11-1-2008

Institutional Repositories and Scholarly Communication

Thomas L. Walker II
Marshall University, walkert@marshall.edu

Follow this and additional works at: http://mds.marshall.edu/lib_faculty
Part of the Education Commons, and the Library and Information Science Commons

Recommended Citation
http://mds.marshall.edu/lib_faculty/20

This Presentation is brought to you for free and open access by the Libraries at Marshall Digital Scholar. It has been accepted for inclusion in Libraries Faculty Research by an authorized administrator of Marshall Digital Scholar. For more information, please contact zhangj@marshall.edu.
Institutional Repositories and Scholarly Communication

New Mexico State University
Thom Walker
December 12, 2008
"It is worth remembering (...) that communication changes have always occurred: it is simply that current changes are particularly rapid and radical".

Meadows, J. (1997, June)
Clifford A. Lynch

“A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.”

Lynch, C. A. (2003, February)
What is an Institutional Repository?

Institutional repositories—digital collections that capture and preserve the intellectual output of university communities—respond to two strategic issues facing academic institutions:

- they provide a central component in reforming scholarly communication by stimulating innovation in a disaggregated publishing structure; and
- they serve as tangible indicators of an institution’s quality, thus increasing its visibility, prestige, and public value
- have the ability to change disciplinary paradigms
In the early days of the open archive movement the following equation was true for many sites:

Archive → Journal
Article → Open Access to its full text → Self-deposit by the author
Now…

- Directory of Open Access Repositories (OpenDOAR) uses the following categories for repositories:
  - aggregating
  - disciplinary
  - governmental
  - institutional

(OpenDOAR 2007: http://www.opendoar.org/find.php)*

*terrific resource for locating repositories...however, they must be registered with OpenDOAR; some may not be included, i.e. Digital Library of Appalachia
Aggregating

- Archives aggregating data from several subsidiary repositories:
  
  The National Agricultural Library
  http://www.nal.usda.gov/

  Dialnet
  http://dialnet.unirioja.es/

  New England Law Library Consortium
  http://lsr.nellco.org/

  Global Research Archive
  http://gra.pronetos.com/
Disciplinary

- Cross-institutional subject repositories

  - Digital Library of Appalachia
    - http://www.aca-dla.org/
  - arXiv
    - http://www.arxiv.org/
  - ArtDOK
    - http://archiv.ub.uni-heidelberg.de/artdok/
Governmental

- Repositories for governmental data

  Kansas State Government Publications Repository
  http://www.kspace.org/

  Digital Library of National Parliamentary Library of Georgia
  http://www.nplg.gov.ge/dlibrary/
Institutional

- Institutional or departmental repositories

  Archive of Popular American Music
  http://digital.library.ucla.edu/apam/

  The Chopin collection at the University of Chicago
  http://chopin.lib.uchicago.edu/

  DSpaceUNM
  https://repository.unm.edu/dspace/

  New Mexico Tech
  http://dspace.nmt.edu/dspace/
Undetermined

- Repositories whose type has not yet been assessed
- Relatively new terminology
- Not accessible via OpenDOAR
OpenDOAR

- OpenDOAR is an authoritative directory of academic open access repositories. Each OpenDOAR repository has been visited by project staff to check the information that is recorded here. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories.

- The following group of charts are based on OpenDOAR statistics.
Institutional repositories (IR) make up the largest percentage of all repository types.
The largest percentage of repositories are in Europe, with North America second.
Journal Articles Make 60% of repository holdings, followed by theses and dissertations.

<table>
<thead>
<tr>
<th>Content Type</th>
<th>Percentage of Repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal articles</td>
<td>786 = 60%</td>
</tr>
<tr>
<td>Theses and dissertations</td>
<td>644 = 49%</td>
</tr>
<tr>
<td>Unpublished reports and working papers</td>
<td>568 = 43%</td>
</tr>
<tr>
<td>Conference and workshop papers</td>
<td>464 = 35%</td>
</tr>
<tr>
<td>Books, chapters and sections</td>
<td>403 = 31%</td>
</tr>
<tr>
<td>Multimedia and audio-visual materials</td>
<td>307 = 23%</td>
</tr>
<tr>
<td>Other special item types</td>
<td>214 = 16%</td>
</tr>
<tr>
<td>Bibliographic references</td>
<td>190 = 14%</td>
</tr>
<tr>
<td>Learning Objects</td>
<td>190 = 14%</td>
</tr>
<tr>
<td>Datasets</td>
<td>68 = 5%</td>
</tr>
<tr>
<td>Software</td>
<td>27 = 2%</td>
</tr>
<tr>
<td>Patents</td>
<td>23 = 1%</td>
</tr>
</tbody>
</table>

Total = 1292 repositories
56% of Repositories are Multidisciplinary

<table>
<thead>
<tr>
<th>Subjects in OpenDOAR Worldwide</th>
<th>Percentage of Repositories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multidisciplinary</td>
<td>728 = 56%</td>
</tr>
<tr>
<td>Science General</td>
<td>83 = 6%</td>
</tr>
<tr>
<td>Agriculture, Food and Veterinary</td>
<td>24 = 1%</td>
</tr>
<tr>
<td>Biology and Biochemistry</td>
<td>52 = 4%</td>
</tr>
<tr>
<td>Chemistry and Chemical Technology</td>
<td>32 = 2%</td>
</tr>
<tr>
<td>Earth and Planetary Sciences</td>
<td>27 = 2%</td>
</tr>
<tr>
<td>Ecology and Environment</td>
<td>43 = 3%</td>
</tr>
<tr>
<td>Mathematics and Statistics</td>
<td>41 = 3%</td>
</tr>
<tr>
<td>Physics and Astronomy</td>
<td>45 = 3%</td>
</tr>
<tr>
<td>Health and Medicine</td>
<td>93 = 7%</td>
</tr>
<tr>
<td>Technology General</td>
<td>77 = 5%</td>
</tr>
<tr>
<td>Architecture</td>
<td>12 = 0.9%</td>
</tr>
<tr>
<td>Civil Engineering</td>
<td>11 = 0.8%</td>
</tr>
<tr>
<td>Computers and IT</td>
<td>81 = 6%</td>
</tr>
<tr>
<td>Electrical and Electronic Engineering</td>
<td>21 = 1%</td>
</tr>
<tr>
<td>Mechanical Engineering and Materials</td>
<td>29 = 2%</td>
</tr>
<tr>
<td>Arts and Humanities General</td>
<td>34 = 2%</td>
</tr>
<tr>
<td>Fine and Performing Arts</td>
<td>41 = 3%</td>
</tr>
<tr>
<td>Geography and Regional Studies</td>
<td>62 = 4%</td>
</tr>
<tr>
<td>History and Archaeology</td>
<td>94 = 7%</td>
</tr>
<tr>
<td>Language and Literature</td>
<td>30 = 2%</td>
</tr>
<tr>
<td>Philosophy and Religion</td>
<td>31 = 2%</td>
</tr>
<tr>
<td>Social Sciences General</td>
<td>75 = 5%</td>
</tr>
<tr>
<td>Business and Economics</td>
<td>85 = 5%</td>
</tr>
<tr>
<td>Education</td>
<td>41 = 3%</td>
</tr>
<tr>
<td>Law and Politics</td>
<td>68 = 4%</td>
</tr>
<tr>
<td>Library and Information Science</td>
<td>52 = 4%</td>
</tr>
<tr>
<td>Management and Planning</td>
<td>24 = 1%</td>
</tr>
<tr>
<td>Psychology</td>
<td>28 = 1%</td>
</tr>
</tbody>
</table>

Total = 1292 repositories
Objectives of IR

The four main objectives for having an institutional repository are:

- to create global visibility for an institution's scholarly research
- to collect content in a single location
- to provide open access to institutional research output by self-archiving it
- to store and preserve other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature (e.g., theses or technical reports)
Quite Simply!

- IR’s can disseminate ideas and results to the wider community
  - USA accounts for 58% of STM literature*
  - Europe 26%
  - the rest of the world a mere 16%

* STM literature – Scientific, Technical and Medical literature
Institutional Repository Software

- the most popular platforms are:
  - Eprints
  - DSpace
  - Bepress
Eprints

- created in 2000
- developed at the University of Southampton School of Electronics and Computer Science
- result of first Open Archives meeting which developed *Santa Fe Conventions* which eventually became the OAI-PMH*
DSpace

- released in 2002
- developed by MIT and Hewlett-Packard
- open source software package that provides the tools for management of digital assets, and is commonly used as the basis for an institutional repository
- predominant repository software used by respondents*

*January 2006 Survey of ARL members
Bepress

- Berkeley Electronic Press
- founded by academics in 1999
- dedicated to producing products and services to aid scholarly communication
Breakdown of OpenDOAR Software Usage

Usage of Open Access Repository Software

Worldwide

- DSpace (362 = 28%)
- [Unknown] (297 = 23%)
- EPrints (242 = 19%)
- Bepress (51 = 4%)
- OPUS (51 = 4%)
- HTML (24 = 2%)
- Wildfire (23 = 2%)
- ETD-db (22 = 2%)
- [60 Others (220 = 17%)]

Total = 1292 repositories

OpenDOAR 05-Dec-2008
Examples of DSpace Repositories

Communities and Collections
Shown below is a list of communities and the collections and sub-communities within them. Click on a name to view that community or collection home page.

- **Atoll Research Bulletin**
  - Atoll Research Bulletin [515]

- **Center for Folklife and Cultural Heritage**
  - CFCH Staff Publications [1]

- **Museum Conservation Institute**
  - Museum Conservation Institute [4]

- **National Air and Space Museum**
  - NASM Staff Publications [236]

- **National Museum of Natural History**
  - Department of Botany [235]
  - Department of Entomology [172]
  - Department of Invertebrate Zoology [257]
  - Department of Paleobiology [93]
  - Department of Vertebrate Zoology [720]
  - Smithsonian Marine Station [142]
  - **Department of Anthropology**
    - AnthroNotes [17]
    - Department of Anthropology Staff Publications [27]
  - **Department of Mineral Sciences**
    - bulletin of the global Volcanism Network [65]
  - **Department of Vertebrate Zoology**
    - Smithsonian Herpetological Information Service [12]
What is Open Access (OA)?

- literature is...
  - digital
  - online
  - free of charge (typically)
  - and free of most copyright and licensing restrictions.
Defining OA

“open” from the architectural perspective – defining and promoting machine interfaces that facilitate the availability of content from a variety of providers. Openness does not mean “free” or “unlimited” access to the information repositories that conform to the OAI technical framework.
Types of OA

Green OA

- self-archiving, whereby an author places a copy of the scholarly output in one or more OA repositories (these may be an institutional repository (IR), a subject-based repository, or a combination of them). The same item may well also appear in a traditional journal (which may be print, parallel published or electronic only).
Types of OA

Gold OA

- is publishing the article in an OA journal, i.e. an electronic or parallel published journal that allows free of charge access to the articles within it.

- sometimes called “author pays” however, much contest has developed from this because the author rarely pays. It is usually the author’s employer or funder typically pays.
What is OAI?

- Originally conceived as the Santa Fe Conventions
- Initial concepts for harvesting metadata were discussed
- Evolved out of the scholarly communications field to provide interoperability among multiple information sources
- Roots of OAI lie in the Eprint community

The goal of OAI

- establishing a low-entry and well-defined interoperability framework applicable across domains
- one institution connected to another with the ability to allow access to holdings
- creating a cross-cultural, cross-age, diversified network for all to use regardless of predisposed conditioning
- The rich have access to the poor; the poor the rich, and everything in between.

UPS (Universal Preprint Service)

- held in October 1999
- adopted metadata harvesting
- the Santa Fe meeting was sponsored by the Council on Library and Information Resources (CLIR), the Digital Library Federation (DLF), the Scholarly Publishing & Academic Resources Coalition (SPARC), the Association of Research Libraries (ARL) and the Los Alamos National Laboratory (LANL).
What is Metadata Harvesting?

- in order to define Metadata Harvesting, it is necessary to define each part individually in order to create a complete definition
Metadata is...

- Metadata-Structured information about resources (including both digital and non-digital resources). Metadata can be used to help support a wide range of operations on those resources. In the context of services based on metadata harvested via OAI-PMH, the most common operation is discovery and retrieval of resources.
Harvesting is...

- Harvesting - In the OAI context, harvesting refers specifically to the gathering together of metadata from a number of distributed repositories into a combined data store.
So, Metadata Harvesting is...

- gathering together of structured information about resources (including both digital and non-digital resources) from a number of distributed repositories into a combined data store.
IRs and Open Archive

- allow for the dissemination of information among the scholarly community, much more rapidly than traditional print publications.

Scholarly Communication and IR’s

"something extraordinary...in the continuing networked information revolution...if they succeed they will permanently change the landscape of scholarly communication"

As a scholarly communication mode, institutional repositories represent a particularly interesting case to examine as they incorporate both normative and ideological agendas and illustrate how technical products embody social goals and power relationships.
Reasoning behind OA

- the economic model of scholarly publishing has been severely strained by rapidly rising subscription prices and relatively stagnant research library budgets

Lagoze, Carl and Herbert Van de Sompel. 2001
Comprehensive access to the expanding volume of scholarly materials necessary for research and teaching is at risk. Trends in scholarly publishing, especially commercial publisher business models, limit the ability to maintain the breadth and depth of library collections and reduce exposure to and impact of scholars’ work.”
among the most outspoken advocate of change, views such solutions as the first step in radical transformation of scholarly publishing whereby authors reclaim control over their intellectual property and the publishing process

OAI Record Example

```xml
<record xmlns="http://www.openarchives.org/OAI/1.1/OAI_ListRecords">
  <header>
    <identifier>oai:spurlock:cultural:59738</identifier>
    <datestamp>2001-11-15</datestamp>
  </header>
  <metadata>
    <dc:identifier xsi:schemaLocation="http://purl.org/dc/elements/1.1/
http://www.openarchives.org/OAI/1.1/dc.xsd"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
      <type>cultural</type>
      <type>physical object</type>
      <type>original</type>
      <title>Cloth Fragment</title>
      <subject>Unclassifiable Artifacts : Artifact Remnant : Cloth Fragment</subject>
      <format>24.6 x 16.3 x 1 cm (14 g)</format>
      <coverage>German</coverage>
      <coverage>Europe, West</coverage>
      <coverage>Germany</coverage>
      <date>Early 20th c. CE</date>
      <description>Materials: Textile--Linen</description>
      <description>Manufacturing Process: Weaving--Machine, Printing</description>
      <description>Munsell Color ID: Dark Greenish Blue (7.5B 3/4) -Background. Yellowish Gray (5Y 8.5/2) -Bird.</description>
      <description>Rectangular white linen fragment with floral and animal designs on printed turquoise background.</description>
    </dc:identifier>
    <rights>Spurlock Museum, University of Illinois at Urbana-Champaign</rights>
    <publisher>Spurlock Museum, University of Illinois at Urbana-Champaign</publisher>
    <identifier>1925.07.0044</identifier>
    <identifier>http://www.spurlock.uic.edu/Lasso.cgi?id=artifacts&layout=details&filename=20number=1925.07.0044&search</identifier>
  </metadata>
</record>
```
Harry Potter and the prisoner of Azkaban / by J.K. Rowling; illustrations...

000 00842cam 22002178a 45e
001 770508
005 19991210141017.0
008 990902s1999 nyu c 000 1 eng
010 ___ |a 99023982
035 ___ |a (OCoLC)41266045
040 ___ |a DLC |c DLC |d ABF |d KLG
020 ___ |a 0439136350 (hc)
020 ___ |a 0439136369 (pb)
050 00 |a PZ7.R79835 |b Ham 1999
049 ___ |a KLGG
100 1_ |a Rowling, J. K.
245 10 |a Harry Potter and the prisoner of Azkaban / |c by J.K. Rowling; illustrations by Mary Grandpré.
300 ___ |a ix, 435 p. : |b ill. ; |c 24 cm.
500 ___ |a Sequel to: Harry Potter and the Chamber of Secrets.
520 ___ |a During his third year at Hogwarts School for Witchcraft and Wizardry, Harry Potter must confront the devious and dangerous wizard responsible for his parents' deaths.
<OAI-PMH xmlns:xsi="http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd"
        xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2008-12-09T18:28:25Z</responseDate>
    <request verb="GetRecord" identifier="oailcoa1.loc.gov.loc.gmd/g3791p.r002300" metadataPrefix="oai_dc">http://memory.loc.gov/cgi-bin/oai2_0</request>
    <GetRecord>
        <record>
            <header>
                <identifier>oailcoa1.loc.gov.loc.gmd/g3791p.r002300</identifier>
                <datestamp>2005-11-21T17:08:59Z</datestamp>
                <setSpec>gmd</setSpec>
            </header>
            <metadata>
                <oai_dc xmlns:xsi="http://www.openarchives.org/OAI/2.0/oai_dc/
                      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
                    <title>New railroad map of the state of Maryland, Delaware, and the District of Columbia. Compiled and drawn by Frank Arnold Gray.</title>
                    <creator>Gray, Frank Arnold</creator>
                    <subject>Railroads—Middle Atlantic States—Maps</subject>
                    <description>Shows drainage, canals, stations, cities and towns, counties, canals, roads completed, narrow gauge and proposed railroads with names of lines. Includes list of railroads.</description>
                    <description>Scale 1:633,600</description>
                    <description>LC Railroad maps, 230</description>
                    <description>Description derived from published bibliography</description>
                    <publisher>Philadelphia</publisher>
                    <date>1876</date>
                    <type>image</type>
                    <type>map</type>
                    <cartographic>
                        <identifier>http://hdl.loc.gov/loc.gmd/g3791p.r002300</identifier>
                    </cartographic>
                </oai_dc>
            </metadata>
        </record>
    </GetRecord>
</OAI-PMH>
<marc:datatag="245" ind1="1" ind2="0">
  <marc:subfield code="a">New railroad map of the state of Maryland, Delaware, and the District of Columbia. Compiled and drawn by Frank Arnold Gray.\</marc:subfield>
</marc:datatag>

<marc:datatag="260" ind1="1" ind2="0">
  <marc:subfield code="c">Philadelphia, \</marc:subfield>
  <marc:subfield code="d">1876, \</marc:subfield>
</marc:datatag>

<marc:datatag="300" ind1="1" ind2="0">
  <marc:subfield code="a">col. map</marc:subfield>
  <marc:subfield code="c">39 x 62 cm.</marc:subfield>
</marc:datatag>

<marc:datatag="500" ind1="1" ind2="0">
  <marc:subfield code="a">Scale 1:633,600.</marc:subfield>
</marc:datatag>

<marc:datatag="510" ind1="1" ind2="0">
  <marc:subfield code="a">LC Railroad maps.</marc:subfield>
  <marc:subfield code="c">230</marc:subfield>
</marc:datatag>

<marc:datatag="520" ind1="1" ind2="0">
  <marc:subfield code="a">Shows drainage, canals, stations, cities and towns, counties, canals, roads completed, narrow gauge and proposed railroads with names of lines. Includes list of railroads.</marc:subfield>
</marc:datatag>

<marc:datatag="530" ind1="1" ind2="0">
  <marc:subfield code="a">Available also through the Library of Congress Web site as a raster image.</marc:subfield>
</marc:datatag>
References


Suber, P. Open Access Overview (definition, introduction)