About things Open - a quick primer on openness

Nancy John
Digital Publishing Librarian
University of Illinois at Chicago
February 2007

Budapest Open Access Initiative

- BOAI
- 5 years old this month
- http://www.soros.org/openaccess/
- A statement supporting open access
- Advocacy:
  - Self-archiving
  - Open access journals
- UIC signature is in process; in the meantime, individuals can sign on to support

SPARC - Scholarly Publishing and Academic Resources Coalition

- Headquartered at ARL since 1998; SPARC Europe at Oxford since 2003
- Current director: Heather Joseph; SPARC Europe: David Prosser
- Membership organization
- http://www.arl.org/sparc/
- Working to support open access through free journals, more reasonably priced journals, developing tools and economic models
- UIC is a founding member

DOAJ and OpenDOAR

- Directory of Open Access Journals (http://www.doaj.org)
- Open Directory of Open Access Repositories (http://www.opendoar.org)
  - a comprehensive and authoritative list of institutional and subject-based repositories, as well as archives set up by funding agencies
  - a joint collaboration between the University of Nottingham in the UK and the University of Lund in Sweden
  - funded by the Open Society Institute (OSI), along with the JISC, CURL, and SPARC Europe.

Self-archiving

- Open Access self-archiving (Green Road)
- Open Access publishing (Gold Road)
- SHERPA: Securing a Hybrid Environment for Research Preservation and Access
  - Project Romeo (Rights MEtadata for Open archiving)(http://www.sherpa.ac.uk/romeo.php)
  - JULIET Research Funders Archiving Mandates and Guidelines (http://www.sherpa.ac.uk/juliet/index.php)

The Open Archives Initiative - an overview (with specific reference to the UIC Library)

Nancy John
Digital Publishing Librarian
University of Illinois at Chicago
February 2007

Why this presentation?

- Both the publishing systems (OJS and Indigo) we've implemented this year, and the conference management system we used for the FM conference, make use of this technology
- Staff expressed interest during the recent Indigo sessions, so this is to meet that interest
- Sessions/briefings on other topics (e.g. LOCKSS, digital migration, Dublin Core / Metadata, XML, EADs etc.) can be requested at any time (at least until September 2007) - nrj@uic.edu

What is OAI?

- "The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication" -- from www.openarchives.org

OAI governance

- Sponsors
  - CNI (Coalition for Networked Information)
  - Mellon Foundation
  - Digital Library Federation
  - National Science Foundation
- Executive
  - Carl Lagoze, Cornell
  - Herbert Van de Sompel, LANL

OAI is changing

- Originally the idea was to federate disparate repositories
- The technology though is quite powerful and is being adapted quickly to link and federate all kinds of data
- OAI is in a state of flux - it's a little unclear what it's really about because it's changing
- Currently there's lots of research and development activity going on about OAI

What's the fuss about?

- When the idea for creating databases of library digital objects, of university faculty intellectual output etc. first was proposed, the biggest obstacle was the lack of a single front door
- Who wants to do 10 or 100 or 1000 separate searches?
- Would the owners of the databases have to specialize and host the subject materials of others?
- Couldn't there be an OCLC for these files? A portal? Could Z39.50 work?
- But would people deposit the records? How could we rationalize varying standards of practice?

Too much stuff, too much work

- The spectre of 100000s of items
- The notion that the records would be good enough but not 'catalog perfect' (e.g. authority control issues)
- The nagging question: can't the computer do this work for us?
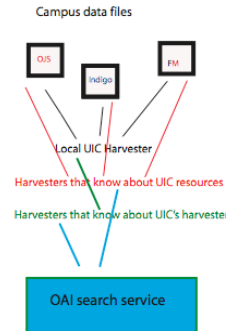
The answer:OAI

- Three parts
  - Databases that could output their cataloging in a standard format automatically, much like a Web server serves up a Web page
  - Software (called a harvester) that could ask a database for its output and then store it, much like a Web spider
  - A service that could build the retrieved and stored data into a user-friendly way of searching the content of 1000s of databases, much like Google or Yahoo!
- Furthermore, if the data output contained a Web link for each object, the services would provide one-click access to the objects

The three-part model

- To participate all one has to do is to be a data provider
- Some participants will make and run data harvesters
- Some participants will make services using the data

Harvesting

– Note: one can
   harvest from many
   sites locally and
   create a single OAI
   data provider for an
   entire campus, or
   one can have many
   data providers on a
   campus

Campus data files

OJS Indigo FM

Local UIC Harvester

Harvesters that know about UIC resources

Harvesters that know about UIC's harvester

OAI search service

So how does this work?

- There is a Protocol for Metadata Harvesting
  (OAI-PMH)
- There are two classes of participants in the
  OAI-PMH framework:
  - Data Providers administer systems that support
    the OAI-PMH as a means of exposing metadata
  - Service Providers use metadata harvested via the
    OAI-PMH as a basis for building value-added
    services.

Duties of Data Providers

- Host a repository that is network accessible and can process the 6 OAI-PMH requests according to the protocol
- In response to a request from a harvester, the repository provides standard xml that contains the metadata
  - Metadata can exist pre-formatted or be generated on-the-fly by the data provider
  - The metadata must be a 'known' format - e.g. unqualified Dublin Core, METS
- Register at openarchives.org and pass their conformity test
- http://www.openarchives.org/Register/BrowseSites

What's a record look like?

- See the last page of the handout

And the harvester …

- Can ask for all records
- Records that match a certain variable:
  - New (creation datestamp after mm/dd/yyyy)
  - Updates/modifications (creation before but update datestamp after)
  - Subject: math
- Uses http:
  - http://arXiv.org/oai2? verb=GetRecord&identifier=oai:arXiv.org:cs/011201 7&metadataPrefix=oai_dc

The Service

- Really, this is what this is all about
- Many services but we'll only look at three
- The list of registered providers:
  http://www.openarchives.org/service/listproviders.html

Services

- OAIster
  http://oaister.umdl.umich.edu/o/oaister/index.html
- D-LIST http://dlist.sir.arizona.edu/
- OpenDOAR: http://www.opendoar.org

OAI @ UIC

- Open Journal System:
  - Behavior and Social Issues
  - First Monday
- Indigo
- EADs
- Carberry Collection

```xml
<header>
  <identifier>oai:arXiv:cs/0112017</identifier>
  <datestamp>2002-02-28</datestamp>
  <setSpec>cs</setSpec>
  <setSpec>math</setSpec>
</header>
<metadata>
 <oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
   <dc:title>Using Structural Metadata to Localize Experience of Digital
            Content</dc:title>
   <dc:creator>Dushay, Naomi</dc:creator>
   <dc:subject>Digital Libraries</dc:subject>
   <dc:description>With the increasing technical sophistication of both
    information consumers and providers, there is increasing demand for
    more meaningful experiences of digital information. We present a
    framework that separates digital object experience, or rendering,
    from digital object storage and manipulation, so the
    rendering can be tailored to particular communities of users.
   </dc:description>
   <dc:description>Comment: 23 pages including 2 appendices,
                8 figures</dc:description>
   <dc:date>2001-12-14</dc:date>
   <dc:type>e-print</dc:type>
   <dc:identifier>http://arXiv.org/abs/cs/0112017</dc:identifier>
 </oai_dc:dc>
</metadata>
<about>
  <provenance
     xmlns="http://www.openarchives.org/OAI/2.0/provenance"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/provenance
     http://www.openarchives.org/OAI/2.0/provenance.xsd">
    <originDescription harvestDate="2002-02-02T14:10:02Z" altered="true">
      <baseURL>http://the.oa.org</baseURL>
      <identifier>oai:r2:klik001</identifier>
      <datestamp>2002-01-01</datestamp>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </originDescription>
  </provenance>
</about>
```