

## From ILS to Repository and Back: Data Interoperability

[www.niso.org/news/events/2010/datainterop](http://www.niso.org/news/events/2010/datainterop)



- **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

NISO Webinar • January 13, 2010



## NISO 2010 Events

<http://www.niso.org/news/events/2010/>

Going to ALA? Visit the NISO website to find out about events highlighting NISO projects and other standards work.

### 2010

- January 13 (Webinar): From ILS to Repository and Back: Data Interoperability
  - February 8 (Open Call): Single Sign-On Authentication Working Group Update - **Free!**
  - February 10 (Webinar): What It Takes To Make It Last: E-Resources Preservation
  - March 8 (Open Call): ERM Data Standards & Best Practices Review Working Group Update - **Free!**
  - March Two-Part Webinar: Identifiers: New Problems, New Solutions
  - ★ March 10: What's in a Name? Latest Developments in Identifiers
  - ★ March 17: Content Identification: What's New with ISO Identifiers, ISBN and ISTC
  - March 23 (Atlanta, GA): Discovery to Delivery: Creating a First-Class User Experience
- \* Buy 4 webinars, get 2 free  
\* Buy 6 webinars, get 7 free

NISO Webinar • June 10, 2009



## Key Findings, Implications and Recommendations

Presentation by Kathleen Menzies, Centre for Digital Library Research, Strathclyde University, Glasgow, Scotland.



NISO Webinar • January 13, 2010

## 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

## 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

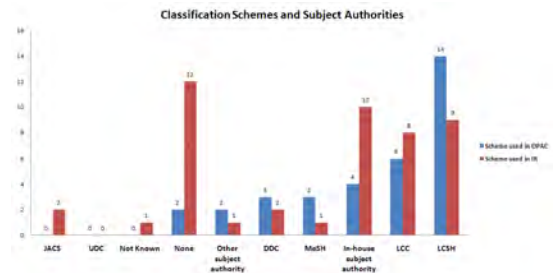
- There is **significant scope overlap** (81%) for *all* item types (we listed 21 based on SWAP) held in IRs and OPACs.
- Duplication at both record and item level is frequent, especially for print/electronic copies of theses and bibliographic data about/links to journal articles.
- Authority control 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

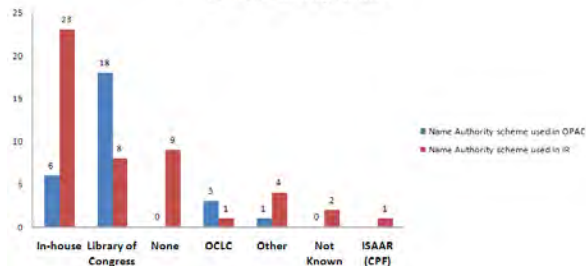


Scope overlap may only be partial. However, the scoping distinctions and boundaries for IRs and OPACs are becoming increasingly blurred. Many IRs contain bibliographic data only while OPACs link to the full-text version. Instances of both partial and full scope overlap are significant.

## Authority control



## Name 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.

## Services

- 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.

## What is the potential?

- 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.

## Requirements

- 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.



## Case Study Institutions

### Cambridge University Library

- Federated structure.
- Of 108 libraries, 89 contribute records to the CUL catalogue, Newton.
- LMS is Voyager, with WebVoyage acting as the OPAC.
- The main IR at Cambridge is DSpace@Cambridge.
- There are several other non-library IRs administered at departmental level. Some use their own classification schemes.

### Glasgow University Library

- LMS is Millennium, with WebPAC Pro acting as the OPAC. All 6 affiliate libraries contribute records.
- A Resource Discovery Platform has recently been installed.
- The GUL administers 3 IRs: Enlighten, the Glasgow Theses Service and DSpace.



## Encouraging Findings

- Strong links between LMS and HR systems for pulling in patron records info (including barcodes for authentication).
- Use of Institutional IDs within LMSs and IRs for author names.
- Links being established with Current Research Information Systems for metadata sharing.
- SFX link resolvers used within IR to pull in holdings info from LMS electronic databases and journals.
- Shared workflows between LMS and IR staff, who all work within the library.
- Use of Resource Discovery Platforms.

## Issues and problems

- 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.



## Resource Discovery Platforms



## Only a partial solution

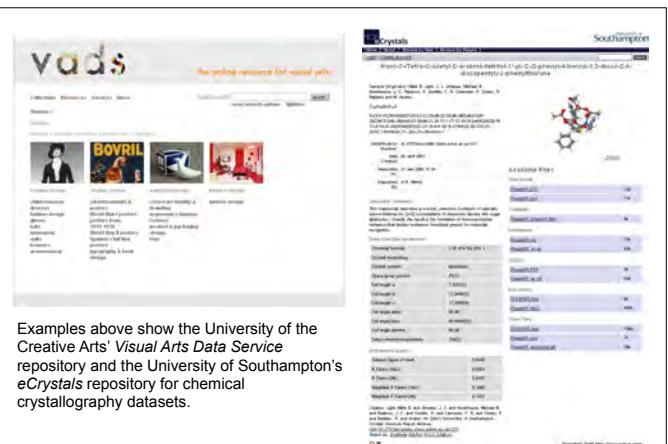


## Aspects of duplication

- 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.

## "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).



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.

## 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.

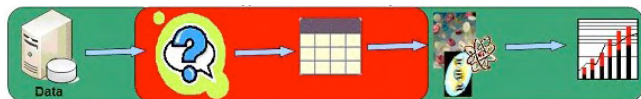
## Recommendations to Universities (cont...)

- 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.



## Thank-you!



### Contact Info:

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



## IDEALS at the University of Illinois: A Case Study of Integration Between an IR and Library Discovery Systems

Sarah L. Shreeves  
University of Illinois at Urbana-Champaign

NISO Webinar: From ILS to Repository and Back: Data Interoperability  
January 13, 2009

## 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

The screenshot shows the IDEALS website with a navigation menu on the left (Browse, My Account, Information) and a main content area. The main area includes a 'Welcome to IDEALS' message, a 'Top Downloads this Month' list, and a 'Recent Additions' list with details on dates and authors.

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

# Scholarship and Research

The screenshot shows a library website with various search results and a large orange text box overlaid on the center. The text box contains the question: 'What of this content would traditionally appear in library discovery systems? What SHOULD appear in library discovery systems?' Below the text box, it says 'Also audio, video, software, undergraduate research'.

# Scholarship and Research

The screenshot shows a library website with various search results and a large orange text box overlaid on the center. The text box contains the question: 'What of this content would traditionally appear in library discovery systems? What SHOULD appear in library discovery systems?' Below the text box, it says 'Also audio, video, software'.

# 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>

Label in Deposit Interface	Metadata Element	Depositor or Software Supplied	Recommendation for Use	Repeatable?	Free Text/Controlled	Display Label
Title	dc: title	Depositor Supplied	Required: The main title of the resource. An additional title may also be supplied (for example, a translation)	Yes, depositor may supply alternate title	Free text	Title
Date Published / Disseminated	dc: date issued	Depositor supplied	Required: Date of publication or first public distribution	No	Controlled	Issue Date
Not available	dc: date accessioned	Software supplied	Required: Date resource was first deposited into IDEALS	No	Controlled	Not Displayed
Not available	dc: date available	Software supplied	Required: Date resource was made available in IDEALS. This may be different from date accessioned if resource went through workflow process	No	Controlled	Date Available in IDEALS
Subject	dc: subject	Depositor supplied	Required: At least one subject term is required	Yes	Free text (has auto-complete)	Subject
Not available	dc: identifier uri	Software supplied	Required	Yes: Depositor may point to additional URLs in the identifier field if desired	Controlled	URI
Not available	dc: description provenance	Software supplied	Required	Yes	System generated text that may be altered if needed	Not Displayed
Author or Creator	dc: creator	Depositor supplied	Highly recommended for resources with an author or creator	Yes	Free Text (has auto-complete)	Author(s)
Genre of Resource	dc: type genre	Depositor supplied	Highly recommended for all resources	Yes	Controlled	Genre
Primary Language of Resource	dc: language	Depositor supplied	Highly recommended for all textual resources	No	Controlled	Language
Abstract or Summary	dc: description abstract	Depositor supplied	Highly recommended for all resources	No	Free text	Abstract
Additional	dc: contributor	Depositor	Recommended for all resources	Yes	Free text (has auto-complete)	Contributor(s)

## Wide Range of Metadata Quality

**Title:** Colin Mochrie vs. Jesus H. Christ: Messages About Masculinities and Fame in Online Video Conversations

**Authors:** Kendall, Lori

**Keywords:** video  
masculinities  
persistent conversation  
animation

**Issue Date:** 2007-Jan-7

**Publisher:** IEEE

**Citation Information:** hicc, p. 76b, 40th Annual Hawaii International Conference on System Sciences (HICSS'07), 2007

**Abstract / Summary:** With the availability of relatively easy-to-use tools for online video creation and distribution, people are increasingly producing videos not just for artistic expression, but also as a form of communication. This article examines two types of online video conversations, one within a specific online subculture, and one a series of responses to one particularly well-known video. Videos lend themselves to the expression of ambiguity, contradictory meanings, taboo topics, and emotions. Within these two video conversations, topics include discussions of masculinity, identity, and fame.

**URI:** <http://hdl.handle.net/2142/705>

**Type of Resource:** text

**Genre of Resource:** article

**Publication Status:** published or submitted for publication

**Peer-Reviewed:** is peer reviewed

**Title:** 3D scene modeling for Distributed Video Coding

**Authors:** Maitre, Matthieu  
Guillemot, Christine  
Morin, Luce

**Keywords:** Video coding  
Stereo vision

**Issue Date:** 2006-Oct

**Publisher:** IEEE

**Citation Information:** International Conference on Image Processing (ICIP)

**Abstract / Summary:** The compression efficiency of Distributed Video-Coding (DVC) suffers from the necessity of transmitting a large number of key-frames which are intra-coded. This paper describes a new 3D model-basedDVC approach which reduces the keyframe frequency. The decoder first recovers a 3D model from the key-frames. It then predicts the intermediate frames by projecting it onto 2D image planes and applying image-based rendering techniques. This paper also introduces a new quasi-DVC method relying on a limited point tracking at the encoder. It greatly improves the prediction PSNR, while only slightly increasing the encoder complexity. It also allows the encoder to adaptively select the key-frames based on the video motion-content.

**URI:** <http://hdl.handle.net/2142/159>

**Type of Resource:** text

**Genre of Resource:** conference paper

**Publication Status:** published or submitted for publication

**Peer-Reviewed:** is peer reviewed

## Interactions/Integration with Library Discovery Systems

### 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

#### Economic and fiscal impact analysis : a primer

Bookmark or cite this item: <http://hdl.handle.net/2142/9627>

File	Description	Format
<a href="#">economicscalm00camp.pdf (size)</a>	[no description provided]	PDF

**Title:** Economic and fiscal impact analysis : a primer

**Author(s):** Campbell, Harrison S.; Colfield, John Brevard, 1953.

**Contributor(s):** University of Illinois at Urbana-Champaign. Agricultural Experiment Station.

**Subject(s):** Urban impact analysis

**Geographic Coverage:** Illinois

**Issue Date:** 1992

**Publisher:** [Urbana, Ill.] : Agricultural Experiment Station, College of Agriculture, University of Illinois at Urbana-Champaign.

**Series/Report:** University of Illinois bulletin. 807

**Type:** Text

**Language:** English

**URI:** [http://hdl.handle.net/10111/14UCUCA\\_Archive/economicscalm00camp](http://hdl.handle.net/10111/14UCUCA_Archive/economicscalm00camp)  
<http://hdl.handle.net/2142/9627>

**Rights Information:** Copyright 1992 Board of Trustees University of Illinois

**Date Available in IDEALS:** 2008-05-29

**Has Version(s):** [http://hdl.handle.net/10111/14UCUCA\\_economicscalm00camp](http://hdl.handle.net/10111/14UCUCA_economicscalm00camp)  
<http://www.archive.org/details/economicscalm00camp>

**Identifier in Online Catalog:** 3497695

**OCLC Identifier:** (OCoLC)ocm28276751



## 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

The screenshot shows the IDEALS interface. At the top, there's a blue header with 'Title' and 'Print or Microform Coverage'. Below it, a yellow box highlights 'Illinois Classical Studies' with a sub-link '1976-Present in IDEALS (Illinois Digital Environment for Access to Learning and Scholarship)(with a five volume delay)'. Below that, another yellow box highlights 'Library trends' with a sub-link '1962 Present in IDEALS (Illinois Digital Environment for Access to Learning and Scholarship?)'. A red box highlights the '1962 Present in IDEALS' link.

But what about discovery of all the preprints, data sets, presentations, software, etc?

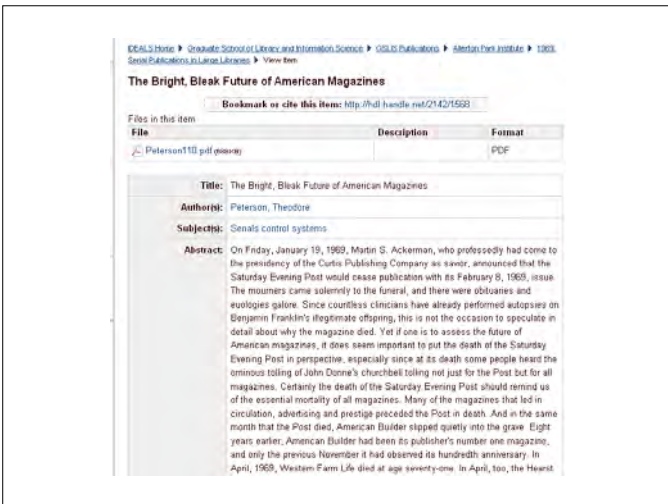
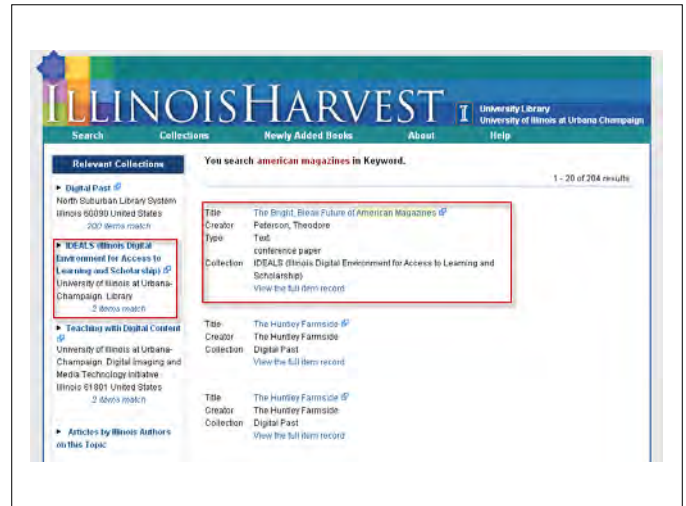
The screenshot shows the University of Illinois at Urbana-Champaign library search page. The search bar contains 'mobile air conditioning' and 'Multi-Subject Resources'. The search results show '24 ARTICLE MATCHES', '20 ARTICLE MATCHES', '229 ARTICLE MATCHES', and '38 ARTICLE MATCHES'. There are also 'QUICK LINKS' for various library services.

The screenshot shows the 'Easy Search Results' page for the search term 'mobile air conditioning'. It lists various search engines and their results:

- Multi-Topic -- Journal and Magazine Articles:
  - Academic Search Premier (EBSCO): 58 ARTICLE MATCHES
  - Academic Onefile (InfoTrac): 20 ARTICLE MATCHES
  - Scopus: 229 ARTICLE MATCHES
  - Web of Science (Social Sci. Sciences, Medicine, Humanities, Engineering): 38 ARTICLE MATCHES
- Books, Music, Documents, Media in UIUC & Illinois Libraries and Full-Text Books:
  - University of Illinois Library Catalog: 24 BOOK MATCHES
  - i-Share Library Catalog [state-wide Illinois, academic]: 28 BOOK MATCHES
  - Springer E-Books: 778 CHAPTERS No Titles: 20 PROTOCOLS
  - Google Books: 7200 E-BOOK MATCHES
  - IDEALS - UIUC Reports, Papers: 124 MATCHES
- Web Search Engines:
  - Google: 23800000 PAGE MATCHES
  - Bing: 7040000 PAGE MATCHES
  - Google Scholar: 51400 PAGE MATCHES
  - Scirus Academic Web Search: 143579 PAGE MATCHES

The screenshot shows the IDEALS search results for the search term 'mobile air conditioning'. It displays the search scope, conjunction, search type, and search for fields. The search results show 124 results, with the first few items listed:

- Three-Zone Modelling of a Mobile Air Conditioning Condenser: Karamlik, M.J. (Air Conditioning and Refrigeration Center, College of Engineering, University of Illinois at Urbana-Champaign, 1991-04). PDF (448)
- Dynamic Modeling of a Mobile Air Conditioning System From Actual Performance Data: Schenk, C.R. (Air Conditioning and Refrigeration Center, College of Engineering, University of Illinois at Urbana-Champaign, 1992-01). PDF (418)
- Refrigerant Charge Loss Detection for a Mobile Air Conditioning System: Johnston, C.E.; Miller, N.R.; Dunn, W.E. (Air Conditioning and Refrigeration Center, College of Engineering, University of Illinois at Urbana-Champaign, 1997-07). PDF (888)
- Improved R134A Mobile Air Conditioning Systems



## 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?

## Contact information

Sarah L. Shreeves  
 IDEALS Coordinator  
 Scholarly Commons Coordinator  
 University of Illinois at Urbana-Champaign  
[sshreeve@illinois.edu](mailto:sshreeve@illinois.edu)  
 217-244-3877

This work is licensed under a  
 Creative Commons Attribution 3.0 United States License.

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](http://www.niso.org/news/events/2010/datainterop)