaining to systems development, workflows, the selection of technologies, the agendas of the institutions and the concerns of those working with LMSs and IRs on a daily basis, were made clearer and more meaningful to the project team.

4 days were spent speaking to staff from the Cambridge University Library and 2 days with staff from the University of Glasgow Library. In both instances these discussions were supplemented with follow-up questions and discussions via email.

At the outset of each visit a Microsoft PowerPoint presentation¹ was given explaining the nature and purpose of OCRIS. This detailed some of the information which had been gathered in Workpackages 1 and 2 and outlined to participants what was hoped for from their involvement. The presentation was also used as a structural aid to inform discussion and concluded with specific questions to be answered by those present. Informal discussions with participants (both as a group and with selected individuals) occurred and the OCRIS team were shown relevant documentation pertaining to catalogues, repositories and workflows. To provide further insight, the staff interfaces of both "DSpace@Cambridge" and "Enlighten" at Glasgow were demonstrated for the project team.

5. Outputs and Results

The main outputs and results of the project are:

- A list of IRs and OPACs being used by UK HEIs, identifying (within responding institutions), those which do/do not interoperate/integrate;
- A list of services currently offered or desired by those working within responding HEIs;
- Analysis of a purposive sample of HEI IRs and OPACs which share or duplicate content or metadata;
- Case studies on the socio-political, technical and workflow-related issues pertaining to IR and LMS/OPAC use and development within the libraries of the Universities of Cambridge and Glasgow;
- Recommendations for the development of possible further links between library OPACs and Institutional Repositories, identifying the benefits of such links to various stakeholder groups.

In line with the project Aims and Objectives we also report on:

- The extent to which repository content is in scope for institutional library OPACs, and the extent to which it is already recorded there;
- The potential for improvements in the links (e.g. using link resolver technology) from repositories and/or OPACs to other institutional services, such as finance or research administration.

In the sections below the following are considered:

1. IR and LMS systems software across all 85 institutions in scope for OCRIS
2. Interoperability of the systems described to us by questionnaire respondents
3. Services currently supported by the LMSs and IRs of questionnaire respondents
4. Resource discovery services and tools in place or pending in responding HEIs.
5. Scope overlap of the LMSs and IRs of questionnaire respondents
6. Authority Control
7. Summary of analysis of in-depth studies
8. Summary of main findings of the Case Studies

¹ Available to JISC on request.

5.1. LMS and IR systems and software

Below are details pertaining to the LMSs and OPACs used by the libraries in scope for OCRIS. This information was gathered through desk research.

LMS/OPAC	Vendor	# of installations	Interoperability or integration features ²
Aleph/WebOPAC	ExLibris	16	Open architectures used. Web services based APIs.
Alto/Prism	Talis	22	'Talis Platform' (used by the externally hosted Prism OPAC User Interface) comprises web services using standard protocols.
Heritage OPAC/Heritage	ISOxford	5	None detailed.
Horizon/HiP (discontinued).	SirsiDynix	1	None detailed.
Millennium/UNICat	Innovative Interfaces, Inc.	2	Web services to share data; Service Oriented Architecture; support for record import from other systems; RSS feeds. XML harvester.
Millennium/WebPAC Pro	Innovative Interfaces, Inc.	15	As above.
Millennium/INNOPAC	Innovative Interfaces, Inc.	1	As above.
Unicorn/iLink (Unicorn is now known as Symphony).	SirsiDynix	14	Web services API for use with Unicorn modules.
Voyager/WebVoyage	Ex Libris	11	Web services based APIs. Open architecture used.
CALM for Archives	Axiell	3	None listed.
Cheshire for Archives	Berkeley and the University of Liverpool	1	Fully documented API; SRU, OAI-PMH interfaces.

Table 1 - Interoperability and integration enabling features of LMSs used by in scope HEIs.

Not all of these systems are being developed with a clear and comprehensive focus on interoperability.

While the notion of linking library systems is now accepted, integration between IRs and LMSs is generally being built from scratch into new "vertical search" platforms and products, enabled by the use of Web Services and APIs within the core LMS. This is part of a "de-coupling" of systems, with a focus on the end user; these Resource Discovery Platforms (RDPs) can be placed "on top" of the LMS, supplementing or replacing the "traditional" OPAC. The interoperability/integration services of some well known examples are summarised below:

RDP	Vendor/Developer	OAI-PMH and interoperability features
Encore	Innovative Interfaces, Inc.	Offers OAI Harvesting Services.
Primo	Ex Libris	OAI-PMH, SRU/SRW and Z39.50 support.
AquaBrowser	Medialab	Data connectors to harvest from OAI repositories.
Talis Platform	Talis	Web Services, Z39.50, Marc21, Onix, OAI-PMH, RSS.

² All of these systems have Z39.50 interfaces although recent versions of Talis Prism do *not* support Z39.50 searching.

VuFind ³	Villanova University, Falvey Memorial Library	Provides an OAI content provider service to allow for harvesting of records.
---------------------	---	--

Table 2 - OAI-PMH and interoperability focussed features of RDPs used by in scope HEIs.

In the chart below, the continuing dominance of EPrints as the IR software of choice (with DSpace a distant second) within the 85 institutions and across their 115 IRs is clearly visible:

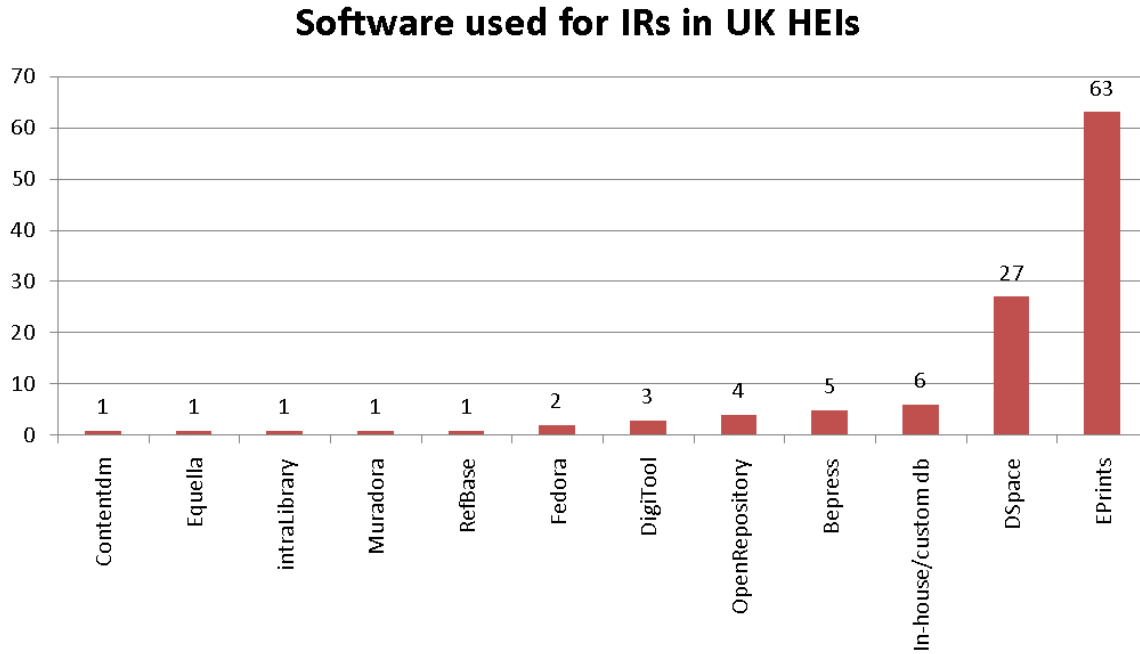


Figure 4 - The various software packages used by UK HEI Institutional Repositories.

³ VuFind is an Open Source platform and offers many of the same functions as the other RDPs listed above but it is entirely modular and customisable. In the United States, it is already being successfully used as an interface for the Union Catalogue of Minnesota State Colleges and Universities on top of an Aleph LMS. See: <http://plus.mnpals.net/> for more information. [Accessed 30th September 2009].

Although not currently used in any UK HEIs identified by OCRIS, another example of an Open Source platform which enables combined access to LMS, IR and other records is the 'Blacklight' OPAC, currently used by the University of Virginia library. See: http://projectblacklight.org/?page_id=2 for more information. [Accessed 30th September 2009].

5.2. Interoperability

The following sections are based on responses to the OCRIS questionnaires. Totals were as follows:

Questionnaire	IR	LMS	Admin systems
Number of individuals responding	35	19	7
Number of institutions responding	31	16	7
Percentage of total institutions contacted	41.2 percent	18.8 percent	8.2 percent

Table 3 - Number of responses received to the OCRIS questionnaires, by group.

Administrative departments responding, by role were:

Department	Number responding to questionnaire
Information Systems and Services	1
Research Support	2
Research Assessment	2
Finance	1
Internal Audit	1

Table 4 - Number of responses received to the OCRIS admin systems questionnaire, by department.

Whilst undoubtedly the IR community found the aims of OCRIS more relevant to their own attitudes and priorities than did other stakeholders, the active participation in the project by professionals working with other institutional information systems indicates *some* wider awareness of the issues and benefits of interoperability.

However, given the low percentage responses for the non-IR stakeholders contacted, many of the figures and charts below should be taken to be illustrative rather than representative.

5.2.1. Administrative Systems

When asked, "Do your systems use Open Standards?"

- 2 respondents answered Yes
- 1 answered No
- The remaining 4 answered Unknown.

When asked, "Do you interoperate with the LMS?"

- 3 respondents answered No
- 2 answered Yes
- 2 said Not Known.

In response to the question, "Do you interoperate with the IR?"

- 3 answered Yes
- 1 No
- 2 Not Known
- 1 pending.

This is illustrated below:

Admin Systems interoperating with IRs and LMSs

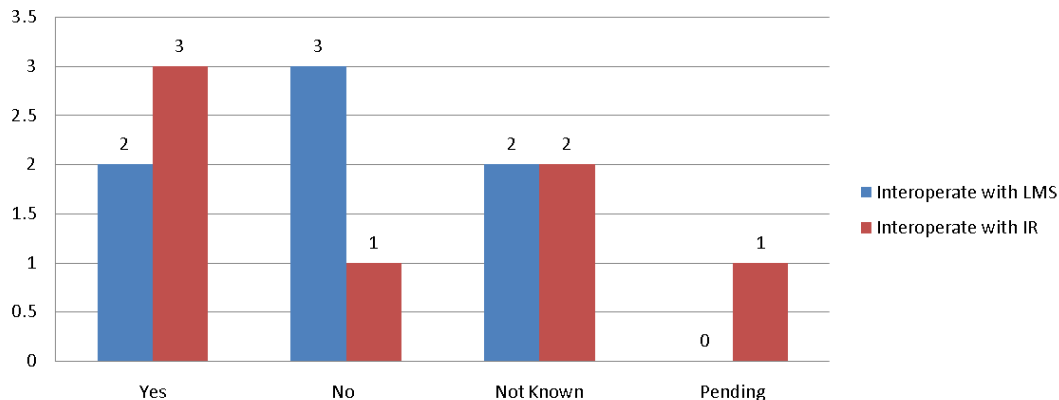


Figure 5 - Answers given to OCRIS' questions about interoperability of admin and library systems.

Of those whose systems interoperate with the IR: 1 works within Research Assessment; 1 within Information Systems and Services; 1 within Internal Audit; and 1 (the respondent who answered "pending") within Research Support.⁴

Of those whose systems do interoperate with the LMS, 1 works within Information Systems and Services, the other with Internal Audit.

The greater number of systems interoperating with IRs *may* suggest that administrative departments find the information held in IRs to be more relevant or beneficial to the work carried out in conjunction with their own systems, than they do the information contained in LMSs. However, far more data would need to be gathered in order to determine if this observation is accurate.

No respondents said that they found improving interoperability to be an issue of no importance; but there was some uncertainty. 4 respondents answered "Yes" when asked "Is improving interoperability important?" and 3 selected "Not Known".

One respondent stated: *"We are currently reviewing our research information management systems. We are looking at both buying an off-the-shelf product that would allow interoperability between all of our systems, and at the feasibility of updating our current system to allow interoperability. Our aim is to have a complete, fully integrated research information system by the end of next year."*

Another wrote: *"We currently have an in-house Research Expertise system used for research admin/assessment which links to our IR and other corporate systems as indicated. We are about to (i.e. have chosen supplier and aim to implement by end of the year) replace this with a proprietary system using CERIF as the core data model and CERIF-XML as the exchange format between systems [www.eurocris.org]. We harvest data from 3rd party sources such as ISI and PubMed for journals and similar output; but we manually input metadata on books; so what would be useful would be authoritative source(s) of metadata for non journal items that could be harvested in CERIF-XML; similarly it would be useful to have access to authoritative lists of journals and publishers in this way."*

Enthusiasm and knowledge about interoperability is clearly being demonstrated; this highlights the need for future work to be done allowing these issues to be investigated in more depth, perhaps without specific reference being made to library systems and perhaps focussed on Current Research Information Systems (CRISs) or Research Management Systems (RMSs).

⁴ Note: these department titles were selected from a list of options provided in our questionnaire (of those occurring most frequently within HEIs, according to our desk research) and are not necessarily the official titles of the departments where respondents work.

5.2.2. LMSs and IRs

One of the most significant questions for the OCRIS project was "Do your Library Management Systems and Institutional Repositories interoperate?" Answers were as follows:

Interoperability between LMSs and IRs

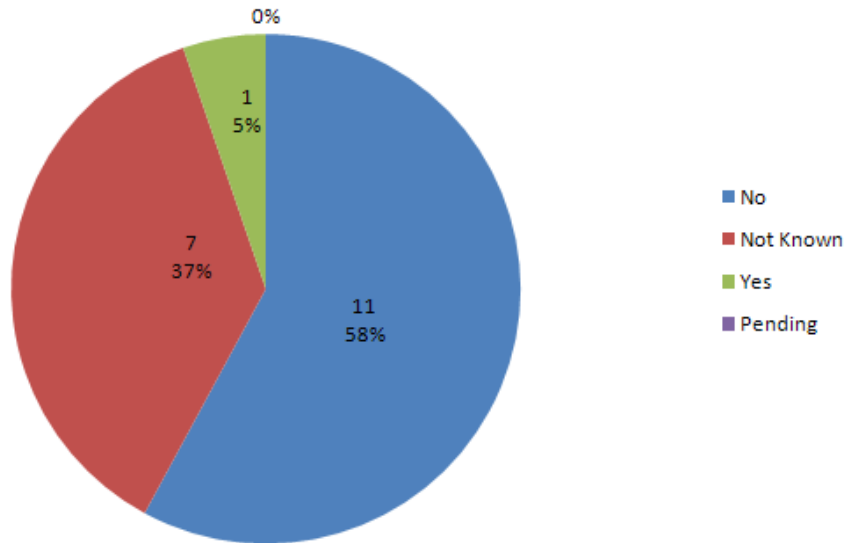


Figure 6 - Responses to the question, "Does the LMS interoperate with the IR?"

Interoperability between IRs and LMSs

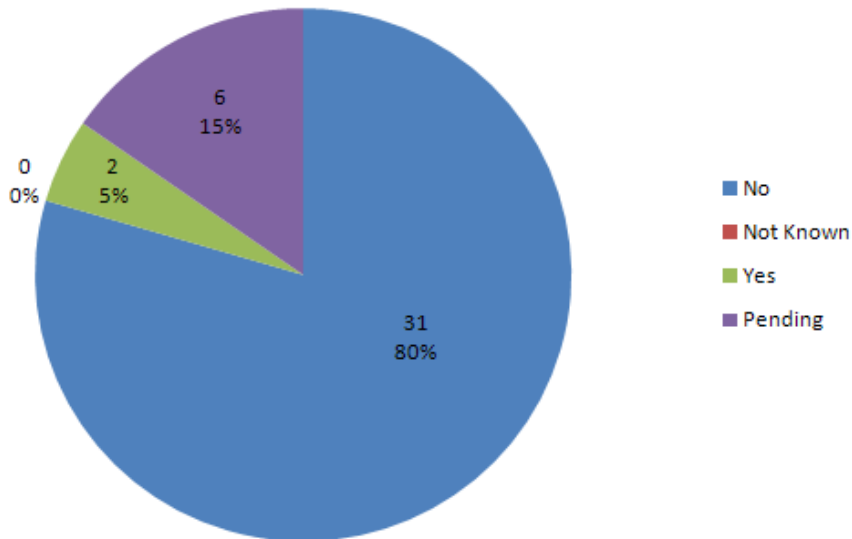


Figure 7 - Responses to the question, "Does the IR interoperate with the LMS/OPAC?"

The totals shown in the above pie charts appear neatly illustrative. However, calculating percentages and numbers for the entire group of respondents (i.e. combining both the LMS and IR questionnaire answers) was complicated by both multiple entries for the same institution and contradictory answers between the two groups, relating to the same systems.

This is illustrated in the table below, which shows answers given to the questions "Does the LMS interoperate with the Institutional Repository?" (Column 2) and "Does the IR interoperate with the LMS/OPAC?" (Column 3). The data which had to be cleaned up or otherwise resolved is shaded.

Institution	LMS Response	IR Response
Sussex, University of	Yes	Pending
Northampton, University of	No	No
Bangor University	Not Known	No
Nottingham, University of	Not Known	No
Nottingham, University of	n/a	No
Nottingham, University of	n/a	No
Lincoln, University of	No	No
Aberystwyth, University of	Not Known	Pending
Warwick, University of	No	Pending
Brighton, University of	No	No
Birmingham, University of	No	No
Birmingham, University of	n/a	No
Birmingham, University of	n/a	No
East Anglia, University of	No	Yes
East Anglia, University of	No	n/a
Glasgow, University of	Not Known	Pending
Glasgow, University of	Not Known	n/a
Glasgow, University of	No	n/a
Cambridge, University of	No	No
St Andrews, University of	Not Known	Pending

Table 5 - Conflicting questionnaire responses.

Devising a "resolving" methodology based on fair supposition, it was decided that the 9 instances⁵ of contradiction would form a new category called "Contradictory answers". To achieve this, firmer answers were privileged over vague ones (i.e. No over Not Known; Yes over Pending) with the vague answer being re-assigned to "Contradictory answers". Where both Yes and No were indicated, both were re-assigned. Pending was selected over No as it seems to indicate greater knowledge. 2 No, 2 Yes and 5 Not Known answers were reassigned in this way. Admittedly there is margin for error.

The introduction of this new category illustrates the difficulties faced when trying to interpret some of the questionnaire responses with precision (which may be an inevitable consequence of asking more than one individual to answer the same question in expectation of the same answer).

⁵ 9 despite only 8 pairs of conflicting answers because of the case of East Anglia where both a Yes and No answer were included in the set.

The table below summarises the two ways of looking at the data described above:

Answer	Interoperability across whole group (totals for each with no correction)
No	42
Not Known	7
Yes	3
Pending	6
TOTAL	58
Answer	Interoperability across whole group (new category added)
No	40
Not Known	2
Yes	1
Pending	6
Contradictory	9
TOTAL	58

Table 6 - Interoperability of LMSs and IRs across the group of HEIs - both before and after the introduction of a new category to resolve conflicting data.

The corrected total (with the introduction of our new category) gives the following pie chart:

Interoperability across whole group (correction attempted and new category added)

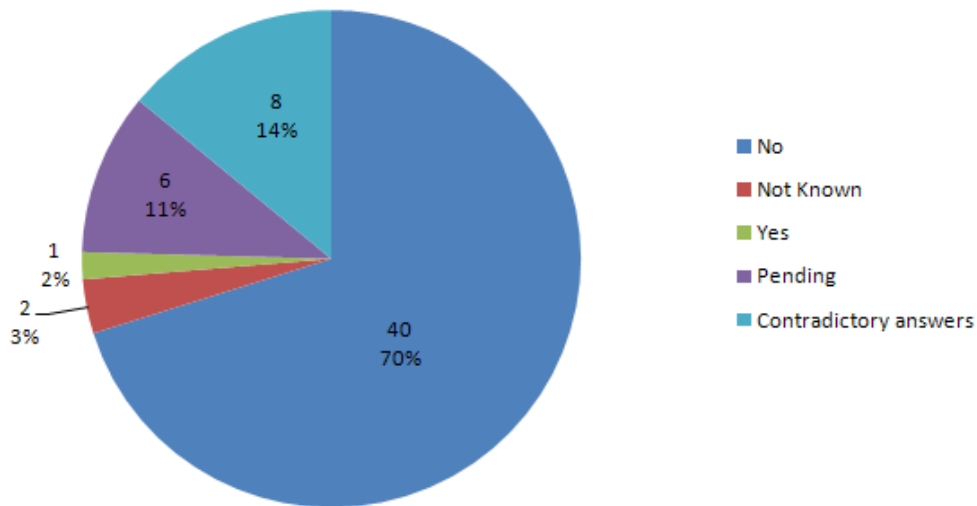


Figure 8 - Interoperability between IRs and LMSs.

Even disregarding the 14 percent of responses which reveal the conflicting perceptions of staff on this issue and even with some margin for error, it is clear that the vast majority of systems do NOT interoperate with one another. Only two percent of respondents answered in the affirmative. The responses of those who answered that interoperability was "pending" might be added, giving the still low figure of 8 percent.

Which *other* institutional systems do LMSs and IRs interoperate with? The response was modest yet diverse. Results are presented below:

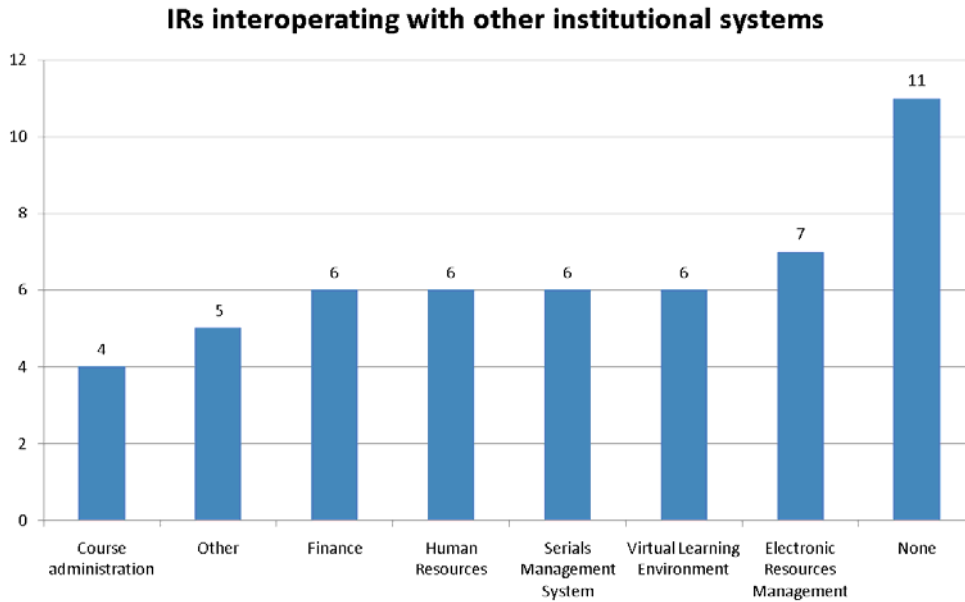


Figure 9 - Various institutional systems with which IRs interoperate.

Although 31 institutions responded to the IR questionnaire, only 20 of those (64.5 percent, each representing 1 IR) interoperate with other systems; the rest indicated that they currently interoperate with none. Of those 20, 7 interoperate with only 1 other system, 6 with 2, 6 with 3 and 1 with 4.

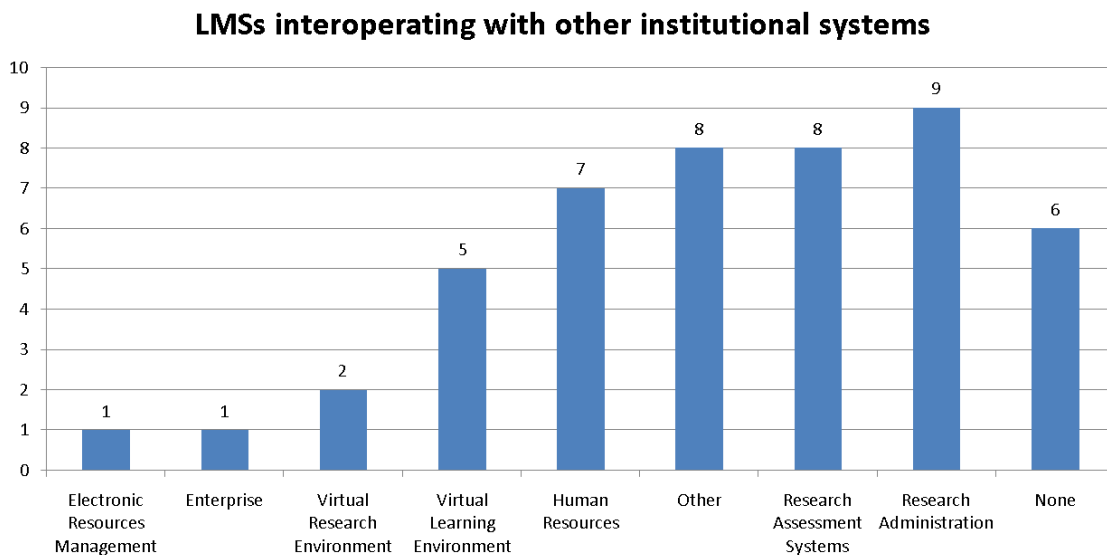


Figure 10 - Various institutional systems with which LMSs interoperate.

Similarly, 13 of those responding to the LMS questionnaire out of 19 (68.4 percent) stated that they currently interoperate with other institutional systems. Of those 13, 6 interoperate with 4 other systems, 3 with 3, 1 with 2 and 3 with 1.

5.3. Services currently supported

As previously noted, the list of services used in the OCRIS questionnaire was derived from a recent JISC report - *Linking UK Repositories* (Swan and Awre, 2006) - which considered the core aspects of a user-needs based national repository infrastructure tailored around the requirements of a variety of

key stakeholders, including "end users (researchers across all disciplines, teachers and learners); research administrators, employers and funders, intermediaries working on behalf of any of these groups, and institutions or other bodies with repositories" (page 17).

This perspective fits extremely well with that of the OCRIS project. While the ontology proposed in *Linking UK Repositories* extends beyond the boundaries of HE, looks at candidate services in greater depth than does OCRIS, and proposes a layered service model, it nevertheless gave the services section of the questionnaires strong theoretical grounding; not least as resource limitations meant the OCRIS team were unable to consult directly with the range of end users listed above.

5.3.1. Institutional Repositories

The services currently supported by IRs are as follows:

Service	Number of IRs supporting service
Providing advice on Open Access	33
Manual metadata creation and enhancement	30
Providing advice on IPR	27
Usage statistics	27
Providing technical help	21
Reporting for use by IR Manager and associated staff	17
Access and authentication	17
Verification of metadata	15
Verification of the digital object itself	14
Verification of content (e.g. document format)	13
Preservation	12
Collaborative or cross-Institutional Repositories at national level	10
Citation	10
Other statistics services (e.g. subject coverage; number of authors)	10
Automated metadata creation and enhancement	9
Seeking submissions from staff/students not currently included in scope	9
Digitisation	8
Collaborative or cross-Institutional Repositories at international level	7
Reporting for specific administrative department (e.g. R&I)	7
Research and quality assessment (other than REF)	6
Repository construction or hosting	5
Bridging and mapping	5
Research Excellence Framework (REF)	5
Publishing	4
Meta-analysis	3
Subject specific repositories gathering primary content	2
Technology transfer/business advice	2
Other	1
Format specific repositories gathering primary content	1
Interim repositories for authors with no institutional repository	1
Media-specific repositories gathering primary content	1

Table 7 - Services currently supported by IRs.

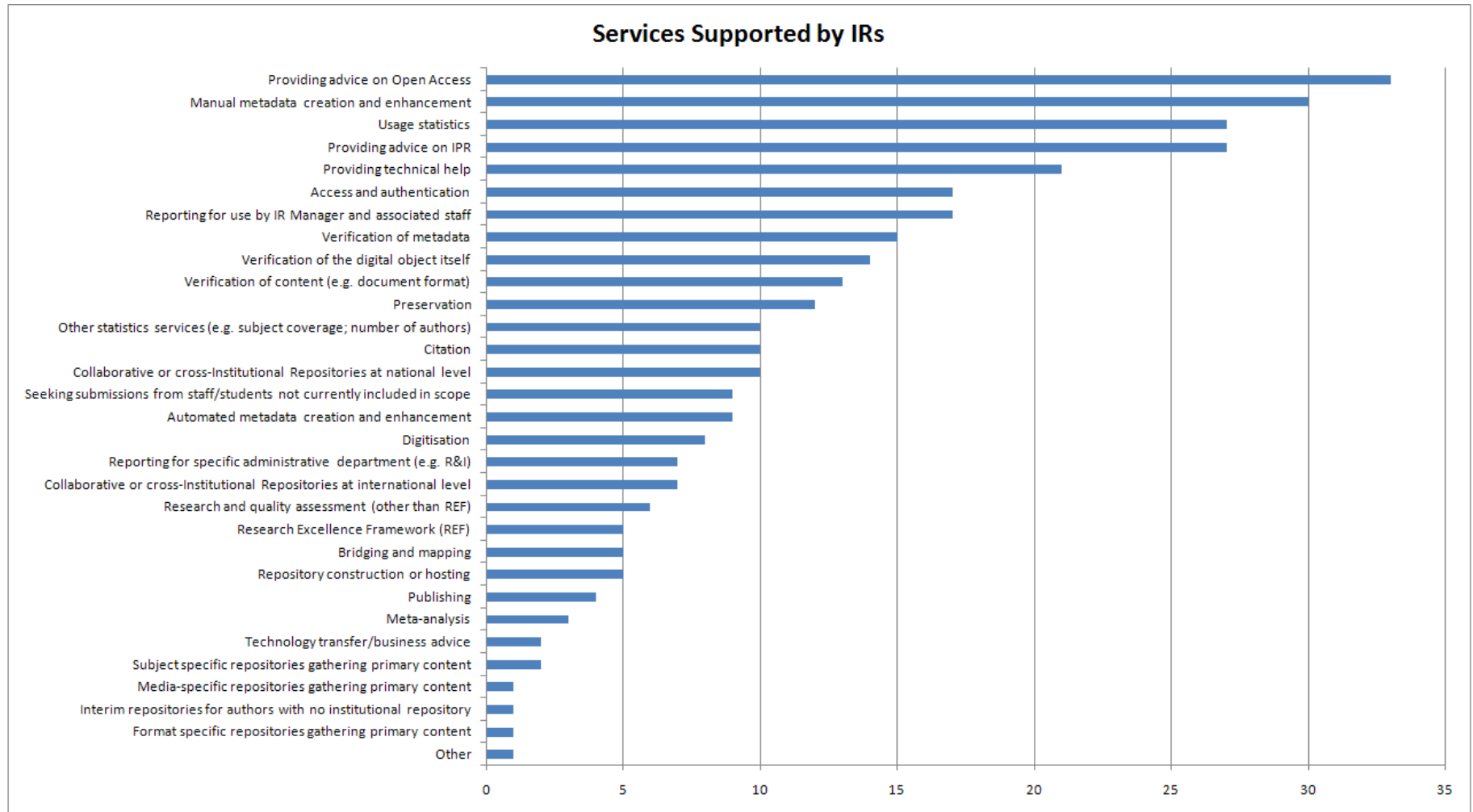


Figure 11 - Services currently supported by IRs.

Project Acronym: OCRIS
 Version: 1
 Contact: Kathleen Menzies or Gordon Dunsire
 Date: August 2009

Services pending for IRs

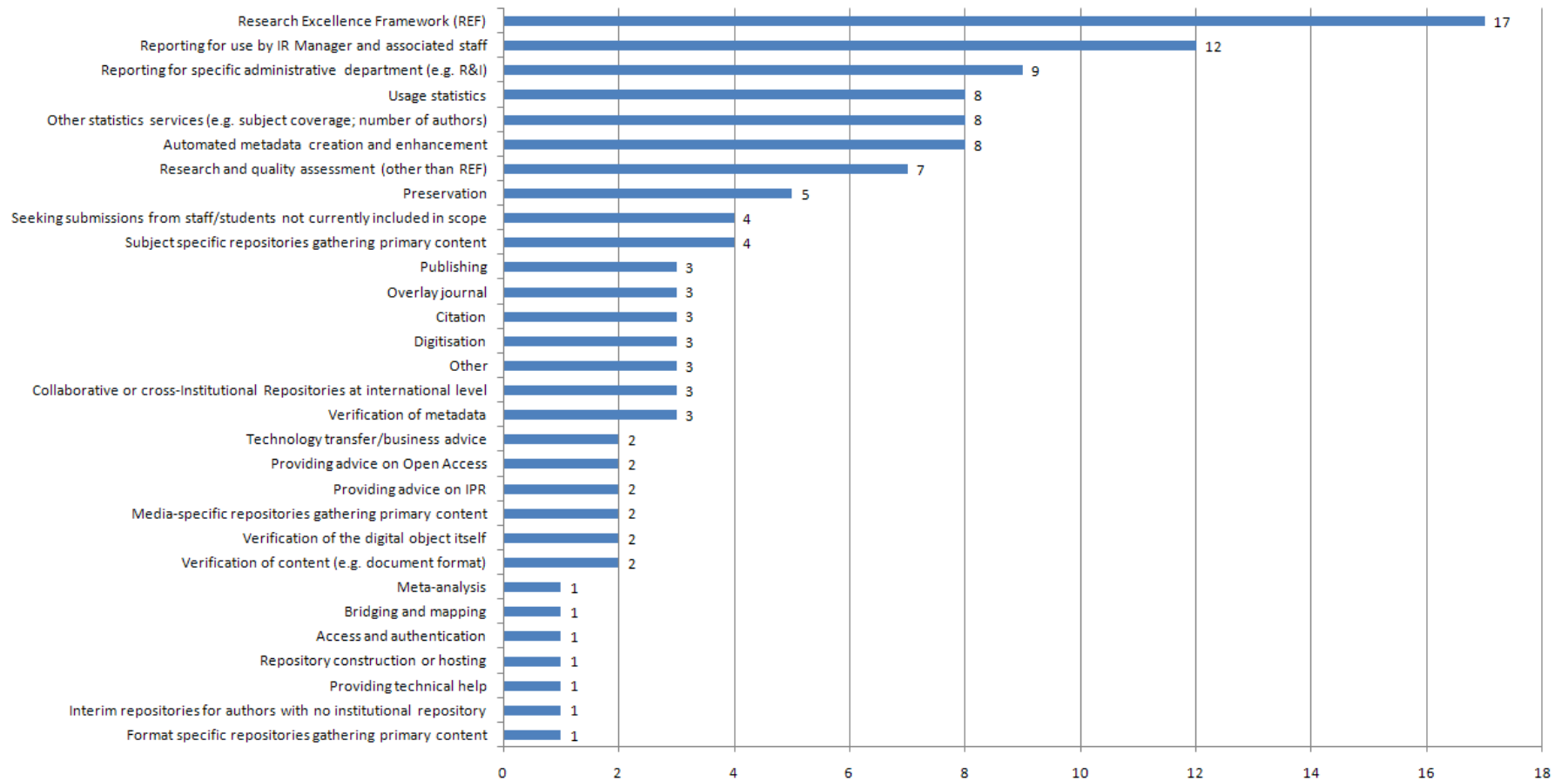


Figure 12 - Services which will be supported by IRs.

5.3.2. Library Management Systems

The services currently supported by LMSs are as follows:

Service	Number of LMSs supporting service
Usage statistics	14
Manual metadata creation and enhancement	12
Verification of metadata	8
Access and authentication	8
Citation	5
Other statistics services (e.g. subject coverage)	5
Automated metadata creation and enhancement	4
Verification of content (e.g. document format)	4
Verification of the digital object itself	2
Digitisation	2
Reporting for specific administrative department (e.g. R&I)	2
Reporting for use by Manager and associated staff	2
Other	1
Bridging and mapping	1
Collaborative or cross-Institutional Repositories at international level	0
Collaborative or cross-Institutional Repositories at national level	0
Format specific repositories gathering primary content	0
Interim repositories for authors with no institutional repository	0
Media-specific repositories gathering primary content	0
Subject specific repositories gathering primary content	0
Providing advice on IPR	0
Providing advice on Open Access	0
Providing technical help	0
Repository construction or hosting	0
Meta-analysis	0
Preservation	0
Publishing	0
Research and quality assessment (other than REF)	0
Research Excellence Framework (REF)	0
Seeking submissions from staff/students not currently included in scope	0
Technology transfer/business advice	0

Table 8 - Services currently supported by LMSs.

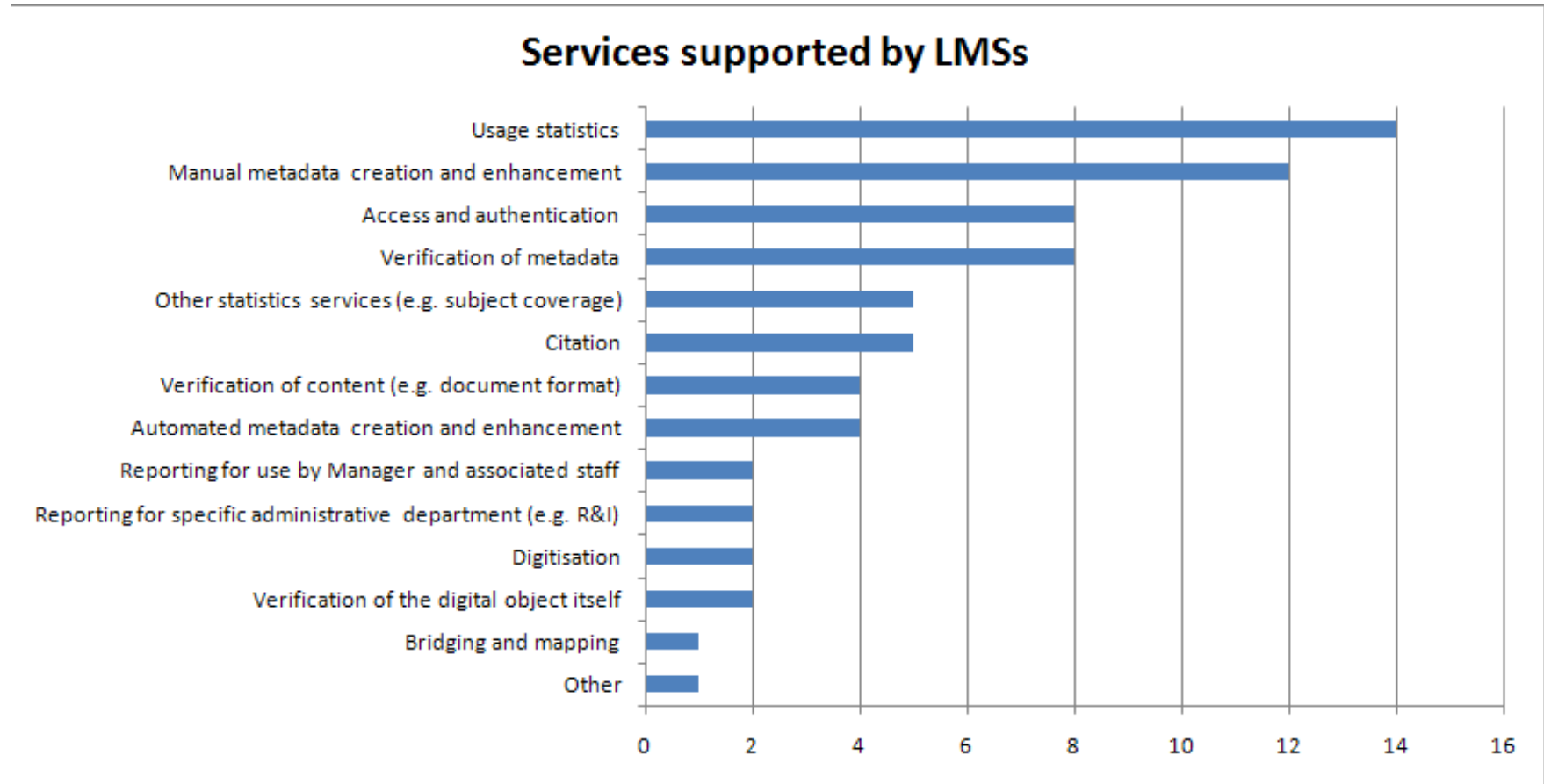


Figure 13 - Services offered by LMSs.

5.3.3. Key findings regarding services offered by LMSs and IRs

It is particularly noteworthy that for both LMSs and IRs, manual metadata creation and enhancement services (63 and 85 percent, respectively) are the second most frequent type of service *currently* supported after the generation of usage statistics (73 and 77 percent). Reporting services for specific administrative departments is uncommon but twice as frequent in IRs as LMSs (10 percent, 20 percent). Providing advice on IPR (77 per cent) and Open Access is, as might be expected, extremely common (94 per cent) within IRs.

Stripping out IR-centric services clearly not appropriate to LMSs, it becomes clear that only a small number of LMSs are beginning to incorporate services which go beyond what might be considered their "traditional" role - such as citation analysis, digitisation and the verification of digital objects.

Basic services, such as arranging for the provision of adequate, high-quality metadata and monitoring performance, clearly remain more significant parts of the workflow within both IRs and LMSs than do other added-value services.

Although we cannot ascertain from the data exactly how "pending" other services are (i.e. will they be available within the repository soon or are they merely "ideals"?) it is immediately visible how major an influence planning for the REF currently is within UK HE. 17 out of 26 institutions (65 percent) plan to offer services in support of the REF. 3 of the 9 remaining from this 26 already fulfil this function; thus 77 percent of UK HEI repositories responding to the OCRIS questionnaire support or plan to support activities relating to the REF.

There is also some desire to establish "reporting services" for IR managers (46 percent) and other departments (34.6 percent), to generate usage statistics (31 percent) other statistics (31 percent) and to automate metadata enhancement (31 percent).

5.4. Integration services and tools

OCRIS used the term "integration service" as shorthand for some of the services becoming popular - or already - popular within the library systems community, such as metasearch link resolvers and software as a service (SOAS).

Integration services in place or pending in IRs

Integration Service	Number of respondents using or planning to use integration service
Metasearch/Linking	7
Software as a Service	1
Vertical Searching	4
Web Service/API	10
Other	4

Table 9 - Integration services in place or pending in IRs.

Integration services in place or pending in LMSs

Integration Service	Number of respondents using or planning to use integration service
Metasearch/Linking	13
Software as a Service	3
Vertical Searching	4
Web Service/API	7
Other	2

Table 10 - Integration services in place or pending in LMSs..

Use or planned use of Integration Services

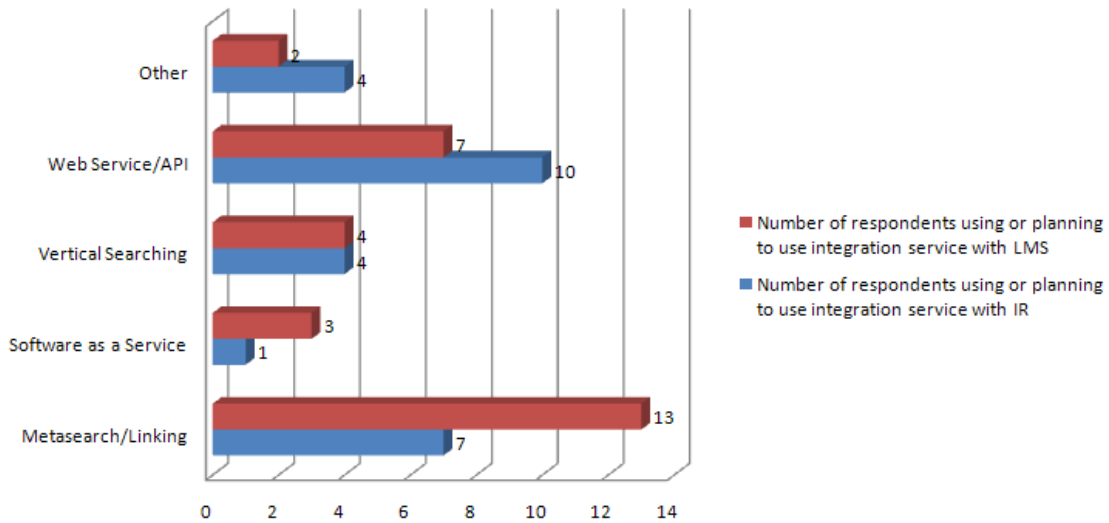


Figure 14 - Use or planned use of what OCRIS terms "integration services" within LMSs and IRs.

LMSs make greater use of metasearch and link resolvers than do IRs; as discussed further below, some IR developers are starting to realise the potential of using existing library link resolver knowledge databases within their own systems.

Vertical search is something currently being marketed heavily at the LMS community by vendors, and is much discussed; it is a positive sign that this technology is also being considered by those working with IRs but perhaps surprising that not many respondents overall are using or planning to use vertical search (only 17 percent of the total institutions responding) at present.

Web services and Application Programming Interfaces (APIs) are more popular (36.2 percent of institutions) – but what are these being used for? Investigating this in more depth with follow up questions would be enlightening and would allow better understanding of the gathered data.

5.5. Scope Overlap

The Venn diagram below illustrates the substantial amount of "overlap" (81 percent) between items which respondents listed as being in scope for their OPACs and IRs. The inner circle shows the set of items common to both groups, the outer one those *only* in scope for IRs.

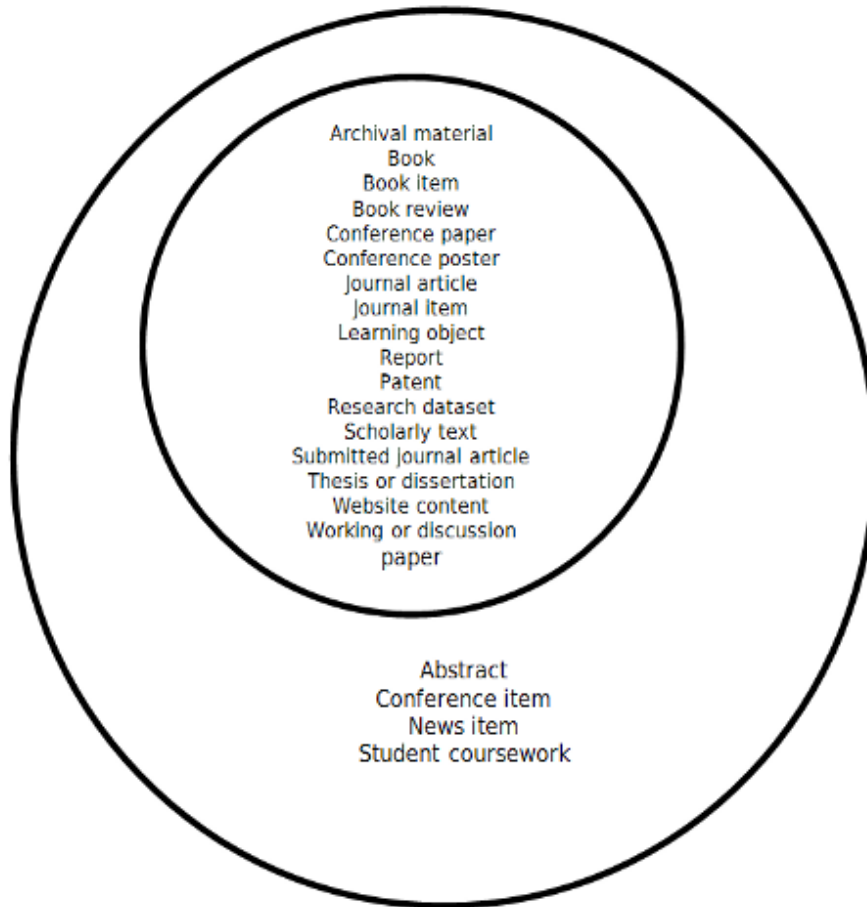


Figure 15 - Venn diagram showing scope overlap between IRs and LMSs in HEIs.

The significant finding here is that one set is *completely contained* within the other; there is not merely *overlap* – every item type in scope for the OPACs of UK HEIs are now also in scope for IRs. There are only 4 item types unique to IRs.

It is worth qualifying this with an understanding that scope overlap may only be partial - in many instances, an IR will hold a full-text item while the OPAC will hold only *metadata* about that item, for example, in the case of patents or theses. However, the scoping distinctions and boundaries for IRs and OPACs are becoming increasingly blurred, with many IRs containing bibliographic data and OPACs linking to full text. Thus instances of both partial and full scope overlap are significant.

Some respondents found using the Scholarly Works Application Profile (SWAP) (UKOLN and JISC CETIS, 2008) classification of item types problematic (although guidance was provided on the OCRIS website) as SWAP is not currently supported widely by either LMSs or IRs thus the language it uses is not familiar. The fact that "Conference items" are unique to IRs while "Conference papers" and "Conference posters" are not, may suggest that this data is unreliable.

This is discussed further in the Conclusions and Recommendations.

4 respondents state that they plan to extend the scope of their OPACs; the items currently found to be only in scope for IRs are, perhaps, likely to be recorded in some LMSs soon. 16 respondents to the IR questionnaire state that there are plans to extend scope for their systems.

5.6. Authority Control

The following charts illustrate the classification, subject and name authority schemes being used by the LMSs and IRs of those responding to the OCRIS questionnaire:

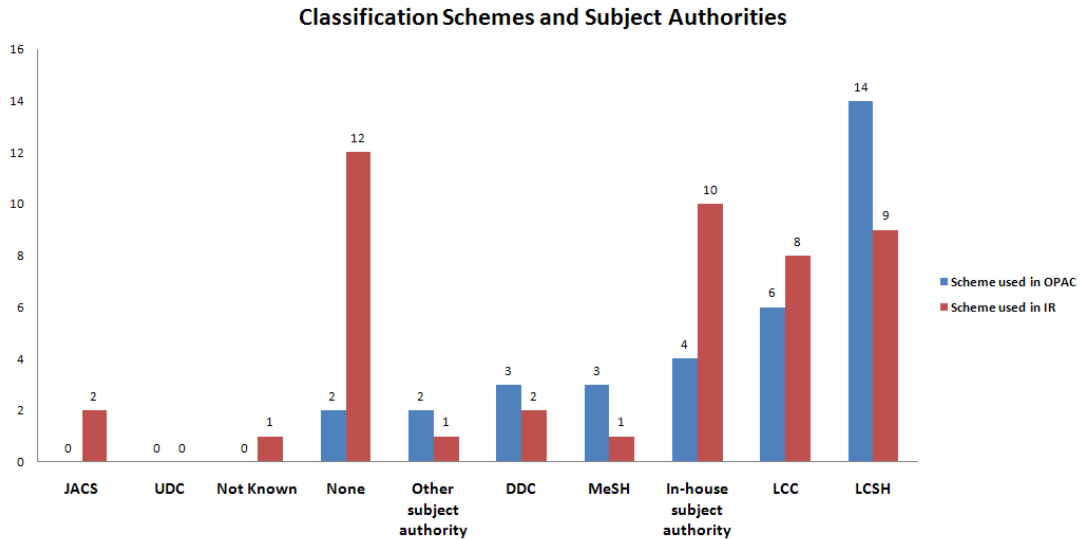


Figure 16 - Classification and subject schemes being used within IRs and LMSs.

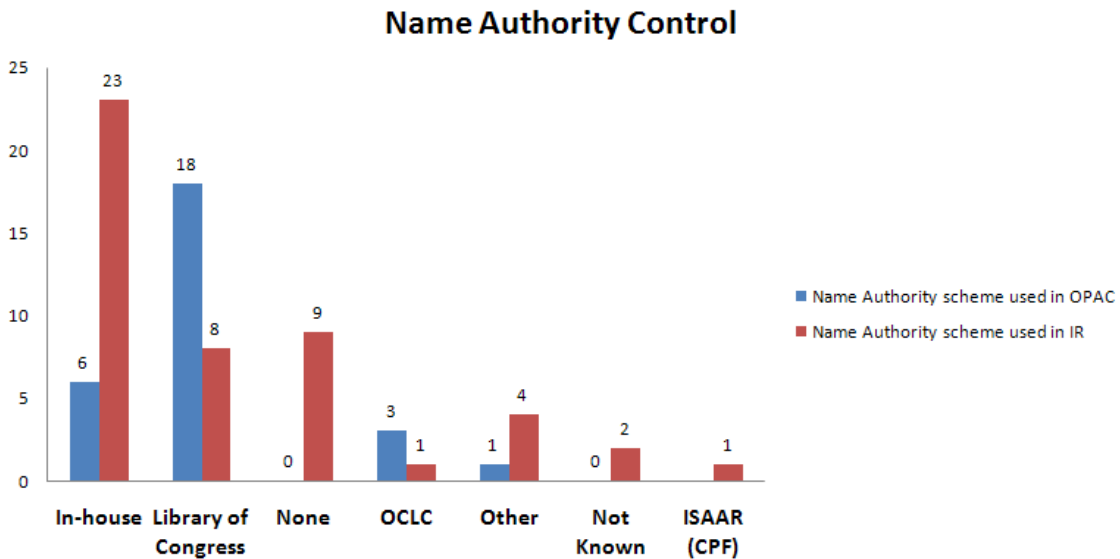


Figure 17 - Name authority schemes being used within IRs and LMSs.

The comparative lack of subject authority control within IRs as opposed to LMSs and the greater extent to which in-house name authorities are being used in IRs is clearly significant in terms of semantic interoperability and the sharing of metadata. Library of Congress schemes remain fairly

popular compared to others. Top-level LCC comes bundled with most versions of EPrints, explaining its comparative frequency within IRs.

Regardless that mappings between many LCC numbers and corresponding LCSH terms are available (despite no intrinsic relationship), it must be stated that there appears to have been some confusion amongst our respondents about LCC and LCSH; some of those responding stated that they use LCSH for subject control, yet (in a few instances) when visiting their repositories, no evidence of this was found – rather, LCC was the standard being used and implemented in "browse by subject" menus.

6. Summary of analysis of in-depth studies⁶

6.1. Introduction

IRs and OPACs from 9 HEIs and 1 library serving 3 research organisations were selected for in depth investigation in relation to scope overlap, the application and presentation of standards and duplication at both record and item level.

- Cambridge and Glasgow acted as the OCRIS case study institutions within Workpackage 3 thus they were selected for inclusion.
- Hull is unusual in using Fedora as the basis for its IR, combined with a Muradora interface.
- Southampton is notable for having many Institutional Repositories (10 in total), many of which pertain to specific subjects or disciplines.
- The National Marine Biological Library was selected as an interesting example of an inter-disciplinary but subject-specific organisation with an IR and catalogue relevant to HEI research. This allowed OCRIS to explore the project's aims and objectives beyond the boundaries of a single University.
- Strathclyde was an early advocate of Open Access repositories and is the "home" library of the OCRIS team. Being able to talk directly to staff involved in the development of its IR was beneficial.
- The University College of the Creative Arts administer the Visual Arts Data Service (VADS), an interesting example of a format and discipline specific repository, using custom software.
- The metadata provided for bilingual interfaces was considered in relation to the 3 HEIs which contribute to the Welsh Repositories Network (WRN): the Universities of Aberystwyth, Cardiff and Glamorgan.

All 10 are discussed more fully in Appendix 4, where a range of illustrative screenshots, and the URLs for the relevant IRs and OPACs, are provided. Here, findings have been condensed to highlight the most typical, or unusual, examples.

6.2. Main findings

The in-depth investigations unearthed significant scope overlap across the group for both theses and journal articles. To qualify: that links to full-text journal articles are often available via the LMS (through a combination of subscription databases and link resolvers) effectively constitutes *item level* duplication, when copies of those same articles are held in the IR; while the full-text of theses may be available only in the IR, when the bibliographic data is also available in the OPAC, overlap occurs at *record level*.

As well as an apparent lack of "joined up thinking" in cataloguing workflows, there are inconsistencies and a lack of precision with statements made on IR and OPAC webpages about the types of item in scope.

The inconsistent application of authority control (or indeed, a complete lack of authority control) in IRs is prevalent, leading to difficulties when trying to find all publications by a specific author or outputs from a particular department or research group by searching across OPACs and IRs. The same

⁶ All information taken from websites was correct at the time when these studies were carried out (July/August 2009). Inevitably some information will have changed since that time.

is true when trying to find IR items according to subject. Inadequate or incompatible metadata is generally the root cause of this fragmentation.

Resource Discovery Platforms (RDPs) - or, "vertical search" applications - at times reveal inadequate metadata and present search results in ways perhaps not best suited to the scholarly needs of end users. Although the data harvesters built into these products are useful for searching across systems (and accordingly, "linking" their items and records), they are, at present, only a partial solution to the problems posed by limited interoperability.

6.3. Descriptions of scope

Item types (variously termed "Format", "Medium" or "Material Type") are usually only viewable via "Advanced Search" menus and are generally derived from fields inbuilt into the system software (e.g. MARC or Dublin Core encoding). In LMSs these fields are often modified to include administrative information relating to circulation control or item location. This means there is little commonality in how items are described across HEI OPACs and scope is difficult to accurately discern, especially as scope may extend beyond the items *currently* catalogued thus will not be revealed by indices of fields already within. Further, the item types listed may not correspond to how they are described elsewhere on the website and levels of granularity vary.

Usually there are no readily accessible statements of scope (either general or detailed) provided on OPAC or IR webpages and sometimes none at all. Users have to rely on making use of the search and browse options, setting advanced search delimiters to ascertain what item types are available to them.

Similarly, repositories may use default fields (such as "Dublin Core Metadata Initiative Type Vocabulary" terms or derivatives thereof) to list what is in scope. In fact, the repository may not seek to restrict the type of format/content in scope - unless it is clearly instantiated along format or subject lines (for example, chemical datasets; visual arts images) where explanations of scope tend to be clearer. Item types listed may simply be those supplied with the repository software thus do not convey an accurate picture of what items are present, prevalent or actively sought by the repository.

While the use of these fields in the construction of menus is intended to help end users limit their searches, the lack of shared vocabularies or standards means that there is little correspondence between item/material/format types listed in OPACs and IRs, thus when searching both systems independently it can be confusing trying to discern what the commonalities and differences are.

To take a typical example: at **Cambridge University Library (CUL)**, items searchable in the OPAC are given as:

- Book
- Serial
- Electronic journal
- Electronic resource
- Disc (CD/DVD)
- Music Score
- Map
- Non-musical Recording
- Musical Recording
- Archive/Manuscript
- Kit
- Mixed Material/Collection
- Mixed Material
- Visual Material

But naturally the true scope extends beyond this list of 14 and is more complex. Further investigation of the website reveals what is recorded in Newton (the CUL library catalogue) in more detail:

- all printed books published from 1978 onwards, with the exception of Official Publications

- selected Official Publications published since 1999
- printed books published before 1978 considered to be of academic importance at the time of acquisition
- all print journals
- all electronic journals
- atlases published after 1977
- maps catalogued since August 2000
- sheet music and recorded music catalogued after 1990
- microfilms and microfiches published after 1977
- audio-visual material published after 1977
- music manuscripts

All printed books and journals in the 4 CUL dependent libraries are also recorded.

A further additional note on scope states that "Coverage of books published prior to 1978 in the University Library is incomplete. Books published prior to this date, and considered at the time of acquisition to have been of secondary academic importance, are not included in Newton...[]..Coverage of music manuscripts in this catalogue is complete, and where such manuscripts contain two or more pieces, there are in most cases catalogue entries for each piece. This catalogue does not include theses or non-music manuscripts, which are in a separate Newton catalogue, University Library Manuscripts and Theses." (University of Cambridge, 2009a).

DSpace@Cambridge lists the following item types as being in scope:

- Article
- Audio
- Book or Book chapter
- BW Image
- Colour Image
- Dataset
- Drawn image
- Image
- Journal Article - Published Version
- Journal Article - Submitted Version
- Map
- Other
- Preprint
- Presentation
- Software
- Table
- Technical Report
- Thesis

These are derived from recognised "dc.item" terms instantiated within the repository; but similarly to the OPAC, the scope is in actuality broader than this, with all types of material being sought.

Item types may not be described consistently across library webpages, particularly when you move between homepage and search interface. At **Glasgow University** we found that "Published conference papers" became "Conference Proceedings" while "Book chapters" became "Book Sections". Specificity (granularity) accordingly also becomes mixed, with "peer-reviewed journal articles" the description used on some pages and the more general "articles" on others.

Terminology is possibly being tailored to suit the function of the webpages on which these listings appear, becoming, for example, more specific, the "deeper" into the repository a user moves. However, this may reduce clarity for users unfamiliar with the terminology.

Resource Discovery Platforms add another level of complexity: while theoretically capable of bringing together items from across systems, they tend also to use the MARC "item type" fields to structure faceted menus and search results lists. Where metadata is inadequate or limited, this is then revealed to users. At **Glasgow**, the following material types are listed in the OPAC's advanced search options (note how "Not Coded" is revealed):

- PRINT BOOK
- E-BOOK
- PRINT JOURNAL
- GU EXAM PAPER
- GU THESIS
- MUSIC SCORE
- MAP/ATLAS
- VIDEO
- SOUND
- EXTRACT
- E-EXTRACT
- KITS
- FINDING AID
- ARTWORK
- PHOTO SLIDE(S)
- NOT CODED
- DATABASE
- SOFTWARE
- WEBSITE
- E-SCORE
- E-MAP
- E-THESIS
- E-VIDEO
- E-SOUND
- DATABASE (CD)

Using **Glasgow's** "Quicksearch" service (which uses the Encore Resource Discovery Platform), the list of formats and collections being searched only appears *after* you have entered a search term in an advanced search field; these properties cannot then be specified to limit initial searches, only to *refine* them. As only those containing items matching a search term are recorded ("Search Found in"), it is hard to discern what the *full* range of these formats and collections are.

There is little commonality across HEIs in the description of IR or OPAC scope.

The **University of Hull** and the **University of Glasgow** both use Millennium as their LMS with WebPAC Pro as the OPAC. MARC is the metadata format standard supported by both yet the "material types" listed by both OPACs are quite different. In Hull's case these are:

- ARTEFACT
- BOOK
- BOOK WITH DISK
- CASSETTE
- CD AUDIO
- CHARTS
- COMPUTER FILE
- DVD
- EBOOKS
- EJOURNALS

- KIT
- MICROFORM
- MS MAP
- PERIODICAL
- PLAYBILL
- PRINT MAP
- PRINTED MUSIC
- SLIDE
- SOUND RECORD
- SPOKEN RECORD
- TAPE WITH BOOK
- THESIS
- VIDEO
- XEROX

Naturally these differences in terminology can be seen to reflect - to an extent - the differing collections of each library and the item types chosen for inclusion; however some terms presumably refer to the same type of item (e.g. print map and map/atlas; book and print book; e-book/ebook; sound/sound record).

When details given are specific to the administrative set-up of the individual library (e.g. "CD-ROM - 3 day loan" or "Fine Art Store - 1 day loan"), this provincialises the description; when terms are too vague - for example, "Academic research output", "Artefact", "Teaching materials" and "Computer files" - it can be hard to establish equivalence.

There is more commonality in description among IRs because, whether EPrints or DSpace is being used, often the same default "dc.type" and "dc.item" values are instantiated. Again, however, these are often modified to match other DC-based metadata standards such as DRIVER or EThOS, or to reflect local needs.

There is an inherent tension in how item/format types are described - if standard fields deriving from cataloguing schemes or repository software are used, this can assist in keeping simple the construction of search/browse lists, allowing search/advanced search limiters to be implemented fairly easily. As long as different terms used by different systems are referred to in the same way by the underlying code (for example, a numeric ID) *semantic* interoperability will be guaranteed. However, uniformity for the end user will only be ensured if item types are referred to in the same way at the *presentation level*. Describing item types in more detail, to reflect the collections of specific libraries or repositories, is clearly beneficial to end users, but if basic item types are not the same, there is no direct matching and this may complicate the search experience. Consistency and clarity for end users searching both within and across institutional systems would, at present, require mapping or conversion exercises.

Independent of software package used, *statements* made by IRs on scope/item type are often all-purpose or "boilerplate", offering only minimal information on the range of material types and versions deposited. For example, a scoping statement that originated on OpenDOAR (<http://www.opendoar.org/tools/en/policies.php>) is used by more than one investigated repository. This statement suggests that essentially all "all types of materials" *may* be in scope for an IR. End users have to wade through the search options to find what item types are currently included or supported but may not achieve certainty when they do so.

6.4. Overlap

As the Venn diagram on page 22 shows there is a high percentage of overlap (approximately 82 percent) amongst item types which respondents understood to be in scope for their IRs and OPACs.

Here, the report focuses on the "duplication" of both theses and journal articles, the two item types which OCRIS's investigations found to be the most prominent and prevalent areas of overlap.

6.5. Journal articles

There are many readily discerned examples of duplication at both item and bibliographic level, across UK HEIs. Two of these are considered below.

Taking a pre-print titled "Changing gender relations and environmental knowledges in Upper Egypt" from the **Glasgow** DSpace Service repository (intended for working papers, technical reports and other grey literature) Glasgow's other IRs and OPACs were interrogated:

Please use this identifier to cite or link to this item: <http://hdl.handle.net/1905/12>

Title: Changing gender relations and environmental knowledges in Upper Egypt.
Authors: Briggs, John
Sharp, Joanne
Hamed, Nabila
Yacoub, Hoda
Date of Issue: 2002
URI: <http://hdl.handle.net/1905/12>
Item Type: Preprint
Appears in Collections: [Research Group: Human Geography](#)

Files in This Item:

File	Size	Format	
jbriggs001.pdf	83Kb	Adobe PDF	View/Open

This item is protected by original copyright:

[View Licence](#)

[Show full item record](#)

Figure 18 - A pre-print held in Glasgow's DSpace Service.

As a pre-print, you can find this article (under its original title) in the **DSpace** repository (<https://DSpace.gla.ac.uk/handle/1905/12>). As a post-print, the article (under its revised title) can be found in the **Enlighten** repository (<http://eprints.gla.ac.uk/170/>):

Changing women's roles, changing environmental knowledges: evidence from Upper Egypt

Briggs, J. and Sharp, J. and Hamed, N. and Yacoub, H. (2003) *Changing women's roles, changing environmental knowledges: evidence from Upper Egypt*. *Geographical Journal*, 169 (4). pp. 313-325. ISSN 0016-7398



Publisher's URL (DOI): <http://dx.doi.org/10.1111/j.0016-7398.2003.00095.x>

Project Acronym: OCRIS
Contact: Kathleen Menzies or Gordon Dunsire
Date: October 2009

Item Type:	Article
Status:	Published
Refereed:	Yes
Authors:	Briggs, J. and Sharp, J. and Hamed, N. and Yacoub, H.
Subjects:	H Social Sciences > HQ The family, Marriage, Woman H Social Sciences > HC Economic History and Conditions
Faculty/Department:	Faculty of Physical Sciences > Department of Geographical and Earth Sciences
Publisher:	Blackwell Publishing
ISSN:	0016-7398
Copyright Holders:	Copyright © 2003 Blackwell Publishing
First Published:	First published in <i>Geographical Journal</i> 169 (4) 313 - 325
Publisher Policy:	Reproduced with the permission of the Publisher.
Unique ID:	glaseprints:2003-170
ID Code:	170
Deposited On:	07 Apr 2004
Last Modified:	05 Dec 2008 12:32

Repository Staff Only: [item control page](#)

Figure 19 - Partial record for the post-print version held in Enlighten.

The Library OPAC also has a record for the *Geographical Journal* in which the article was published.⁷ The post-print is available electronically by various means (Expanded Academic ASAP, the JSTOR Arts and Sciences II Collection, the SocINDEX with Full Text and the Wiley-Blackwell Full Collection).

There is clear overlap between both IRs and the OPAC in relation to information about and copies of journal articles. Partly this foregrounds the necessity for clear notes on *version* with all IR deposits (this is already occurring). It also flags up the fact that ensuring links exist between pre- and post-print versions and records can help end users find what they are looking for more readily, navigating through a "seamless web" of information rather than encountering "digital dead ends".

Repositories may state that the 'full text is not available via the repository' or similar, when only providing bibliographic data or an abstract. However, even when the full text is available via library subscriptions (physical or electronic) links do not tend to be instantiated. At times, users are even directed to expensive subscription sites which are unlikely to be useful to them.

Performing an advanced search for *Design* magazine in the Journal title field of the **University College of the Creative Arts** library catalogue retrieves collections located in each of the 5 libraries throughout UCCA: Canterbury, Epsom, Farnham, Maidstone and Rochester. Holdings information for the magazine is available.⁸

Records retrieved are for volumes of the magazine published from 1949, as illustrated below:

Design [Change Catalog Display](#)

ISSN: 0011-9245

Title: [Design.](#)

Publication info: London : Design Council.

Current frequency: Quarterly, 1995-

⁷

<http://eleanor.lib.gla.ac.uk/search~S6/?t%22the+geographical+journal%22/tgeographical+journal/1%2C5%2C9%2CB/frameset&FF=tgeographical+journal+online&1%2C1%2C/indexsort=->

⁸

<http://librarycatalogue.ucreative.ac.uk/uhtbin/cgiirsi.exe/APY2qAWNzW/CANTERBURY/245710023/2/1000>.

Volume/date range: Rochester: 1961-1999 (Quiet Study Room)

Volume/date range: Canterbury: 1960-1998

Volume/date range: Maidstone: 1949-1999 (nos 164, 166-167, 176, 184, 187, 191, 193, 201, 208-209, 211-212, 218-219, 225, 232-233, 241, 243-249, 460-461, 464, 467, 485, 525, 534-536, 542, 544 missing.)

Volume/date range: Epsom: No.2 Feb 1949 to No.545, May 1994 (incomplete)

Volume/date range: Farnham: 1949 - 1951 then 1962 - 1998

General Note: Indexed in ABM, Art Abs, DAAI.

General Note: Publication ceased with no. 545, May 1994, resumed 1995, ceased after Spring 1999.

Figure 20 - UCCA bibliographic records for the magazine *Design*, published from 1949 - 1999.

Online runs of the magazine are not however recorded in the OPAC. Design Online is an electronic library containing digitised versions of *Design* for the years 1965 to 1974. Through Design Online, VADS offers access to digitized copies of *Design* magazine between these dates, provided by the London College of Communication. It is part of the Digitisation in Art and Design (DIAD) project <http://www.vads.ac.uk/resources/DIAD.html> which, according to its webpages aims "to provide a digitised record of seminal journals in design and applied arts subjects within the Higher Education sector".

Design Online only retrieves records from the years 1965 to 1974 however it supports access to "around 100 pages in each magazine, which are available as full screen size black-and-white or colour images". The magazines are therefore a searchable resource through the VADS web interface.



Figure 21 - The Search by Year interface of Design Online.

Although there is partial duplication in scope between IR and OPAC here, the OPAC provides more extensive coverage to this journal whilst the IR provides enhanced functionality to end users.

Are such examples *actually* duplication when the item is only *linked to* via a library's subscription packages? OPAC records may reveal that access to electronic journal subscriptions is restricted to authorised users only. Yet full-text versions of many of the articles in a journal may be held in an

Open Access IR. Similarly, some copies may offer enhanced functionality, with annotative tools or the ability to save to a workspace. Different descriptive metadata may be associated with each one.

It is worth considering some conceptual questions here: if items are clearly purposed in different ways or targeted at distinct types of user; if one copy is access restricted and another freely available; if one record is for a print copy and the other for a digital one, *are these examples of true duplication?* They are the same work, but - to use the terms of the Functional Requirements for Bibliographic Records (IFLA, 1998) - are different *manifestations* of it. A similar situation arises with the distinction between "pre-print" and "post-print" - as titles may change between submission and final publication, they may be considered to be different manifestations of the same work. In some cases these might even be regarded as different works.

6.6. Theses

Links between different copies of high-level theses tend to be in place across the group of HEIs investigated. The LMS will link to the electronic copy in the IR and most end users within the IR (one assumes) would be aware that physical copies of these will be held with the library, although in some cases explicit links go from IR to LMS.

It is especially useful if an IR refers to the print copy of a thesis in cases where "embargoes" have been placed on certain content for copyright reasons. An example can be found within Enlighten, **Glasgow University's** central repository:

Item Type:	Thesis (PhD)
Qualification Level:	Doctoral
Keywords:	gender, body, cinema, spectatorship, embodiment, physicality, sport, feminism, textual analysis
Subjects:	<u>G Geography. Anthropology. Recreation > GV Recreation Leisure</u> <u>P Language and Literature > PN Literature (General) > PN1993 Motion Pictures</u>
Faculty and Department:	<u>Arts > Department of Theatre, Film and Television Studies</u>
Supervisor:	Geraghty, Prof. Christine and Boyle, Dr. Karen
Date of Award:	2009
Embargo Date:	03 March 2012
Additional Information:	Due to copyright restrictions the full text of this thesis cannot be made available online. Access to the printed version is available once any embargo periods have expired.
Printed Thesis Information:	http://eleanor.lib.gla.ac.uk/record=b2658734
Unique ID:	glathesis:2009-611
Copyright:	Copyright of this thesis is held by the author.
Deposited On:	03 March 2009

Figure 22 - Doctoral Thesis defended at the TFTS department at Glasgow, with embargo period imposed.

6.7. Subject headings

Within LMSs alone one may encounter many different subject heading indexes due to the amount of services made available by HEI libraries. Subject fields may be searched as part of a "keyword search" facility (where the user effectively guesses which term may return a result) while SFX or Electronic Journal subscription packages may provide their own, automatically generated lists to help guide users towards resources. What is meant by "Subject" may therefore vary depending on which library service you are accessing; subject terms may relate only to subscription journal titles or they may relate to the Library of Congress Subject Headings contained in the OPAC's bibliographic records.

While this can be confusing, the problem within IRs tends to be the opposite – rather than *many* types of subject heading or keyword being used, there are generally too few or none at all. Where they are in place these tend to be derived from user-generated metadata attached to deposited items thus they are error prone and rife with inconsistency, no recognised schema or controlled vocabulary being used. Navigation menus tend to use Library of Congress Classification (LCC) to broadly enumerate topics rather than Library of Congress Subject Headings (LCSH) which provide information on actual subjects. Yet often LCC indices are referred to as "browse by subject" menus.

Thus terms used within IRs will not necessarily find equivalents in the OPAC.

Those administering departmental, discipline or subject based IRs often modify or adapt subject schemes to suit the needs of those using them. For example:

Southampton's inter-disciplinary CogPrints repository acts as an "electronic archive for self-archive papers in any area of Psychology, Neuroscience, and Linguistics, and many areas of Computer Science...Biology...Medicine...[and] Anthropology... as well as any other portions of the physical, social and mathematical sciences that are pertinent to the study of cognition." (University of Southampton, 2009). Once inside the user finds that it has adopted a mixed and granular approach, with in-house terms seeming to be based on a controlled vocabulary. Narrower terms or sub-divisions are used within 8 top-level subject areas, in a 2 tier hierarchy. The specificity of items in the collection is clear. Yet as can be seen in the figure below, journal titles are used within "subject" menus alongside actual subject headings (<http://cogprints.org/view/subjects/>):

Browse by Subject

Please select a value to browse from the list below.

- [Subject Areas](#) (3358)
 - [Biology](#) (522)
 - [Animal Behavior](#) (132)
 - [Animal Cognition](#) (116)
 - [Behavioral Biology](#) (88)
 - [Cognitive Archeology](#) (1)
 - [Ecology](#) (15)
 - [Ethology](#) (75)
 - [Evolution](#) (173)
 - [Population Biology](#) (12)
 - [Primateology](#) (29)
 - [Sociobiology](#) (60)
 - [Theoretical Biology](#) (143)
 - [Computer Science](#) (1016)
 - [Artificial Intelligence](#) (504)
 - [Complexity Theory](#) (123)
 - [Dynamical Systems](#) (136)
 - [Human Computer Interaction](#) (49)
 - [Language](#) (148)
 - [Machine Learning](#) (200)
 - [Machine Vision](#) (62)
 - [Neural Nets](#) (242)
 - [Robotics](#) (206)
 - [Speech](#) (26)
 - [Statistical Models](#) (118)
 - [Electronic Publishing](#) (67)
 - [Archives](#) (61)
 - [Copyright](#) (14)
 - [Economics](#) (33)
 - [Peer Review](#) (19)
 - [JOURNALS](#) (406)
 - [Behavioral & Brain Sciences](#) (19)
 - [Calicut Medical Journal](#) (9)
 - [Indian Pacing and Electrophysiology Journal](#) (165)
 - [Internet Health](#) (2)

Figure 23 - Extract from the CogPrints repository's "Browse by Subject" menu.

6.8. Name Authority Control and finding publications by specific authors, research groups and departments

As illustrated on page 27 of this report, most IR questionnaire respondents indicated that they utilise either in-house name authority control lists, or none whatsoever.

Given that Southampton has 10 IRs (most but not all arranged on a discipline/subject basis), it presents a clear opportunity to compare and contrast a variegated suite of Institutional Repositories. How might the way in which standards are implemented and descriptions used impact upon the implications of duplication?

Although the records contained in the School of Electronics and Computer Science repository (**ECS e-Prints**) are clearly in scope for **e-Prints Soton** (the institution-wide IR), there are significantly more publications in the ECS's own "Nanoscale Systems Integration Group" folder (1746) than there are in the corresponding section of e-Prints Soton (25).

Taking the year 2007 as a test sample, we found there to be only 3 record duplicates across the two repositories. The citation details displayed to users are almost identical:

In e-Prints Soton:

de Planque, M.R.R., Raussens, V., Antoranz Contera, S., Rijkers, D.T.S., Liskamp, R.M.J., Ruyschaert, J-M., Ryan, J.F., Separovic, F. and Watts, A. (2007) [β-sheet structured β-amyloid\(1-40\) perturbs phosphatidylcholine model membranes](#). *Journal of Molecular Biology*, 368, (4), 982-997. ([doi:10.1016/j.jmb.2007.02.063](#))

Ramanujan, C.S., Sumitomo, K., de Planque, M.R.R., Hibino, H., Torimitsu, K. and Ryan, J.F. (2007) [Self-assembly of vesicle nanoarrays on Si: a potential new route to high-density functional protein arrays](#). *Applied Physics Letters*, 90, (3), 033901-033903. ([doi:10.1063/1.2431774](#))

Gonzalez, D.C., Kiziroglou, M.E., Li, X., Zhukov, A.A., Fangohr, H., de Groot, P.A.J., Barttlet, P.N. and de Groot, C.H. (2007) [Long range ordering in self-assembled Ni arrays on patterned Si](#). *Journal of Magnetism and Magnetic Materials*, 316, (2), e78-e81. ([doi:10.1016/j.jmmm.2007.02.059](#))

Figure 24 - An edited version of results retrieved from e-Prints Soton, showing duplication of bibliographic records between e-Prints and ECS e-Prints (fig.25).

In ECS EPrints:

de Planque, M. R. R., Raussens, V., Antoranz Contera, S., Rijkers, D. T. S., Liskamp, R. M. J., Ruyschaert, J. M., Ryan, J. F., Separovic, F. and Watts, A. (2007) [β-Sheet Structured β-Amyloid\(1-40\) Perturbs Phosphatidylcholine Model Membranes](#). *Journal of Molecular Biology*, 368 (4). pp. 982-997.

Ramanujan, C. S., Sumitomo, K., de Planque, M. R. R., Hibino, H., Torimitsu, K. and Ryan, J. F. (2007) [Self-assembly of vesicle nanoarrays on Si: A potential route to high-density functional protein arrays](#). *Applied Physics Letters*, 90 (3). 033901.

Gonzalez, D. C., Kiziroglou, M. E., Li, X., Zhukov, A. A., Fangohr, H., de Groot, P. A. J., Barttlet, P. N. and de Groot, C. H. (2007) [Long range ordering in self-assembled Ni arrays on patterned Si](#). *Journal of Magnetism and Magnetic Materials*, 316 . e78-e81. (In Press).

Figure 25 – The corresponding records in ECS e-Prints.

Neither system makes the full-text available though both provide links to the official location of the full-text articles at Science Direct. Perhaps data is being shared by both systems or drawn from a common source.

The "Nanoscale Systems Integration Group" is referred to on its webpages as both "Nano" and "Nano Research Group" while it is referred to across both IRs by different naming conventions; hence it is unclear what the definitive name for this research group is within the University.

e-Prints Soton reflects accurately the structure of the University, as shown below.

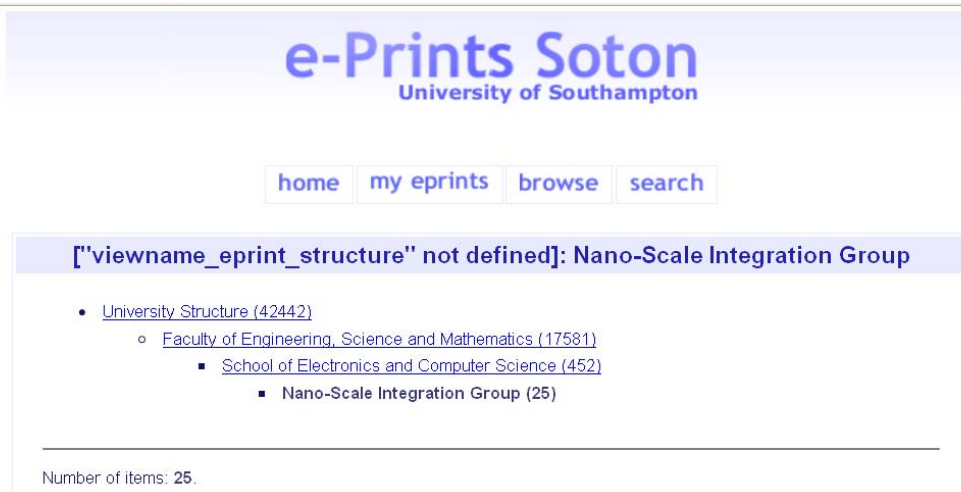


Figure 26 - Extract from the University Structure hierarchy used in e-Prints Soton.

ECS EPrints, with its scope tied to a specific school (Electronics and Computer Science), organises items by Research Group, an example of which is shown below:



Figure 27 - Extract from CogPrints with Research Group structures implemented as a menu.

Although these names match the way the groups are referred to internally, the departments of which they are a part (hence the structure of the University, useful to external data harvesters) is not referred to in the records. Only the general "ECS Groups" label is applied. On the main webpage of the ECS Department, the term "Pervasive Systems Centre" (complete with extra "s") is used.

It is useful to widen the discussion at this point and consider whether duplication and scope overlap in fact serves a useful purpose in some cases, especially in an institution with multiple repositories. Different sites may be used by different types of user/community as some may feel more at home (as depositors or users) with a departmental IR rather than an institutional one, which may be tailored around their specific needs (whether in terms of format supported or subject granularity reflected or tools built in). Some users may prefer electronic journals catalogues where they can look at what other content appears in a given issue, rather than being limited to individual records; some researchers may prefer the interfaces of a "classic catalogue" and others those of an RDP. Copies may be different manifestations or offer different functionalities.

There is also the option of offering users different "views" of the same repository (i.e. giving the appearance that a customised sub-set of a database is a distinct, separate service).

Developers of **Strathclyde's** Strathprints repository informed us that, while they do not consider it an ideal solution, they have:

"Recently started to use a new method for handling different instances of the same work, e.g. a journal article and conference paper, so that only the article appears to Strathprints users, but the conference paper is still stored in the database and could be exported, e.g. for complete CVs or staff web pages. That has only just started. The aim is to reduce duplication for users but not lose any information. The Notes field in the primary 'live' record contains a reference to the variant form."

Of course, this is based on certain informed assumptions about user requirement:

"Most articles are available online, either within Strathprints or externally. Most conference papers are not. And if they are, they are often buried within a single huge PDF of abstracts and papers. Also the term 'finished article' is relevant - our assumption, and it is an assumption, is that where the items differ, the article is more complete, and designed for reading rather than being listened to. However,

in most cases, the deposited document, or the external link, are exactly the same, so it is unnecessary duplication to have two identical documents attached to records where the titles, authors and abstracts are identical.

If both documents are required, and are different, then there is no problem about including two versions of a document to one metadata record, that includes the detail about both article and conference paper. Book chapters are often better than conference papers as they have an ISBN and so should be more easily located in print if not available online."⁹

The key is creating links between central IRs and specialised instances, between IRs and other systems such as LMSs, and between the different copies and versions of items so that users can gain access to as full a picture as possible of the range of items produced and held by their institutions.

These links, however they are achieved at the technical level, must be underpinned by standardised, consistent metadata and controlled vocabularies.

⁹ Personal correspondence via email, July 2009.

7. Summary of main findings of case studies

Cambridge University and the University of Glasgow

7.1. Rationale

The parallel OCRIS case studies carried out within two research intensive institutions - Cambridge University and the University of Glasgow - presented an opportunity to discuss in depth many of the issues addressed by the project and which confront the professionals who work, as a matter of course, with LMSs, OPACs and IRs as they attempt to meet the challenges facing academic libraries of providing an integrated e-infrastructure for research.

A key goal of the Case Studies was to discern how the systems under examination relate to the wider agendas and priorities of the institutions to which they belong, and what were their actual or perceived roles in terms of institutional mission, research strategy, administrative requirements, and the curation/preservation of university publications and other research output. Further consideration was also given as to how such issues impact on library resources, workflows and staff development.

In accordance with the stated terms of the OCRIS Project, the following was discussed with case study participants:

- Technical issues involved in the implementation and running of library systems both discretely and with regard to integration and interoperability.
- Metadata schemes and standards used
- Views of staff (and, by proxy, users) with regard to service and service development.
- Workflows and procedures currently governing development and use of IRs and OPACs and how these might be refined or improved.
- The socio-political and economic factors and imperatives informing the strategic aims and objectives of HEIs; how these affect collection policy, development policy, work practices and the attitudes of staff and other stakeholders.

Both HEIs are research-intensive and members of the Russell Group; their University Libraries are members of Research Libraries UK (RLUK) and both are involved in the HEFCE REF Pilot Exercise. The structural and operational organisation of the two Universities and their respective libraries is, however, significantly different.

7.2. Introduction

It is unsurprising that within both Cambridge University Library (CUL) and the University of Glasgow Library (UGL), staff demonstrated a highly informed expertise and showed ample awareness of the issues under consideration by OCRIS: duplication of content (at both item and record level), the compatibility or otherwise of format and content standards and, crucially, the need for systems that share information as part of a broad, interoperable "ecosystem" across not only library services but institutions as a whole.

That the workflows and practices involved in addressing these issues are complicated and at times hindered by factors outwith the control of library staff – economic, political and cultural factors; research assessment obligations; the technical capabilities of the systems with which they work; the quality and availability of externally sourced bibliographic data – encourages debate and collaboration. Thus, while there are many pressures upon staff there is also much innovation and continuous, iterative assessment and development.

7.3. Systems and Scopes

The Library Management System in use at **Cambridge University Library** is Voyager, with WebVoyage acting as the OPAC. Newton is the major online catalogue for the libraries of the University, including the CUL and its 4 dependants as well as most faculty, departmental and college

libraries. Of Cambridge's 108 libraries, 89 contribute records. All items describable by MARC21 and AACR2 are within scope. Further:

"Due to the size and complexity of the libraries in Cambridge, the catalogue is currently divided into several smaller catalogues, each holding information on a specific set of libraries or library collections. A test Universal Catalogue interface provides a means of searching all of the smaller catalogues." (University of Cambridge, 2005).

The main IR at Cambridge is DSpace@Cambridge, which seeks to gather all types of content from across the University. It was established in 2003 and is based on DSpace version 1.45.1. Although DSpace@Cambridge is an established service, it is funded on a temporary basis. The current funding runs out in 2012 and a strong business case will be promoted in order to justify its continued existence, making it clear that the value is worth more than the cost.

There are several other IRs at Cambridge administered at departmental level. These are the Computer Laboratory Technical Reports Repository - which uses a custom-built database and the Cambridge University Engineering Department Publications Database (CUED), an EPrints installation holding bibliographic data only. The Department of Earth Sciences has a similar EPrints set-up as does the Computer Laboratory.

The Library Management System in use at **University of Glasgow Library** is Millennium, with WebPAC Pro acting as the OPAC. Again, all item types permissible according to MARC21 are in scope. Each of the 6 other UGL affiliate libraries contribute records to the main catalogue.

The UGL administers 3 IRs: the Glasgow Theses Service is a collection of full text Higher Degree theses successfully defended at the University of Glasgow. The service does not contain *all* theses defended at the University of Glasgow but rather, around 600 of them (and rising). DSpace is for 'grey literature' such as working papers and technical reports and consists of various self-defined communities and sub-communities. Enlighten (formerly known as the "Glasgow EPrints Service") is the University of Glasgow's core Institutional Repository service for published research material and is therefore its key repository. Indeed, staff at UGL consider Enlighten the *only* Glasgow IR, the other two being specialised instances with far narrower remits.

At the **UGL**, the Encore Resource Discovery Platform operates on top of the LMS while systems staff at **CUL** are in the midst of assessing recent tenders made for procurement of their own RDP.

7.4. Key points of comparison

7.4.1. The Impact of Research Assessment

Both Cambridge and Glasgow are among the 22 institutions involved in the HEFCE Research Excellence Framework (REF) Pilot project. (HEFCE, 2009).

The existing Research Assessment Exercise (RAE) and the forthcoming REF have brought into sharp relief specific, system level issues impacting upon the gathering of publications data, the generation of reports and the development of efficient tools and workflows through interoperability/integration. Library systems are implicated in markedly different ways within each University.

At **Glasgow**, the library's Bibliometrics Working Group was heavily involved in compiling publications lists for HEFCE as part of the REF pilot. Research assessment preparations have strengthened the position from which the library argues for the necessity of their main repository (Enlighten) and the need for mandates (successfully lobbied for and instituted over the past two years) requiring staff (and students defending high-level degrees) to deposit their work. This has acted as a catalyst for fostering relationships with a range of departments, including Research and Enterprise, leading to joint projects such as Enrich (JISC, 2009a) exploring how the IR might "hook up" with other systems.

Further, the University is *"intending to use Enlighten as the platform for its publications database. Bibliographic details of all peer-reviewed and published research outputs will now be added to this*

database, with links to the full text where available. This will "ensure that research outputs are prepared and curated in a way which helps maximise the value that they have for the university in terms of the external use of bibliometric data e.g. league tables, post-2008 RAE." (University of Glasgow, 2008).

Institutional "buy-in" has been heavy (Ashworth, 2004) and delivering on Enlighten is now one of the Key Performance Indicators (KPIs) for assessing the progress of the University's research strategy. Enlighten acts as the University's Publications Database and is tied firmly into its official Publications Policy.

At **Cambridge**, the Research Policy Committee oversaw the REF pilot activities of the Academic Division. The Management Information Services Division - who deal with central administrative systems as part of the University Offices' Unified Administrative Service – do not oversee library systems (which have not been involved in the REF pilot although staff were debriefed). Thus the repository has not been as central to REF planning as many centralised systems have.

The Research Policy Committee found that for the database they are developing, "an interface with CHRIS [Cambridge Human Resources Information System], was the minimum needed to deliver the pilot data collection *and* the full submission expected in 2010...[]...Evolution of the REF might require interfaces with CamSIS [Cambridge Student Information System] and with CUFS [Cambridge University Finance System], but it was not expected that they would report through Symplectic¹⁰ and nor was additional functionality for Symplectic assumed."

The Committee also found that "For extension of the REF to the humanities and social sciences there was little expectation of any system that could capture research eligible books. There was little prospect of any alternative system." (University of Cambridge, 2009b).

The absence of a pivotal role for library systems in data gathering is notable as one assumes the IR (and LMS) would be valuable in capturing such publications/outputs data.

Both the LMS and IR contain bibliographic data about books and book items and DSpace@Cambridge staff are hoping to position themselves more centrally, partly by building links with the Symplectic Publications System currently being road-tested at Cambridge. The goal is to allow researchers who are managing their publications in Symplectic to also have the option to upload the full text for deposit into DSpace by re-using metadata. The two systems would be linked using web services based on Atom and the SWORD protocol.¹¹

The Centre for Applied Research in Educational Technologies (CARET) at Cambridge have begun to integrate DSpace@Cambridge with Sakai/CamTools (a software suite and tools allowing staff and students to share research and learning materials) to encourage increased use of DSpace¹² while a current project is developing applications allowing publication and simultaneous deposit of a paper into DSpace and other research infrastructure systems including a REF publications/citations database, incorporating editing and text mining tools. (JISC, 2009c).

Interestingly, DSpace@Cambridge will be *provided* with bibliographic data by the REF database rather than providing it, pulling in information to enrich the metadata associated with the research publications deposited with them; this is because the REF system team has already negotiated contracts with organisations like Web of Science, allowing them to reuse metadata for repository purposes. DSpace@Cambridge staff investigated getting relevant publications data directly from WoS - this was very expensive thus not feasible within current repository budgets.

¹⁰ Symplectic's publications management software (designed for academic institutions) is currently being used within Cambridge. More information can be found at: <http://www.symplectic.co.uk/> [Accessed 01 October 2009].

¹¹ See: <http://www.swordapp.org/> [Accessed 06 August, 2009].

¹² For more information, see: <https://camtools.cam.ac.uk/access/wiki/site/jisc-ctrep/home.html> [Accessed 26th August, 2009].

At **Cambridge** it was pointed out that funding bodies or assessment exercises may ask for sets of statistics only meaningful in a specific instance; these could not and would not be built into an interface or database design; thus, such requests should not be made entirely central to systems development; other, local priorities must remain equally meaningful.

7.4.2. Systems Issues

Library Management Systems can present technical problems when interoperability or other improvements are sought.

At **Cambridge**, the desire to provide a single point of access to holdings information, bringing together their fragmented databases and catalogues did not prove possible because of the limitations of the Voyager hardware and software; there was a clear lack of scalability thus the need to search groups of catalogues within Newton remains. This signals the problems inherent in trying to establish institution-wide interoperability.

At a lower level, the de-duplication module offered by Voyager was felt to be inadequate (for instance, sometimes both detected duplicates are "thrown out" and system re-builds after de-duplication are slow, making workflows at times convoluted). Complex weighting mechanisms being developed by Ex Libris to detect whether items described by separate records were the same never worked beyond a low percentage.

Glasgow was approached by its LMS vendor several years ago when that vendor was developing a prototype resource discovery interface. For Glasgow, integration with the Electronic Resources Management (ERM) system was vital; however, due to the fact that within the group of libraries contributing to testing and development, the majority were not concerned with ERM, for the vendor it was not deemed a priority – hence not all of the interoperability/integration features desired and required by UGL were reflected in the eventual product.

Similarly, a project was undertaken attempting to make archival resources discoverable through the main catalogue using a "metadata builder" package. The intention was to harvest records and create corresponding MARC records to describe items encoded in EAD; however trying to cross-walk the two standards (which have no direct correlation) was near impossible, and there were numerous problems with authority control.

At the same time, LMSs have many advantages over IRs. Cataloguing modules and tools are sophisticated in comparison to those offered by most IRs, such as data duplication algorithms, auto-complete functions or bibliographic and authority editing tools. However, IRs are beginning to play catch up (for instance, EPrints can better discriminate between print and electronic editions of journals).

7.4.3. Authority Control and Standards

In **both libraries**, there were common concerns raised about inadequate authority control (of both names and subjects) and problems with bibliographic records. These problems, which are barriers to developing efficient cataloguing workflows, are perceived most readily by those working with the LMS and OPAC (the "traditional" cataloguers). Despite the fact that these problems may be more prevalent and pronounced in IRs, their staff - as a body - may not yet have the experience to identify such issues.

One may consider importing pre-existing bibliographic records into the LMS, for example, to improve efficiency; this is complicated by the fact that the authorities used by major database services such as the Library of Congress or WorldCat do not necessarily match those being used by any given library thus do not accommodate local requirements.

Further, there are problems with using internally produced authority lists for staff names and departments: there are multiple links within and between faculties and even within single departments it is always possible that there may be more than one member of staff with the same name, causing ambiguity. Reflecting these hierarchies and disambiguating authors can be resource intensive,

especially across systems. When searching amongst various catalogues one commonly encounters issues both of inconsistency and ambiguity because of different standards being used for names or because of the different ways in which those standards are applied.

With the best will in the world, accurate (and by implication, shareable or interoperable) records cannot be maintained if: 1. Standards are not well understood and applied, 2. De-duplication processes and the cleaning or updating of bibliographic records becomes too time-consuming, 3. Externally sourced records prove unreliable, 4. Resources are too thinly spread.

Within **both repository teams**, it was felt that the authority lists supplied by default with repository software (such as top level Library of Congress Classifications (LCC) or the Norwegian science index) or which are readily available publicly are not always appropriate to or specific enough for the communities using them. Such problems are most tangible when we consider the needs of higher level academics. Many faculties require controlled vocabularies, with sufficient granularity built-in, for the effective description and retrieval of items relevant to their specific subject area.

At **Cambridge** such complications are writ large because of the sheer number of libraries within the federation, the lack of common standards and also the need to update and modify many older catalogue records conforming to out-moded standards¹³ which is resource-intensive. Some older records in the Voyager database are brief and do not contain data in the fields used for "match point" detection of duplicates; thus these on occasion remain visible to end users.

At **Glasgow** the structure is far less complex; however, there are many collections held within departments which are not catalogued fully (or at all) and for which there are not sufficient resources allowing them to be added to the main catalogue. Archival holdings are kept entirely separate and only *some* Special Collections records are visible within the main catalogue.

In **both libraries** there were frustrations at inadequate resourcing putting pressure on bibliographic services staff who are increasingly involved in creating records for both the OPAC and the IR.

Again, in **both libraries** it was felt that while major author identification projects such as the Virtual International Authority File (VIAF) (National Library of Australia et al. (2009)), IFLA's Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) or new schemes such as Resource Description and Access (RDA) (Joint Steering Committee for Development of RDA, 2009) are of interest and could in future prove invaluable, they cannot be built into working practices which do not allow for "waiting and seeing".

7.4.4. External sourcing of bibliographic records

Taking advantage of major databases which make their records available to others either freely or at a cost (such as PubMed, Nielsen, Scopus and OCLC) can drastically improve workflows within the LMS, making it possible for more items to be made visible in the catalogue more quickly, with a clear adherence to recognised standards. At **both institutions** there is a desire to utilise these "clean" bibliographic records.

Yet it would be extremely helpful in making such practices scalable if bibliographic data could be sourced from as few places as possible, and as easily as possible, with the clear assurance of "best quality". This would reduce time spent editing or improving records and would, in turn, encourage institutions to share their own records, making them visible beyond institutional boundaries, potentially as a part of a web of "Open Data". Licensing agreements could be established nationally for the use and re-use of metadata, in keeping with the recommendations of the recent RIN report *Creating catalogues: bibliographic records in a networked world*. (Research Information Network, 2009).

¹³ For example, as part of the Tower Project (Cambridge, 2009b).

7.4.5. Links between LMSs/OPACs and IRs

At **Glasgow**, the UGL team are looking at harvesting records from the EPrints repository for use within their OPAC, using the harvester supplied with Encore. At **Cambridge**, it is a requirement of the tender that whatever RDP they are supplied with be able to harvest records from DSpace for use in the OPAC.

A "find more" feature, which uses the library's WebBridge link resolver is already instantiated within **Glasgow's** EPrints test environment, pointing users accessing a journal article to print and electronic library holdings of the same journal; this will be developed and will go live within the IR in the near future.

SFX link resolvers are used within **DSpace@Cambridge**, with an SFX link to "eresources@cambridge" being pulled up at the bottom of the page for searching electronic journals and databases. There are problems with the utility of this feature; often it is rendered irrelevant because the resolver doesn't "know" whether the Library's subscription package holds any related items or information. In the majority of cases there are no other versions of the DSpace item available and if they did exist, the SFX resolver and its "Knowledge Base" would not be aware of them. One exception would be for these containing 3rd party copyright material. If permissions cannot be granted to release this content electronically, the SFX button would be a useful tool to locate the complete print copy.

Glasgow are beginning to see online course materials (normally managed within the LMS) being deposited in Enlighten as well as in the OPAC. Here, the workflow and the correspondence between the IR record and the existing OPAC record is not straightforward, nor does it necessarily imply interoperability since in many cases the original entry is a static file. However, information is being shared and made visible to a wider range of users as a result and the sharing of records should only be increased, even if this to some extent might be viewed as "duplication".

7.4.6. Links with other institutional systems

At **both libraries** there are strong existing links between the LMS and HR systems; financial information is self-contained within both LMSs.

The LMS at **Glasgow** "pulls" data in from HR as the basis for its Staff Patron Records and derives data from the Student Records System (SRS) for student Patron Records. This is undertaken by batch loading records on an overnight basis thus it does not constitute "real-time" interoperability. Again, web services might be utilised to enable this information sharing and processing to be done more quickly were a need for increased speed deemed beneficial.

Similarly, at **Cambridge**, the barcodes on student and staff cards are fed into the LMS from CHRIS (the Cambridge Human Resources Information System) and used for patron authentication and account management. Local copies of patron records are stored by the system across various institutional databases meaning users registered with more than one library are thereby presented with just a single patron record aggregated from details held across the systems.

As previously mentioned, links are emerging between both IRs and research administration systems. Staff at **Glasgow** are currently involved in a JISC-funded project which will establish links and interoperability between the university's Research Systems and Enlighten to retrieve data on the range of publications and funders associated with research pools and projects.

At **Cambridge**, some integration has been established between DSpace and the Sakai/CamTools suite; applications are also being developed to allow publication and simultaneous deposit of a paper into DSpace@Cambridge and other research systems (such as Symplectic) through metadata re-use, using web services based on standard protocols.

At **both libraries** centralisation was not necessarily seen as the most effective way to develop interoperability. At Glasgow it was stated that: *"Departments like their autonomy; technology is available to aid their integration but it relies in part on web services being introduced which push and pull data across distributed services rather than on being held centrally."*

At **Cambridge**, workflow inefficiencies would inevitably be encountered if changing from a system of local control to one which relies on the flow of data to-and-from a centralised system. If information were to be directly manipulated at a higher level, staff within individual libraries could find themselves uncertain about the currency, accuracy and fitness-for-purpose of records being pushed or pulled into their systems.

7.4.7. Re-imagining the role of IRs and the LMS

"[There is a need] to make the material much more visible and available through standard library facilities. They should be available through one search facility – the LMS." - Member of staff at UGL.

Staff at both **Cambridge** and **Glasgow** feel that the increasing number of electronic services – particularly IRs – have led to a renewed and positive engagement between end users and their libraries. Digital materials and the systems which make them available help position libraries as central to University life. There is a clear convergence in terms of philosophy and shared skills. Both IRs and LMSs are viewed as potential "shop-fronts", (and also, "data warehouses") making visible content available from, and publications produced by, the library and the University.

Old and new systems are not viewed as competitors; tensions tend only to arise as frustrations about inadequate resourcing to support the workflows of both OPAC and IR.

If the LMS/OPAC is to become a "junction box" - the "share point" or "portal" - for catalogue records, patron records, email and other re-usable data, as staff at **Glasgow** hope it might, a cultural as well as a technological shift will be required. Such an evolution in service delivery would require the integration of new workflows into standard cataloguing procedures, with the necessity to further improve the skills base of cataloguers. This would sit well alongside current plans to expand the functionality and role of Enlighten, with increased integration and interoperability between the two systems, including the harvesting of records from Enlighten into Encore.

At **Cambridge**, similar thoughts were expressed regarding the expansion of both the LMS and the IR, and the types of data they provide. Staff are interested in the idea of providing different, customised interfaces to the LMS for different types of user - for example, library staff, whose non-public Resource File at present uses an interface nearly identical to the public one.

The 'Library Toolbox' at **Cambridge** offers OpenSearch "plug-ins", an iGoogle search gadget and Reference Management tools (the Zotero citation collection and management suite which works with the Firefox web browser and can export Newton results lists to RefWorks and EndNote). RSS feeds are set up to alert users to new print, electronic and science resources being subscribed to and any electronic resources trials being run. The web browser toolbar allows multiple field searching of any Newton catalogue without having to go to the web interface. This is based on the Virginia Tech University Libraries' LibX Open Source framework.

However, in **both cases**, there is no evidence that the development activities occurring within the IR are intended to impact upon or alleviate any difficulties with the LMS or OPAC and their methods of gathering, storing, preserving and making accessible information. Staff at **Cambridge** feel that "IRs may end up with the same level of fragmentation" as other types of system. Staff there did not necessarily feel that duplication between the OPAC and the IR was problematic, being more concerned about whether or not users can find and access content, wherever it may be located.

DSPACE@Cambridge are focussed on positioning themselves more centrally by building a critical mass of publications but also by including cultural heritage materials, becoming something of a Digital Library.

7.4.8. Resource Discovery Platforms

At **both institutions**, the installation of RDPs is motivated not only by a desire to address the expectations of users but also to include and expose Repository content and other types of learning, research and teaching material within the OPAC. This is especially so at **Cambridge University**

Library where it is felt they are "solving a number of problems - not least the multiple catalogue silo situation". These new products, rather than the traditional core LMS or its technologies (such as Z39.50), are being used to push forward interoperability and distributed search.

One of the key benefits of RDPs is that they offer harvesting services. Glasgow is considering how records from Enlighten could be integrated into Encore. Cambridge issued an ITT for an RDP in April of this year, with the product required to:

"...act as a single point of discovery for University collections...It is anticipated that it would also include library materials from digital collections within Cambridge and subscribed electronic material, via a Federated Search service, and by harvesting data from non-bibliographic sources, such as the DSpace Institutional Repository."

The platform would include, among other things, "a means to control de-duplication of data from incoming streams" and "full harvesting and storage of bibliographic and other forms of metadata from a variety of sources, including the [...] Library Management System". Staff could choose to implement only limited de-duplication and could implement an "FRBR-ised" display; with no risk to the underlying data itself (content and carrier are separated) it would be the presentation and arrangement of data that would change. The installation of an RDP is therefore seen as low-risk.

Systems staff have provided detailed specifics about how de-duplication should be carried out to ensure that groups of duplicates will be adequately identified and edited/merged to create "best records" which will be visible to users and "non-best" records, which will be hidden from view (but retained and linked to the "best record").

However, problems may arise with the way data is extracted from records and the way in which attendant search results are presented. Although the full MARC record can often be viewed within a traditional OPAC, only the basic title, author and location (the most reliably encoded fields) are seen in initial results lists. With RDPs, certain fields are more prominently exposed through faceted menus, tag clouds etc. Weaknesses in records may therefore be revealed - for example "not coded" may appear as an option in a menu offering "browse by item type." Perhaps the unearthing of these incomplete or inaccurate records is a good thing in the long-term but only *if resources can be found* to improve and update them; staff already know that they exist.

It cannot be supposed that the user needs which RDPs cater for today, will bear any relation to the user needs of tomorrow.

7.4.9. Development of Identity Management Systems within Institutions

Both Universities are involved in Identity Management activities.

Cambridge's "Identity Management Group" (of which members of CUL's Electronic Services and Systems team are a part) is investigating how HR records relating to staff and students might be better shared across systems, with possibilities including a portable ID that could track an individual throughout their time at Cambridge. This ID would be persistent, even were you to graduate from Cambridge and return at some point as a conference delegate or visiting professor.

There are already both University Cards and Library Cards which use the same number for identification purposes and within the Library, within patron records.

There are multiple "data owners" within Cambridge and devising a comprehensive ID or name list from the Cambridge HR system would be complicated as there is no common login scheme in place across Cambridge, and initials and full names are variously used. College libraries may use their own ID systems on *College* library cards. Many individuals who do not qualify for an institutional email address are still allowed access to Cambridge libraries.

In terms of publications then, there may not be any direct link between a named author and the institution; they may have contributed to a paper or symposium but no name authority would exist for them anywhere at Cambridge.

At **Glasgow**, there are similar discussions regarding the University's Identity Management infrastructure.

UGL staff perceive increased convergence of strategies and greater co-ordination of systems and services, with the IT and Management Information Systems infrastructure steadily becoming "much less nebulous," partly through the application of consistent, institution-wide "digital identities" known as Glasgow Unique Identifiers (GUIDs) administered by the "Identity Vault".¹⁴

The GUID is used by university staff for a variety of services including access to HR and Research Systems and the submission of Time Allocation Schedules (TASs) as part of mandatory HEFCE reporting activities. Within the library, the GUID is tied into a number of services and has replaced Athens as the method by which to access e-books, e-journals and subscription databases.

Questions remain as to whether these IDs have sufficient metadata attached to provide an adequate basis for service development and the elision of ambiguity. In terms of the wider context (i.e. beyond institutional boundaries), these developments do not necessarily make the sharing of records with external systems possible; however they do provide a potentially solid basis for the interoperability of systems *within* Glasgow or *within* Cambridge.

Identifiers created by Identity Management Systems could become the "glue" of what one member of staff at UGL termed a "Junction Box" model, wherein the IR (or another centralised University system) would become a "share point" or "portal" for re-usable data linking staff and student records to publications data and a wide range of other administrative information.

8. Outcomes

OCRIS aimed to:

Survey the extent to which repository content is in scope for institutional library OPACs, and the extent to which it is already recorded there;

The OCRIS survey of scope overlap and item duplication was detailed and intensive. Despite only a moderate response level to the questionnaire, the mixed methodological approach employed proved robust. Combining the questionnaire with case studies, desk research and analysis allowed substantial, meaningful quantitative and qualitative data to be gathered.

Examine the interoperability of OPAC and repository software for the exchange of metadata and other information;

OCRIS' examinations focussed on semantic interoperability and the use of common metadata standards among library systems. Examination on technical aspects of interoperability was limited to some discussion of protocols for data exchange such as OAI-PMH, the use of Open Standards and, within 2 case study institutions, the configuration of Resource Discovery Platforms and link resolvers. Examining technical interoperability was not strength of the project but data sought was successfully gathered.

List the various services to institutional managers, researchers, teachers and learners offered respectively by OPACs and repositories;

The use of an existing JISC output (an ontology of IR services, with some modification) within the online questionnaire proved an invaluable way of formulating questions on services and gathering meaningful answers. The service list provided in this report also considers pending services which adds value to the findings.

Identify the potential for improvements in the links (e.g. using link resolver technology) from repositories and/or OPACs to other institutional services, such as finance or research administration;

¹⁴For more information see: <http://www.gla.ac.uk/services/it/projects/identitymanagement/> and <http://www.gla.ac.uk/services/library/howtofindinformation/accessingresources/glasgowuniqueidentifierguid/> [Accessed 19th August, 2009].

Case studies allowed us to supplement desk research significantly by discussing with professionals, specific instances of links being established between IRs and OPACs and other institutional departments, and what the barriers to making those links can be in real-world situations. These findings were greatly informative to OCRIS' understanding of this area and allowed stronger conclusions to be drawn.

Make recommendations for the development of possible further links between library OPACs and institutional repositories, identifying the benefits to relevant stakeholder groups.

OCRIS in-depth investigations into the IRs and OPACs of 9 Higher Education Institutions and 1 related institution serving 3 HEIs, used qualitative methods and utilised a workflow which allowed for exploration of the subject matter. This enabled many significant observations to be made which would not have been unearthed by either a literature review or from quantitative datasets. Again, the mixed methodology allowed solid recommendations to be formulated.

Benefits of the project to key stakeholders are as follows:

Scholarly and research communities – gain insight into the benefits of allowing OPACs and IRs to interoperate as they relate to scholarly concerns and needs, for example, improving resource discovery and increasing the visibility and linkage of both research outputs and source materials.

University governance and management – gain clarification and insight into how interoperability between administrative systems and library systems will further the aims and objectives of the institution. This points towards efficiency gains, economies of scale, and better value for money.

Librarians, archivists and related information professionals – gain a set of useful findings and recommendations based on a solid body of work which may assist and inform them as they formulate policies, develop systems, and establish cataloguing and other workflows involving their IRs and LMSs.

Repository staff and the repository community – gain a set of useful findings and recommendations based on a solid body of work which may assist and inform them as they establish workflows and policies for cataloguing, resource description and arrangement, particularly in relation to the consistent application of common standards and controlled vocabularies.

HEI administrative departments – will benefit from services which could be developed if the recommendations of OCRIS are enacted, allowing more mutually productive relationships to form between the staff and systems of administrative departments and library departments.

Students and learners – if the recommendations produced by OCRIS are acted upon by UK HEIs, it is hoped that the student population (and other learners with access to the systems and services of the institution) will directly benefit from saved money, improved workflows, more effective and useful systems and services, and improved resource discovery systems.

Groups and organisations affiliated or associated with HEIs (e.g. consortia and umbrella organisations possessing, or with an interest in, IRs and OPACs) – will find HEI library data easier to harvest and share via the developments made to relevant systems as a result of institutions following OCRIS' recommendations.

9. Conclusions

Because of the wide-ranging nature of the OCRIS project, its conclusions are divided into sub-sections.

Interoperability and services

1. Interoperability between IRs and LMSs in UK HEIs is currently rare - only 2 percent of questionnaire respondents state that their systems *definitely* interoperate, with a further 14 percent stating that interoperability is *pending*.
2. Interoperability of either of these system types with *some* type of other institutional system is moderately high, and is slightly higher for LMSs than IRs.

3. Interoperability between LMSs or IRs and a *range* of other institutional systems is limited. It cannot be said that interoperability is substantial or that a wide *variety* of administrative systems interoperate with any individual library system.
4. The REF has clearly been a factor in the establishment or consideration of interoperability between Institutional Repositories and other administrative systems.
5. Services stemming from library systems are limited and narrow, excepting the generation of usage statistics and metadata enhancement services.
6. The generation of reports for specific administrative departments is not a common service offered by either IRs or LMSs.
7. The most popular service offered by IRs remains "advice on Open Access" suggesting perhaps that they are still in their infancy, still require explanation thus have yet to spread their wings in terms of widening their range of services.
8. The use of metasearch/linking tools as well as web services and APIs is moderately popular within LMSs and IRs.

Duplication and scope

1. There is significant scope overlap (81 percent) for *all* item types held in IRs and OPACs.
2. The scoping distinctions and boundaries for IRs and OPACs are becoming increasingly blurred, with many IRs containing bibliographic data and OPACs containing links to full text.
3. Duplication at both record and item level is frequent, especially for print/electronic copies of theses or journal articles.
4. For many OPACs and IRs, *any* type of content is in scope, regardless of the items *currently* recorded or held.
5. Links between print (OPAC) and electronic (IR) copies of theses are frequently instantiated within both systems.
6. Linking for content other than theses is not common. However, within some IRs link resolvers are beginning to be used to direct users to related library holdings.
7. Some HEIs are choosing to expose both OPAC and IR data through the use of Resource Discovery Platforms (RDPs) which offer "vertical search" functionality. The popularity of these systems seems likely to increase.

Authority control and description

1. Authority control within LMSs is high; format and content standards are well supported, with MACHine-Readable Cataloguing (MARC) standards and Library of Congress Subject Headings (LCSH) the most commonly used alongside local authority lists.
2. In IRs there is little authority control for subjects, and only a moderate amount of effective classification. In-house lists are predominantly used for the construction and maintenance of name authorities.
3. Within Institutional Repositories standards are not applied adequately and often not at all.
4. There are frequent inconsistencies and a lack of completeness in statements made on both IR and OPAC webpages about item types and scoping.
5. Modifications allowed to the item and format fields of DSpace or EPrints software, and the presentation of administrative or outdated terms by LMSs within their search/browse lists, undermines consistency, standardisation and clarity for end users, across the sector.
6. The use of vocabularies and standards within library systems in any given HEI is fragmented and disjointed; there is little commonality in resource description.
7. LCC is frequently used within EPrints repositories as the top levels come bundled with the software; however there appears to be some confusion in the IR community about the distinction between Library of Congress Classification (LCC) and Library of Congress

Subject Headings (LCSH). This is clearly significant, suggesting limited knowledge and expertise of professional standards within the IR community.

8. There are many benefits in recognising the Author/Creator/Person field as the metadata element common to all internal HEI systems.
9. There is an increasing awareness of the role that could be played within Institutional Repositories by institutional IDs.
10. There will be an increased role for institutional (or even international, if one is to be more ambitious) personnel and group identity management schemes to enable authentication, access to services and the gathering and compilation of data for internal and external purposes.
 - 10.1. Supporting this can become time-consuming as HR-produced codes or the relationship between a department and the University hierarchy may be subject to constant change. Keeping records up-to-date can therefore be an ongoing challenge.
11. Resource Discovery Platforms (RDPs), because of their nature (for example, the use of "tag clouds") reveal inadequate metadata in catalogue records more readily than do other interfaces.
12. RDPs are only a partial solution to making all relevant items visible to users in one place; they may present search results in ways not best suited to the scholarly needs of higher-level end users and do not necessarily reveal the richness of a library's collections.

Flexibility and working practices

1. The fragmentation or disconnectedness of HE information systems puts a strain on the abilities of cataloguers/bibliographic services staff to work effectively across both IRs and LMSs or for shared workflows to be developed across departments.
2. Lack of resourcing puts a strain on the abilities of cataloguers/bibliographic services staff to work effectively across both IRs and LMSs or for shared workflows to be developed across library departments.
3. Administrative staff and the systems with which they work are not sufficiently involved in cross-departmental collaborations with library systems staff.
4. Batch processing is still considered an effective way to share data between non-library and library systems (particularly HR systems and LMSs).

Dominance of IRs

1. Administrative departments may find the information held in Institutional Repositories to be more directly relevant or beneficial to the work carried out in conjunction with their own systems, than they do the information contained in LMSs. This is likely to be because:
 - A. Both admin systems and IRs are concerned with local, *institutionally produced* data while LMSs record information about a far broader range of resources.
 - B. IRs are often research-centric and privilege recency. The influence of research assessment exercises means that the utility of interoperating with IRs is more immediately apparent to institutional management than is interoperating with the LMS.
 - C. Because IRs are often Open Source, in-house technical developers and experts can be readily leveraged to modify and make available IR data to other systems, provided the skill-sets are there. With LMSs this is not the case as proprietary standards and licenses mean most development needs to be undertaken by vendors.

10. Implications

10.1. Duplication and scholarly needs

How do we define an "Institutional Repository"? Not only "centralised" repositories - many of them primarily designed for the gathering of current research outputs - should be considered IRs. Many HEIs possess a range of discipline, department, content or format specific repositories, including those for use by undergraduate students and learners. All are rich sources of metadata and as a group serve a variety of purposes. These non-central IRs must be better understood and included in any definition if cross-sector enhancements and projects are to be undertaken.

Just as an LMS may be seen to be both an inventory control system and a user-facing front end (i.e. the OPAC) - allowing carrier and content to be usefully separated - so too can the IR. Conceptualisations of Institutional Repositories must recognise the benefits that a diversity of local systems can bring, including the various ways in which contents and metadata might be stored, used, presented and shared.

Unnecessary duplication of data is clearly to be avoided if departments seek to synchronise workflows and use resources efficiently and economically. However, OCRIS contends that duplication and scope overlap is not *by definition* wasteful; often, under certain conditions, it serves a specific purpose, especially in institutions administering multiple repositories or with wide-ranging user communities encompassing diverse sets of preferences and working practices.

Some users may feel more at home (as depositors or users) working with a departmental IR rather than a centralised one, finding it to be tailored better to their specific needs (whether in terms of format support, subject granularity or the added value tools and services built in to it).

Similarly, some users may prefer to find items via an electronic journals catalogue where they can look at what other content appears in a given journal issue, instead of being limited to considering only individual records. One researcher may prefer the interface of a "classic catalogue" and another that of a Resource Discovery Platform (RDP). Copies of items may differ (for example, one may offer enhanced functionality to the end user).

The concepts used by the Functional Requirements for Bibliographic Records (FRBR) should also be considered here; a digital item is a different *manifestation* of a printed work or even a different *expression* of a previous edition of the same work (for example, when we consider pre- and post-prints). In some instances it may even be a different *work* altogether.

Access to as full a picture as possible of the numerous contexts for the reception of research outputs is an important factor for the scholarly community, and one which is as beneficial as access to bibliometrics or other quantitative information on publications.

The findings of other JISC projects must be re-stated. Swan and Awre (2006) have already outlined the case for subject-specific repositories or for repositories run by learned societies (such as the Institute of Physics) and make reference to user demands for subject-specific entry points and discipline-specific services (pp.36- 44) as well as to the different ways in which research communities may choose to make their work available.

We *cannot* consider only the needs of one community, or privilege a commitment to a particular philosophy (i.e. Open Access or the requirements of research assessors) over the working practices and publication preferences of other types of user/researcher.

What is necessary - across the board - is the creation of effective, efficient links between "central" IRs and specialised instances, between IRs, LMSs and OPACs and hence, between the different copies and versions of items within (and ultimately across) HEIs. This must be under-pinned by the re-use of data and open standards where possible. For instance, if a unique bibliographic record is held in a departmental IR or in the OPAC, *it* should be used as the basis for any "duplicate" record required by the central IR, even if each record is subsequently modified to suit local system/user needs. The key is

making *links* between these expressions/copies available to users; making everything findable as the number of places to find things grows ever more varied and potentially baffling is crucial.

Such processes should be built into the workflows of those administering library information systems - all institutional stakeholders concerned with information systems should be involved in discussions about how such good practice can be achieved.

While they may be part of the solution, Resource Discovery Platforms (RDPs) - however effectively they may be able to pull in records via OAI-PMH - cannot yet be seen to fully expose or reflect the richness of either IRs or OPACs.

10.2. Creating links between systems and records

If LMSs supported open standards and protocols such as OAI-PMH, and IRs accurately catalogued their content according to controlled, recognised vocabularies, records held by OPACs and repositories could be harvested and combined under one roof where, making full use of Institutional IDs or Persistent Identifiers, they could become part of an interoperating "Academic Information Domain" (see, for example, Baars, Dijk, Hogenaar and van Meel, 2006) or an institutional "junction box" of information.

Any number of interfaces could be applied to such a data hub, allowing reports to be generated and statistics to be gathered, with data being "pushed" in and "pulled" out by researchers and teachers for use within a plethora of institutional systems, open, re-usable data becoming commonplace.

This is how forward-thinking library staff are beginning to think of their systems; they do not want fragmentation, they do not want tensions to exist between those who work with one system and those who work with another. All serve the same diverse user community. However, across the board, two of the most significant barriers to achieving goals such as those described above are lack of resourcing and the inconsistent application of standards (there being a clear link between these two conditions).

Link resolvers, Resource Discovery Platforms, SRU/SRW and add-ons such as the LibX toolbar¹⁵ are all solutions to achieving technical interoperability and access integration between IRs and LMSs and are already being used already within many HEI libraries. These technologies allow data from a range of systems to be searched and retrieved for users and clearly offer many benefits, addressing the problem of "siloed" or hidden information.

Yet whatever the compatibility of LMSs and IRs at a technical level, the incompatibility of the standards used to describe the items they contain is visible throughout UK HEIs. Often there are no common authority headings, content or format standards being used, thus exchanging data between systems is hindered. This is nothing new - systems librarians have in the past discussed the fact that, however useful Z39.50 is as a *technical* solution (to the transfer of data from one system to another), the resourcing to address *content* issues (which data sharing often makes more explicit) remains inadequate.

The findings of other JISC projects must be restated yet again:

"Given...the significant number of IR uses which may require interfaces to disparate systems, flexibility and support for interoperability standards and common specifications are not an optional extra, but essential criteria for a repository framework" (Green, 2007).

Certainly these issues are well-known and understood within HEI libraries. So why does such inconsistency remain? In the face of economic and administrative pressures it may be hard for cataloguing staff or systems developers to make realistic, manageable workflows out of ambitious "blue-sky" thinking.

¹⁵ SRU/SRW interfaces provide similar functionality to Z39.50 but use rich URLs (in the case of the former) and a SOAP-based web-service (in the case of the latter). LibX is an open source framework for searching across multiple catalogues and databases from a search bar or menu and "from which editions for specific libraries can be built" (LibX, 2009).

Cross-walks or interchange formats for converting or mapping one metadata format to another (for example MARC to DC or ONIX to MARC21) are already available¹⁶. These can be used to assist in record-sharing efforts. The issue of which standards to support is, however, complex.

Existing (and even fairly recent) profiles and standards used within library systems - including the **Dublin Core Library Application Profile (DC-LAP)** and the **Scholarly Works Application Profile (SWAP)** – may be insufficient for various reasons. Simple DC allows items to be described in only limited ways, and has a 'flat' organisational structure. Indeed, this is one issue which SWAP's developers set out to address. Yet SWAP is generally unsupported at present, partly as it is based on a distributed, relational, multi-file arrangement of metadata which does not exist within any repository software. Its terminology is unclear to some library professionals particularly its "Type Vocabulary", which does not chime with item types as they are currently understood by the LMS community.

It may be that such alternative profiles are simply too granular or resource intensive for use within systems at present, however development communities and working groups seek to address these issues and keeping an eye on the latest versions of such standards and profiles is advisable.

IRs are in some ways more flexible in how they handle metadata than are OPACs, largely because the software they use is more up-to-date and better reflects current thinking in the information field, using terminology which takes into account user needs better than the old LMS MARC-derived fields. This may be because of the need to encourage self-deposit by making the process more straightforward for those not trained as cataloguers.

Conversely, this flexibility is off-set by the fact that most IRs do not make use of professional bibliographic staff and are less aware of the need for standards and their consistent application. In this they lag behind LMSs. Ideally the best of both IR and LMS-centred practices and workflows should be combined.

The development and implementation of standardised lists of item type/format for use within libraries and capable of describing the scopes of both OPACs and IRs would both assist users and improve the potential for semantic interoperability and clarity for staff and end users. This would serve multiple purposes: it would improve the potential for disparate systems to interoperate at semantic level while reducing the need for complex mapping exercises; help library staff comprehend more readily the relationships between the systems which they work with and others administered within their domain; and be beneficial to library patrons/end users.

Any such vocabulary would need to reflect adequately, in clear, simple, modern language, the wide range of item types held by libraries so that end users can better understand what they are able to search for within a catalogue or repository. At the same time, it would need to allow room for lower-level descriptions more suitable for users already conversant with the items types being described.

To take one example: it is surely not sufficient anymore to refer simply to "computer file" [derived from MARC Leader 06/code m] as though the fact that these files are not books or microfilms makes no further distinction necessary.¹⁷ Indeed, this was addressed by the Library of Congress many years ago "to allow for the coding of electronic resources for the significant aspect of the content, rather than their physical form". (Weitz, 2006). These developments and refinements are intended to assist end users and should be made visible to them.

¹⁶ A crosswalk is defined by the Dublin Core Metadata Initiative as "A table that maps the relationships and equivalencies between two or more metadata schemes. Crosswalks or metadata mapping support the ability of search engines to search effectively across heterogeneous databases." (<http://dublincore.org/documents/usageguide/glossary.shtml>). Many useful examples of existing crosswalks are linked to from the UKOLN website: <http://www.ukoln.ac.uk/metadata/interoperability/> [Accessed 12th October, 2009].

¹⁷ The extensive list of MIME media types (or internet file types) maintained by the Internet Assigned Numbers Authority illustrates this point and could be useful. See: <http://www.iana.org/assignments/media-types/> [Accessed 7th September, 2009].

Such an undertaking might involve a reassessment of the terms used by application profiles such as DCLAP the or SWAP, which already aim to create and encourage the use of standardised vocabularies for the exchange of information between library systems at the same time as supporting a wide range of recognised encoding schemes. SWAP already incorporates the concepts of the **Functional Requirements for Bibliographic Records (FRBR)** entity-relationship model.

Resource Description and Access (RDA) and the **International Standard for Bibliographic Description (ISBD)** (IFLA, 2007) already have controlled vocabularies of this type and are likely to impact majorly on research libraries once released, however it is not yet known how well these lists of carrier, media and content types will be understood by end users.

The **Library of Congress Authorities and Vocabularies service** "seek[s] to provide access to commonly found standards and vocabularies promulgated by the Library of Congress" by providing "resolvability to values and vocabularies by assigning URIs" while "assert[ing] relationships to similar concepts found in non-Library of Congress authorities and vocabularies whenever possible" (Library of Congress, 2009). They also have a URI for resource *types*.¹⁸

However, IR developers should bear in mind that LCC is not considered the most suitable scheme for reflecting subjects hierarchically (with corresponding negative implications for creating online browse menus) or for classifying any library collection other than that of the Library of Congress itself. (San Segundo, 2007).

Standards supported by any library system must support not only interoperability but the end users whose needs they ultimately serve.

11. Recommendations

To JISC

- A. Formulate and promote a standardised, controlled vocabulary describing item types, formats and scope, for use across HEI libraries and departments administering IRs, created in consultation with staff from those libraries and their institutions.
- B. Promote the development of - and collect - unique numeric institutional identifiers derived from HEI HR systems to be shared amongst institutions. These would be used as a field to which author/creator names can be mapped, within and across institutions. If desirable they might be also mapped against RAE person identifiers or HEFCE institutional identifiers.
- C. Support the interoperability of subject authorities within and between institutions – this might build on the work of existing projects such as OCLC's Terminology Service Pilot¹⁹ or the High-Level Thesaurus Project (HILT).²⁰
- D. Raise awareness of the impact of current work-practices on future needs (in particular, the need to both utilise open standards to support data linkage within and across institutional systems, as the Semantic Web continues to develop).
- E. Create recommendations on where, or consider developing a service through which, libraries can source high-quality, openly available, "clean" bibliographic records with guarantees that these can be freely re-purposed and edited. This would allow HEI libraries greater certainty about the utility and cost-effectiveness of record-sharing.
- F. Investigate further, and in depth, the interoperability of administrative systems and the opinions of staff working with them, without placing undue emphasis on library systems.

¹⁸ See: <http://www.loc.gov/standards/codelists/resourceTypes.xml> for an XML file outlining these. [Accessed 7th September, 2009].

¹⁹ <http://tspilot.oclc.org/resources/index.html> [Accessed 7th September 2009].

²⁰ <http://hilt.cdlr.strath.ac.uk/> [Accessed 7th September 2009].

- G. Investigate the implications for interoperability of tools/services becoming more common within IRs and LMSs such as Web Services, APIs and Resource Discovery Platforms (RDPs).

To Higher Education Institutions

1. Expose all LMS and IR records for harvesting and linking (except in cases where legal requirements restrict such data re-use) via distributed/federated/meta search using technical protocols such as OAI-PMH, Z39.50, SRU/SRW or link resolvers, as appropriate to the technical infrastructure.
2. Improve co-ordination between all departments possessing institutional information-gathering systems and their staff, with support at the highest levels of the institution, in order to develop efficient workflows, reduce un-necessary duplication of effort and formalise collaboration.
3. Align the systems of both libraries and administrative departments, and their attendant data-processing practices, more closely.
4. Consider establishing a centralised system and attendant workflows for cross-checking and cleaning metadata that is to be shared between systems, to ensure quality, usability and re-usability by both internal and external service providers.
5. Consider options other than batch processing (such as web services or applications, underpinned by open standards) where administrative departments are sharing data with library systems.
6. Develop clear policies on the scopes and uses of IRs and OPACs.
7. Present clearly, comprehensively and comprehensibly, to both staff and end users, the scopes of IRs and OPACs.
8. Develop a single scheme for describing item types/formats and scope within OPACs and IRs, with interoperability requirements and local needs fully accounted for.
 - 8.1. This scheme should be tested with a variety of users (teaching staff, researchers and undergraduates) to ensure it speaks to their needs and is understood by them.
 - 8.2. If JISC act on recommendation A above, library staff should be allowed to attend meetings and contribute to the formulation of a cross-institutional scheme, with local needs discussed as part of the activities of the group.
9. Ensure the use of format and content standards within IRs to avoid the need for future "retro-conversion" or "re-keying".
10. Support the interoperability of subject authorities across institutional systems if common ones are not appropriate – this might build on the work of existing mapping and switching projects such as OCLC's Terminology Service Pilot²¹ and the High-Level Thesaurus Project (HILT)²².
11. Use interchange formats and cross-walks based on open standards more widely and extensively, to assist in the sharing and exchange of records conforming to different format and content standards.
12. Reassess the use of Library of Congress Classification (LCC) within IRs; staff should be familiar with the distinction between LCC as a classification system and LCSH as a subject heading system in order to determine whether these schemes meet the needs of their users and whether they accurately reflect their repository collections/items.
13. Establish controlled name authority lists for staff throughout the institution using agreed, recognised standards, to be made available to all relevant departments.
14. Develop (or if already place, make consistent use of) persistent, institutional or departmental IDs, making these available internally and to other institutions.

²¹ <http://tspilot.oclc.org/resources/index.html> [Accessed 7th September 2009].

²² <http://hilt.cdlr.strath.ac.uk/> [Accessed 7th September 2009].

- 14.1. Relevant institutional systems could possibly hold these IDs in the form of a flat file. This would allow data relating to specific individuals to be "pushed" and "pulled" between various systems. The IDs would become the "glue" allowing information to be disambiguated.
- 14.2. These IDs should be built into metadata workflows and be usable by both staff and end users. Person and role information from various institutional systems should be "warehoused" and made available as "a rich source of contextual metadata" (Green, 2007).
15. Recognise that LMSs are a rich source of bibliographic information on books, book items, monographs, conference proceedings and other items authored by institutional staff (which may not be recorded in an IR) hence an interoperable LMS could be leveraged for research assessment data gathering activities or for use within Research Management Systems.

16. References

- Ashworth, S. (2004). *Glasgow University institutional repositories: a case study*. Presentation available online at: www.lmba.lt/ppt/SusanAshworth.ppt [Accessed 06 August 2009].
- Baars, C., Dijk, E., Hogenaar, A. and van Meel, M. (2006). Creating an Academic Information Domain: a Dutch Example. In Bosnjak, A., and Stempfhuber, M. (Eds.). In *Proceedings: Get the Good CRIS Going: Ensuring Quality of Services for the User in the ERA*. 9th International Conference on Current Research Information Systems. 5-7 June 2008, Maribor, Slovenia, pp.75-85. <http://www.onderzoekinformatie.nl/nl/oi/papers/crisaid/%5C%5Camsterdam.knaw.net%5Cbureau%24%5C%5C%5Cchome%5Cchrisb%5CMaribor%5CPaper+EUROCRIS+2008.pdf> [Accessed 4th September, 2009].
- Chaudhri, T. (2009). Assessing FRBR in Dublin Core Application Profiles. In *Ariadne*. Issue 58. January 2009. Online at: <http://www.ariadne.ac.uk/issue58/chaudhri/> [Accessed 11th September, 2009].
- Freire, N., Borbinha, J. and Martins, B. (2008). Consolidation of References to Persons in Bibliographic Databases. In *Proceedings of the 2008 International Conference on Asia-Pacific Digital Libraries*, 2-5 December, 2008, Bali, Indonesia. Online at: <http://www.springerlink.com/content/m8114g7583367087/fulltext.pdf> [Accessed 19th September, 2009].
- Green (2007). *RepoMMan Final Project Report: September 2007*. Online at: <http://ie-repository.jisc.ac.uk/125/> [Accessed 30th August 2009].
- Higher Education Funding Council for England. (2009). *Bibliometrics*. See: <http://www.hefce.ac.uk/research/ref/Biblio/> [Accessed 2nd September, 2009].
- House of Commons Science and Technology Committee. (2004). Scientific Publications: Free for All? In *10th Report of Session 2003-04*, vol. 1: Report, HC 399-1, London: The Stationary Office, July 20, 2004). Available online at: <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm> [Accessed 17th August 2009].
- IFLA. (1998). *Functional Requirements for Bibliographic Records, Final Report* by the IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K.G. Saur.
- IFLA. (2007). *International standard bibliographic description (ISBD)*. Recommended by the ISBD Review Group; approved by the Standing Committee of the IFLA Cataloguing Section. Preliminary consolidated edition. München : K.G. Saur.
- JISC and SCONUL. (2008). 'Appendix II – Vendor Profiles'. In *JISC & SCONUL Library Management Systems Study. An Evaluation and horizon scan of the current library management systems and related systems landscape for UK higher education*. Compiled by Sero Consulting Ltd., Glenaffric Ltd., and Ken Chad Consulting Ltd. pp.120-156. Online at: www.jisc.ac.uk/media/documents/programmes/resourcediscovery/lmsstudy.pdf [Accessed 06 August 2009].
- JISC. (2009a). *Metadata Enrichment for Repositories in a London Institutional Network*. Online at: <http://www.ucl.ac.uk/lis/merlin/about.shtml> [Accessed 2nd September, 2009].
- JISC (2009b). *Enlighten and Research Integration*. Online at: <http://www.jisc.ac.uk/whatwedo/programmes/inf11/enrich.aspx> [Accessed 2nd September, 2009].
- JISC. (2009c). *Institutional Scholarly comms with integrated publication sharing*. Online at: <http://www.jisc.ac.uk/whatwedo/programmes/vre/insscol.aspx> [Accessed 26th August, 2009].
- Joint Steering Committee for Development of RDA. (2009). *Full Draft of Resource Description and Access*. Online at: <http://www.rda-jsc.org/rdafulldraft.html> [Accessed 9th September, 2009].
- LibX. (2009). *LibX - a Browser Plug-In for Libraries*. [Accessed 3rd October, 2009].

- Library of Congress. (2009). *Authorities and Vocabularies. About*. Online at: <http://id.loc.gov/authorities/about.html> [Accessed 7th September, 2009].
- National Library of Australia et al. (2009). *The Virtual International Authority File*. Online at: <http://viaf.org/> [Accessed 12th September, 2009].
- Research Information Network. (2009). *Creating catalogues: bibliographic records in a networked world. A Research Information Network report. June 2009*. Research Information Network: London.
- San Segundo Manuel, Rosa (2007). Some arguments against the suitability of Library of Congress Classification for Spanish Libraries. In Cordeiro, Maria Ines. (Ed). *Extensions and Corrections to the UDC, 30 (2008)*. UDC Consortium: the Hague, pp. 35-45.
- SHERPA (2006). *Distribution of UK Higher Education Repositories*. Page viii. Online at: http://www.sherpa.ac.uk/documents/rep_distrib.html [Accessed 02 August, 2009].
- Swan, A. and Awre, C. (2006). *Linking UK Repositories: Technical and organisational models to support user-oriented services across institutional and other digital repositories*. Online at: http://ie-repository.jisc.ac.uk/10/1/Linking_UK_repositories_report.pdf. [Accessed 6th September, 2009].
- UKOLN and JISC CETIS. (2008). *EPrints Type Vocabulary Encoding Scheme*. Online at: http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme [Accessed 26th August, 2009].
- University of Cambridge. (2007). *About the Newton Catalogue*. Online at: <http://ul-newton.lib.cam.ac.uk/help/about.htm>. [Accessed 21st August, 2009].
- University of Cambridge. (2009a). 'About Newton'. Online at: http://ul-newton.lib.cam.ac.uk/vwebv/ui/en_US/htdocs/help/about.htm [Accessed 24th August, 2009].
- University of Cambridge. (2009b). Tower Project website. Online at: <http://www.lib.cam.ac.uk/deptserv/towerproject/> [Accessed 21st August 2009].
- University of Cambridge. (2009b). *Information Strategy and Services Syndicate. Minutes of the meeting held on Thursday 19 March 2009*. Page 2. Available online at: <http://www.admin.cam.ac.uk/committee/iss/m/minutes/20090319.pdf> [Accessed 21st August 2009].
- University of Glasgow. (2008). *Publications Policy*. Online at: <http://www.lib.gla.ac.uk/enlighten/publicationspolicy/> [Accessed 05 August 2009].
- Wales Higher Education Libraries Forum. (2009). See: <http://whelf.ac.uk/wrn/index.shtml> [Accessed 10th September, 2009].
- Weitz, J. (2006). 'Cataloging Electronic Resources: OCLC-MARC Coding Guidelines.' Online at: <http://www.oclc.org/support/documentation/worldcat/cataloging/electronicresources/#3> [Accessed 11th September, 2009].

Technical Glossary

Key terms used by the project

Interoperability - Generally, interoperability refers to how well two or more systems work together to achieve a common goal; for the purposes of OCRIS this means direct processing by one automated system or sub-system, of data provided by another. This will usually be assisted by the use of standards and standard protocols. There are different types of interoperability. The ones most relevant to this work are **technical (or, syntactic) interoperability** and **semantic interoperability**.

- A. Technical interoperability** - If two or more systems are capable of communicating and exchanging data, they are exhibiting syntactic interoperability. Specified data formats, communication protocols and the like are fundamental. In general, XML or SQL standards provide syntactic interoperability. Technical / Syntactical interoperability is required for any attempts of further interoperability - <http://en.wikipedia.org/wiki/Interoperability>.
- B. Semantic interoperability** - The ability of two or more systems or components to exchange or harmonize cognate subject vocabularies and/or knowledge organization schemes to be used for the purpose of effective and efficient resource discovery without significant loss of lexical or connotative meaning and without special effort by the user - <http://www.und.nodak.edu/dept/library/Departments/abc/SACSEM-SemInGlossary.htm>.

It is also worth noting that " interoperability is often more of an organizational issue: often interoperability has a significant impact on the organizations concerned, raising issues of ownership (do people want to share their data?), labour relations (are people prepared to undergo training?) and usability. In this context, a more apt definition is captured in the term "business process interoperability" - <http://en.wikipedia.org/wiki/Interoperability#Software>.

Metadata - Metadata is data that describes other data. For example, a library catalogue record is metadata, because it describes an item such as a book or DVD. Another example is a data dictionary which describes the structure and content of a database.

Item Types - the terms used by the OCRIS project to describe item types, particularly in its questionnaires, are taken from the Scholarly Works Application Profile (SWAP) EPrints type vocabulary, with some additions.

Abstract: A summary or abstract of a journal or book item, or book.

Archival material: Material created for institutional administrative and management purposes. Includes publicity material such as photographs and graphics, and material pertaining to the history of the institution.

Book: A non-serial publication that is complete in one volume or a designated finite number of volumes. In physical form, a book is a collection of sheets of paper, parchment or other material bound together along one edge within covers. Books are often identified with an ISBN.

Book item: A defined chapter or section of a book, usually with a separate title or number.

Book review: A review of a book.

Conference item: A contribution to a conference, workshop or other event. Where the contribution is a paper or poster, use Conference Paper or Conference Poster as appropriate.

Conference paper: A paper submitted and/or presented at a conference, workshop or other event.

Conference poster: A poster submitted and/or presented at a conference, workshop or other event.

Journal article: An article or paper published in a journal. For pre- and post-prints, use Submitted Journal Article. For book reviews, use Book Review. For news items use News Item. For other types of contribution use Journal Item.

Journal item: A contribution to a journal, other than an article, e.g. an editorial, conference report, debate, letter or response. Where the contribution is an article, use Journal Article.

Learning object: An item created by institutional staff for use in teaching.

News item: A news item.

Other: Any material not categorised by another option.

Patent: A patent or patent application.

Report: A research, statistical or technical report issued by an institution, agency, government body or other organisation.

Research dataset: A set of data gathered during research and subsequently processed for research outputs.

Scholarly text: A scholarly text that is primarily words for reading. Where the text is one of the more specific types listed here, use the more specific term.

Student coursework: Material created by a student during the course of their studies.

Submitted journal article: The author's original manuscript as submitted to and/or accepted by a journal. In the terminology recommended by the [Sherpa project], a submission prior to peer review is a preprint; a submission after peer review is a post-print. Therefore, a Submitted Journal Article can be either a preprint or a post-print. For the version of the article published by the journal, use Journal Article.

Thesis or dissertation: A thesis or dissertation submitted in completion of a course of study at an institution of higher education.

Website content: Material created for institutional websites. For institutional archival material, use Archival material.

Working or discussion paper: A working or discussion paper circulated publicly or among a group of peers. Certain disciplines, for example economics, issue working papers in series. Working or discussion papers may form the basis for a Journal Article or Conference Paper.

Metadata formats:

context_object: A metadata format that represents an Entity that is referenced in a networked environment, along with Entities that constitute the context in which the Referent is referenced.

DC qualified: Dublin core (qualified). Including any modification of simple DC not given as an option: use uketd_dc, primo_dc, SWAP, and UK-LOM as appropriate.

DC simple: Dublin core (simple). Including oai_dc.

DIDL: Digital item declaration language. An XML dialect standardised in the MPEG-21 standard, aimed at defining an open framework for multimedia applications (specifically, a means of protecting rights and permissions).

imsmd: IMS Global Consortium standard for IMS learning resources.

in-house: Local format, including customised database.

MARC: Machine-readable cataloguing. Including UKMARC, UNIMARC, MARC21, MARCXML.

METS: Metadata encoding and transport schema.

MODS: Metadata object description schema. A schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications.

o-ex: XML Schema of the Open digital rights language. An open standard for rights expressions, intended to provide flexible and interoperable mechanisms to support the use of digital content across sectors and communities.

primo_dc: The Dublin core application profile used by the Primo package.

RDF: Resource description framework. A general-purpose language for representing information on the web, using subject-predicate-object expressions (triples) wherein the subject is a URI or a blank node denoting a resource.

rfc1807: This RFC defines a format for bibliographic records describing technical reports.

Scholarly Works Application Profile (SWAP): A Dublin Core Application Profile for describing scholarly works (EPrints) held in institutional repositories. Formerly known as the EPrints Application Profile. For further information see:
http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme.

uketd_dc: UK Electronic Theses and Dissertations Dublin Core. Developed for EThOS. A Qualified DC application XML Schema (based on Dublin Core with some EThOS specific added terms).

UK-LOM: The UK LOM Core is an application profile of the IEEE LOM that has been optimised for use within the context of UK education.

VRA4: VRA Core 4.0 is a data standard for the cultural heritage community. It consists of a metadata element set as well as an initial blueprint of how those elements can be hierarchically structured.

Authority controlled personal/corporate name schemes and standards:

In-house: Local authority file.

ISAAR(CPF): International Standard Archival Authority Record. For Corporate Bodies, Persons and Families

Library of Congress: Library of Congress Name Authority File.

OCLC: OCLC Online Computer Library Center.

Other: Authority file other than LCNAF, ISAAR(CPF), OCLC, VIAF, or local/in-house.

VIAF: Virtual International Authority File.

Authority controlled subject schemes and standards:

DDC: Dewey Decimal Classification.

In-house: Local subject vocabulary.

JACS: Joint Academic Coding System.

LCC: Library of Congress Classification.

LCSH: Library of Congress Subject Headings.

MeSH: Medical Subject Headings.

Other: Subject or classification authority file other than DDC, LCSH, or UDC.

UDC: Universal Decimal Classification.

Technologies and systems referred to in the Final Report

API - Application Programming Interfaces - A formalized set of software calls and routines that can be referenced by an application program in order to access supporting system or network services. An API defines the calling conventions and other information needed for one software module (typically an application program) to utilize the services provided by another software module -
<http://www.cise.ufl.edu/research/ParallelPatterns/glossary.htm>.

ATOM - An XML-based syndication format that is used to publish headlines of the latest updates on blogs and Web sites for retrieval by users and other sites. Based on RSS 2.0, Atom was turned over to the IETF for standardization. Most news aggregators support Atom along with the traditional RSS formats - http://www.pcmag.com/encyclopedia_term/0,2542,t=atom&i=38122,00.asp.

Batch processing - A mode of data processing in which data is gathered over a period of time and aggregated for subsequent sequential processing - <http://www.gbc.hu/english/bszotare1.htm>.

CERIF (Common European Research Information Format) - Developed with the support of the EC (European Commission), CERIF is a standard; technically it is an EU (European Union) Recommendation to member states and is intended to facilitate data exchange and the resolution of schema differences between heterogeneous distributed databases - primarily those of Current Research Information Systems (CRISs).

Its essential features are: (a) it has the concept of objects or entities with attributes such as project, person, organisational unit; (b) it supports n:m relationships between them (and recursively on any of them) using 'linking relations' thus providing rich semantics including roles and time; (c) it is fully internationalised in language and character set; (d) it is extensible without prejudicing the core data model thus providing guaranteed interoperability at least at the core level but not precluding even richer intercommunication. It is designed for use both for data exchange (data file transfer) and for heterogeneous distributed query / result environments - <http://www.eurocris.org/cerif/introduction/>.

Crosswalks - A crosswalk is defined by the Dublin Core Metadata Initiative as "A table that maps the relationships and equivalencies between two or more metadata schemes. Crosswalks or metadata mapping support the ability of search engines to search effectively across heterogeneous databases" - <http://dublincore.org/documents/usageguide/glossary.shtml>). Many useful examples of existing crosswalks are linked to from the UKOLN website: <http://www.ukoln.ac.uk/metadata/interoperability/>.

Current Research Information System (CRIS) - A database or other information system storing data on current research by organizations and people, usually through some kind of project activity, financed by a funding programme.

Digital Object Identifier (DOI) - an alphanumeric name that identifies digital content, such as a journal article. The DOI is paired with the content's electronic address, or URL, in an updateable central directory, and is published in place of the URL in order to avoid broken links while allowing the content to move as needed - <http://www.foodinfoquest.com/glossary>.

Dublin Core Library Application Profile (DCLAP) - a possible application profile that clarifies the use of the Dublin Core Metadata Element Set in libraries and library-related applications and projects. It was originally prepared by the DCMI-Libraries Application Profile drafting committee, a subset of the DCMI-Libraries Working Group - <http://dublincore.org/documents/library-application-profile/>.

Electronic Resources Management (ERM) system - Electronic Resource Management is a system encompassing a wide range of functions throughout the electronic resource life cycle, including but not limited to ordering, acquisition, maintenance and renewal processes, the generation and maintenance of discovery tools such as e-resource web pages, and recording and presentation of license information such as authorized users and permitted uses - <http://www.loc.gov/acq/conser/DLF-eresource-management.html>.

Functional Requirements of Bibliographic Records (FRBR) - FRBR is the product of a study undertaken following the 1990 Stockholm Seminar on Bibliographic Records "to delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs." FRBR does so by means of a conceptual model that identifies and defines: (1) entities of interest to users of bibliographic records; (2) their attributes; and (3) the relationships that operate between them. (Source: Ed Jones, FRBR summary for Summit).

Harvesting (metadata harvesting) - Automatically gathering metadata that is already associated with a resource, and which has been produced via automatic or manual means. Metadata harvested may be attached to a document (e.g., it may be encoded in the header of a Web resource), or it may be found in a metadata registry or database - <http://dublincore.org/groups/tools/glossary.shtml>.

Identity Management System - A system that provides a cluster of services relating to Identity Management. The central service is Authentication. The system may also support other services, such

as Pre-Authentication, Authorisation, Single Sign-On, Identity repository management, a synchronisation management facility, user self-service registration, user self-service capabilities, and audit - <http://www.rogerclarke.com/EC/IdAuthGloss.html>.

iGoogle - (formerly Google Personalized Homepage and Google IG), a service of Google, is a customizable AJAX-based start page or personal web portal. Google originally launched the service in May 2005. Its features include the capability to add web feeds and Google Gadgets (similar to those available on Google Desktop) such as news feeds, content and games - <http://en.wikipedia.org/wiki/IGoogle>.

International Standard for Bibliographic Description (ISBD) - a principal standard to promote universal bibliographic control, to make universally and promptly available, in a form that is internationally acceptable, basic bibliographic data for all published resources in all countries - <http://www.ifla.org/en/about-the-isbd-review-group>.

Interchange format - An interchange format is used to exchange data from one digital content creation tool to another. The goal for such a format is to be able to transfer files back and forth between applications without data loss - <http://www.khronos.org/collada/faq/>. Being automatically parsable into syntactic elements, an interchange format is machine readable, relying on unambiguous syntax and the discrete representation of elements. Examples of interchange formats include the JPEG File Interchange Format (JFIF) and the Data Interchange Format (DIF).

LibX - a browser plug-in for Firefox and Internet Explorer that provides direct access to a library's resources, created by Virginia Tech Libraries and the Virginia Tech Department of Computer Science. LibX is an Open Source framework from which editions for specific libraries can be built. Currently, 660 academic and public libraries have created public LibX editions - <http://www.libx.org/>.

Link Resolvers - Software that converts a logical address or meta-data into the physical URL of the target data - http://www.pcmag.com/encyclopedia_term/0,2542,t=link+resolver&i=58073,00.asp. These URLs are known as **OpenURLs** (see below).

Metasearch (aka distributed or federated search) - Cross-database search tools that can search multiple catalogues, online databases, search engines, or commercial databases. They can often merge and de-duplicate results and provide unified access to a variety of information resources - <http://www.loc.gov/acq/conser/glossary.html>.

MARC (MACHine-Readable Cataloguing) record - A widely used set of format standards (with many national and international variants) for the encoding and exchange of data in machine-readable library catalogue cards. For further information see: <http://www.loc.gov/marc/faq.html>.

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) - A protocol created to facilitate discovery of resources distributed in many repositories or locations. The OAI-PMH achieves this by providing a simple, yet powerful framework for metadata harvesting. Harvesters can incrementally gather records contained in OAI-PMH repositories and use them to create services covering the content of several repositories - <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

ONIX (ONline Information eXchange) - a metadata format for publishers to distribute electronic information about print and electronic books. For further information see: <http://www.editeur.org/83/Overview/>.

Open Architecture - An architecture that employs open standards (i.e. the specifications are made public) for key interfaces within a system, allowing for the adding, upgrading and swapping of components.

Open Standards - Standards that are widely used, consensus based, published and maintained by recognized industry standards organizations - <http://www.acq.osd.mil/osjtf/termsdef.html>.

Open URL - A URL that refers to an article, book or other resource rather than the physical location of the item on the Web. Developed for use in libraries and other repositories of information, the

OpenURL contains the address of an OpenURL Web site combined with the item's identification. The OpenURL Web site serves as a resolver and finds the physical address of the item. Developed by the National Information Standards Organization (NISO), it was turned over to the Online Computer Library Center (OCLC), Dublin, OH (www.oclc.org) in 2006 - http://www.pcmag.com/encyclopedia_term/0,2542,t=OpenURL&i=58119,00.asp.

Plug-in - A software extension that enables added capabilities. Plug-ins are usually available at software developer's web sites - <http://www.c7.ca/glossary/>. Popular examples include those made available by Google, such as 'Wikipedia Search', 'Google Calendar' and 'Google Earth plug-in' which allows users to embed Google Earth into their webpages. Many web-based Content Management Systems use plug-ins to offer extended functionalities on top of their basic services.

Reference Management Software (or, Reference Management Tool) - software for scholars and authors to use for recording and utilising bibliographic citations (references). Once a citation has been recorded, it can be used time and again in generating bibliographies, such as lists of references in scholarly books, articles and essays. The development of reference management packages has been driven by the rapid expansion of scientific literature.

These software packages normally consist of a database in which full bibliographic references can be entered, plus a system for generating selective lists or articles in the different formats required by publishers and scholarly journals. Modern reference management packages can usually be integrated with word processors so that a reference list in the appropriate format is produced automatically as an article is written, reducing the risk that a cited source is not included in the reference list. They will also have a facility for importing the details of publications from bibliographic databases. - http://en.wikipedia.org/wiki/Reference_management_software.

Research Management System (RMS) - a research management system "needs to cover the gamut of activity areas, including researcher expertise and funding opportunity searches, full application and award project management, ethical clearances and post-graduate student supervision, facilitating research partnerships, the recording of research outputs, and captured management of commercialisation initiatives". They provide an overview of an 'ideal' research management system, highlighting the need for integration with other corporate systems in order to manage the "inter-related business processes involved in research administration". The challenge for Research Offices has been to source and implement systems that satisfy a diverse range of functions, often with limited budgets.

The diverse range of functions supported by the research office has meant that customised, in-house developments have commonly been used to satisfy research management needs. Recent quality assessment mechanisms have seen a greater emphasis placed on research output management that have lead in some cases to greater involvement between the Research Management System and institutional digital repositories, most often managed by University libraries. Greater emphasis is also being placed on management level reporting on research performance, requiring the Research Management System to be integrated with corporate information reporting systems, often involving data warehousing and Business Intelligence toolsets (e.g. University of Adelaide ORBIT project). - <http://heswiki.onconfluence.com/display/radwiki/Research+Management+Systems>.

Resource Discovery Platform (RDP) - a service-oriented layer acting on top of a Library Management System (LMS), usually complementing (but sometimes replacing) the traditional OPAC. RDPs are capable of searching across multiple databases or data sources and tend to focus strongly on the presentation layer and User Interface (UI), presenting results in ways familiar to "Google" or "Web 2" users - for example, the use of autosuggest ("Did you mean..?"), faceted navigation menus for results lists, tag clouds or front-cover image retrieval for books and CDs. Examples of RDPs include Encore, developed by Innovative Interfaces Inc. and Primo, developed by Ex Libris.

Resource Description Framework (RDF) - A framework for constructing logical languages that can work together in the Semantic Web. A way of using XML for data rather than just documents. For further information see: <http://www.w3.org/RDF/>.

RSS (Really Simple Syndication) - A syndication format that was developed by Netscape in 1999 and became very popular for aggregating updates to blogs and the news sites - http://www.pcmag.com/encyclopedia_term/0%2C2542%2Ct%3DRSS&i%3D50680%2C00.asp.

SRW - Search/Retrieve Web service. SRW is an XML-based protocol designed to be a low-barrier-to-entry solution for searching and other information retrieval operations across the Internet - <http://www.twason.com/glossary.html>.

SRU - SRU stands for "Search/Retrieve via URL". According to the Library of Congress, "SRU (Search/Retrieve via URL) is a standard search protocol for Internet search queries, utilizing CQL (Common Query Language), a standard query syntax for representing queries." In other words, it is a URL that includes search and retrieval commands - <http://www.white-clouds.com/iclc/cliej/cl24wen.htm>.

SRU/SRW interfaces provide similar functionality to **Z39.50** (see below) but use rich URLs (in the case of the former) and a SOAP-based web-service (in the case of the latter).

Semantic Web - The Web of data with meaning in the sense that a computer program can learn enough about what the data means to process it - <http://www.w3.org/People/Berners-Lee/Weaving/glossary.html>.

Service Oriented Architecture (SOA) - SOA is an IT architecture strategy for business solution (and infrastructure solution) delivery based on the concept of service-orientation - http://blog.elementallinks.net/2005/07/my_soa_definiti.html.

Software as a Service (SaaS) - a model of software deployment whereby a provider licenses an application to customers for use as a service on demand. SaaS software vendors may host the application on their own web servers or download the application to the consumer device, disabling it after use or after the on-demand contract expires. The on-demand function may be handled internally to share licenses within a firm or by a third-party application service provider (ASP) sharing licenses between firms - <http://en.wikipedia.org/wiki/SaaS>.

Switching language - Intermediary terms that serve as a mechanism for moving between vocabularies; unlike links, which are internal, switching language is external to records for the terms being associated - <http://www.und.nodak.edu/dept/library/Departments/abc/SACSEM-SemInGlossary.htm>.

SWORD (Simple Web-service Offering Repository Deposit) - "SWORD is a lightweight protocol for depositing content from one location to another. It is a profile of the Atom Publishing Protocol (known as APP or ATOMPUB). SWORD has been funded by the Joint Information Systems Committee to develop the SWORD profile and a number of demonstration implementations". For further information see: <http://www.swordapp.org/>.

Symplectic Publications System - publications management software (designed for academic institutions) is currently being used within Cambridge. For further information see: <http://www.symplectic.co.uk/>.

Syndication format - A publishing format that lets people view headlines of the latest updates from their favourite blogs and Web sites all from within a single newsreader program. The major syndication formats are RSS and Atom, and most newsreaders support both formats - http://www.pcmag.com/encyclopedia_term/0,2542,t=syndication+format&i=52356,00.asp.

SFX - early and popular example of a **link resolver** for use with **Open URLs** (see below). It is now owned by Ex Libris, the Library Management Systems vendor.

Sakai - "The Sakai Community develops and distributes the Open Source Sakai CLE, an enterprise-ready collaboration and courseware management platform that provides users with a suite of learning, portfolio, library and project tools" - <http://sakaiproject.org/portal>.

Uniform Resource Identifier (URI) - A Uniform Resource Identifier (URI) is a string that identifies a particular location in a file system or on the web. For example, the address of a web page is a URI - <http://library.gnome.org/admin/system-admin-guide/stable/glossary-1.html.en>.

Vertical Searching - A search service which is focused on a particular field, a particular type of information, or a particular information format - <http://www.baseonsearch.co.uk/search-glossary/>.

Web Service - One type of service that can be part of an SOA infrastructure, a web service is defined by a set of technologies that provide platform-independent protocols and standards used for exchanging data between applications - http://www.skywaysoftware.com/resources/terminology/#Web_Service.

WorldCat - "the focal point of OCLC's Web-scale strategy. Both a Web portal to the WorldCat catalogue and a supporting program of data syndication that draws users from other popular Web destinations, it presents a common, relevant and compelling Web presence for libraries that promotes local content and value" - <http://www.oclc.org/WorldCat/web/default.htm>. WorldCat makes available online the holdings information of over 10,000 contributing libraries.

XML - A set of rules for encoding documents electronically. It is defined in the XML 1.0 Specification produced by the W3C and several other related specifications; all are open standards. XML's design goals emphasize simplicity, generality, and usability over the Internet. Although XML's design focuses on documents, it is widely used for the representation of arbitrary data structures, for example in web services. - <http://en.wikipedia.org/wiki/XML>. For further information see: <http://www.w3.org/XML/>.

Zotero - an example of **Reference Management Software**. Zotero is "a free, easy-to-use Firefox extension to help you collect, manage, and cite your research sources."

Z39.50 - A "pre-Web" client/server based service and protocol for Information Retrieval, commonly used by libraries. It specifies procedures and formats for a client to search a database provided by a server, retrieve database records, and perform related information retrieval functions. The protocol addresses communication between information retrieval applications at the client and server; it does not address interaction between the client and the end-user - <http://www.loc.gov/z3950/agency/Z39-50-2003.pdf>.

Appendices

The following appendices to the OCRIS Final Report are available online as separate files.

Appendix 1 - OCRIS Questionnaires and Guidance

Appendix 2 - Case Studies in 2 Higher Education Institution Libraries [not yet available]

Appendix 3 - In-depth studies of a sample of IRs and OPACs