Introduction

- Todd Carpenter, Managing Director, NISO

OCRIS: Online Catalogue and Repository Interoperability Study
- Kathleen Menzies
  Researcher, Centre for Digital Library Research
  University of Strathclyde
- Gordon Dunsire, Head, Centre for Digital Library Research

IDEALS Repository at the University of Illinois at Urbana-Champaign
- Sarah Shreeves
  IDEALS Coordinator, Scholarly Commons Coordinator
  University Library, University of Illinois at Urbana-Champaign

Outline

- Introduction
- Aims of the project and methodology
- Main findings
- Requirements for and barriers to interoperability
- Institutional case studies and good practice
- Aspects of duplication
- Recommendations

Key Findings, Implications and Recommendations
Presentation by Kathleen Menzies, Centre for Digital Library Research, Strathclyde University, Glasgow, Scotland.

Project overview

- 3-month project funded by Joint Information Systems Committee (JISC)
- Survey of UK Higher Education Institutions
- Carried out at Centre for Digital Library Research, University of Strathclyde
- Project team
  —Duncan Birrell, Gordon Dunsire, Kathleen Menzies

Aims and Objectives

- 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.
Methodology
- Quantitative and qualitative – a questionnaire sent to staff at 85 Universities.
  - Staff managing/working with the LMS and OPAC
  - Staff managing/working with the IR
  - Staff in administrative departments
  - 1149 individuals contacted in total.
  - Those working with IRs were the most responsive: 42 percent.
- 2 in-depth case studies within the main libraries at: University of Cambridge and University of Glasgow.
- In-depth desk research into the IRs and OPACs of 10 HEIs.

Findings
Scopes and Standards
- There is significant scope overlap (81%) for all item types (we listed 21 based on SWAP) held in IRs and OPACs.
- Duplication at both record and item level is frequent, especially for print/electronic copies of theses and bibliographic data about links to journal articles.
- Authority control within LMSs is very high; format and content standards are well supported; within IRs standards are not applied adequately and often not at all.
- The use of vocabularies and standards within library systems in any given HEI is fragmented and disjointed, as are the majority of HEI systems, in all regards.
- Some institutions are considering solutions.

Scope overlap between IRs and OPACs
Scope overlap may only be partial. However, the scoping distinctions and boundaries for IRs and OPACs are becoming increasingly blurred. Many IRs contain bibliographic data only while OPACs link to the full-text version. Instances of both partial and full scope overlap are significant.

Authority control

Interoperability
- The majority of IRs/LMSs (65%, approx.) DO interoperate with at least one other administrative system (HR, Finance, Research Assessment, VLEs) but not yet with a range of systems.
- Only 2% of IRs and LMSs currently interoperate with each other. In 11% of cases interoperability is pending. This is based on a 41% (31) response rate from the 85 in-scope HEIs.
- Links between IRs and Research Management Systems are often discussed – but what about LMSs? They are a rich source of bibliographic information on non-journal items.
- There is clearly a need for such information to be accessible to other systems, especially when items recorded are staff outputs.
The Research Excellence Framework (REF) has been a major factor in the discussions about interoperability. 65% of IRs plan to develop services in support of the REF. 11% already do so. At present, IR services are generally limited to the generation of usage statistics, advice on Open Access and metadata enhancement/verification. Many IRs wish or plan to extend these services. Most LMSs focus most strongly on end users and resource discovery.

A one-stop shop and a shop window - both administrative and a user interface. The same data pool could be used for a variety of purposes, repurposed to suit various needs. Data could be shared efficiently within and across HEIs, and with external agencies. Many institutions are growing increasingly aware of the utility of unique person or author IDs as a common referent. Subject and format-specific IRs - not just centralised IRs - have a valuable role in serving specific communities.

Systems and services are only as good as the weakest link. Aggregation services searching within or across HEIs require a robust, flexible model for data exchange. Open standards are essential. Standards must be consistent for sharing and translation. Services must be supported by systems which provide all users with what they need and which reflect their expectations.

Lack of resourcing puts a strain on the abilities of bibliographic services to work across both IRs and LMSs. Most IRs are Open Source. Technical skills needed to develop or customise systems may not be there. Proprietary systems make this prohibitive or expensive, especially for LMSs. Many admin systems are bespoke, customised to very specific local needs. Administrative staff are not sufficiently involved in cross-departmental collaborations with library systems staff. Lack of top-down leadership or clear policies informing strategy, procurement, and workflows.
Good practice at Strathclyde

- Name authority control being implemented within Strathprints (University of Strathclyde IR), which is jointly administered with the library.
- ‘Notes’ field within Strathprints records link users from ‘main’ record to other versions stored in the database.
- Enhancing Repository Infrastructure for Scotland (ERIS) project is looking at improving and standardising subject metadata.
- Installation of Resource Discovery Platform (RDPs); IR and LMS can be cross-searched through one interface.

Notes field within Strathprints records link users from 'main' record to other versions stored in the database.

Enhancing Repository Infrastructure for Scotland (ERIS) project is looking at improving and standardising subject metadata.

Installation of Resource Discovery Platform (RDPs); IR and LMS can be cross-searched through one interface.

Only a partial solution

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 true examples of duplication?

If items are purposed in different ways or targeted at distinct types of user, aren’t these positive examples of duplication?

Duplication and scope overlap serve a useful purpose in some cases.

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

Aspects of duplication

Examples above show the University of the Creative Arts’ Visual Arts Data Service repository and the University of Southampton’s eCrystals repository for chemical crystallography datasets.

“Designated Communities”

- Different sites used by different types of user community (whether as depositors or end users).
- Some may feel more at home with a departmental IR rather than an institutional one.
- Some may prefer electronic journals catalogues where they can look at what other content appears in a given issue.
- The menus in a “classic catalogue” might be preferable to those of an RDP.
- Copies may be different manifestations or offer different functionalities.
- Can set up IR to offer users different views of the same repository (i.e., giving the appearance of a distinct, separate service).
Recommendations to Universities

- Improve coordination between all departments, with support at the highest levels in order to develop efficient workflows, reduce duplication of effort and formalise collaboration.
- Align the systems of libraries and administrative departments more closely.
- Consider establishing a centralised system and workflows for cross-checking and cleaning metadata to be shared between systems, ensuring quality, usability, and reusability.
- Expose all LMS and IR records for harvesting and linking via distributed/federated/meta search using technical protocols such as OAI-PMH, Z39.50, SRU/SRW, or link resolvers.
- Use common, open standards wherever possible.

Recommendations to Universities (cont...)

- Unique numeric, institutional identifiers derived from HR systems to be shared amongst institutions. These could be:
  - Used as a field to map author/creator names
  - Mapped against RAE person identifiers or HEFCE institutional identifiers.
- Formulate a controlled vocabulary describing item types, formats and scope, for use across HEI libraries, created in consultation with staff.
  - Keep up to date with existing/new standards or vocabularies (e.g., MIME media types, RDA, FRBR).

The LMS is more relevant than ever!

Recognise and discuss LMSs as a rich source of bibliographic information that could be leveraged for research management/assessment systems and multiple types of data sharing.

Contact Info:
Kathleen Menzies
Researcher
Centre for Digital Library Research,
University of Strathclyde
Email: klmenzies@cis.strath.ac.uk
Tel: +44 (0) 141 548 4752

Thank-you!

Outline

- Basic interoperability/integration issues for IRs
- IDEALS background, collection policy, metadata policy
- Interactions with library systems
- Challenges and areas for further work
Issues in the IR environment

- Large range of types and granularity of material from diverse disciplines
- Reliance on author deposit and description - tension between ease of deposit and full description
- Reliance on full text indexing / discovery via Google and other search engines
- Limited resources for technical development and resource description

Quick Facts:
- Joint initiative of Library and CITES (central IT)
- DSpace platform (1.5.2)
- Over 13,000 items
- Over 1.2 million downloads since July 2007
- Some representation from about 70% of colleges on campus (note: not faculty)
- Most material from physical sciences, agricultural sciences, library and information science, social sciences
- Faculty, researchers, graduate students can deposit at any time

IDEALS: http://www.ideals.illinois.edu/

Scholarship and Research

Also audio, video, software, undergraduate research

What of this content would traditionally appear in library discovery systems?

What SHOULD appear in library discovery systems?

IDEALS Metadata Policy

- Use Qualified Dublin Core - out of the box DSpace with few corrections / additions
- Very few required fields: title, date of publication, a subject keyword, type of publication
- No built-in controlled vocabularies / authority control but some help provided during deposit process
- Enforce a few encoding schemes for date, language, type

http://services.ideals.illinois.edu/wiki/bin/view/IDEALS/MetadataPolicy
Wide Range of Metadata Quality

Example One: Digitized Monographic Type Material

- Transform MARC records into QDC for IDEALS
  - Standard mapping using XSLT
  - Always include Bib ID and OCLC Identifier
  - Do not use all information in MARC record (physical description, for example)
  - Include extra information – notably copyright
- Pushing Handles (persistent links) back to OPAC
  - Use Bibliographic ID to generate regular report
- Cataloging and Metadata staff update MARC records
- Show up in downstream applications like WorldCat
Example Two: Digitized Serial Runs

- Transform MARC record into QDC for IDEALS (sometimes)
- Each volume (and usually issue) has its own record in IDEALS so add volume, year information (often automated)
- Push handle for COLLECTION to OPAC or to Online Resources List
- Done as a one-off, manual effort
- If in MARC record will show up in downstream applications like WorldCat
Example Three:

- Material first published in IDEALS
- In planning stages – not yet integrated
- Theses, dissertations, grey literature – challenge of identifying appropriate collections
- Harvest sets of records via OAI-PMH
- Map MODS record into MARC format
- Make available in OPAC
- Will appear in downstream applications like WorldCat

But what about discovery of all the preprints, data sets, presentations, software, etc?
Challenges/Areas for Further Work

- What metadata should go where?
- Differences in quality, granularity, and needs
- Lack of organizational coordination – big library, very decentralized
- Lack of infrastructure to ease integration – author identifiers for example
- Need to do better tracking of where people are coming from – does it matter?

From ILS to Repository and Back: Data Interoperability

Questions?

All questions will be posted with presenter answers on the NISO website following the webinar:

www.niso.org/news/events/2010/datainterop