OAI and ODL
Building Digital Libraries from Components

Ryan Richardson <ryanr@vt.edu>
Virginia Tech DLRL
18 September 2003
Outline

1. Introduction to OAI
2. Definitions and Concepts
3. OAI Protocol for Metadata Harvesting
4. Introduction to ODL
5. OAI and ODL Components
1. Introduction to OAI

• What is the Open Archives Initiative?
  – Group of people and organizations dedicated to solving problems of digital library interoperability by developing simple protocols.

• Major Accomplishment:
  – Protocol for Metadata Harvesting (OAI-PMH)
1.1. What is the OAI-PMH?

- What is the Protocol for Metadata Harvesting?
  - Network protocol to transfer metadata from one archive to another
    - Any metadata (XML-encoded data records)
    - In a continuous stream
    - As simply as possible
1.2. General System Strategy

- Services
- Metadata Harvesting
- Document Model
1.3. Case Study: AmericanSouth

- Digital library of resources related to Southern history and culture
- Multiple independent university-based collections of electronic documents

Diagram:

```
Emory       OAI Protocol for Metadata Harvesting   American South.Org portal
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UTK</td>
<td></td>
</tr>
<tr>
<td>Virginia Tech</td>
<td></td>
</tr>
</tbody>
</table>
```

OAI & ODL - CS6604
1.4. Versions of OAI-PMH

• v1.0 January 2001
• v1.1 July 2001
  – Minor revision from v1.0
• v2.0 June 2002
  – Mostly syntactical changes
  – These notes are based on version 2.0!
2. Definitions / Concepts

• Basic Principles
  – What is an Open Archive?
  – Harvesting vs. Federation
  – Data and Service Providers

• Underlying Technology
  – HTTP and XML

• Protocol Policies
  – What is a record?
  – Multiplicity of Metadata
  – Sets
  – Datestamp, Harvesting and Flow Control
2.1. What is an Open Archive?

- Any WWW-based system that can be accessed through the well-defined interface of the Open Archives Protocol for Metadata Harvesting
- ... aka OAI-Compliant Repository
- No implications for:
  - Physical storage of data
  - Cost of data
  - Metadata and data formats
  - Access control to server
2.2. Harvesting vs Federation

- Competing approaches to interoperability
  - Federation is when services are run remotely on remote data (e.g. Meta-searching)
  - Harvesting is when data/metadata is transferred from the remote source to the destination where the services are located (e.g. Union catalogues)

- Federation requires more effort at each remote source but is easier for the local system and vice versa for harvesting

- OAI currently focuses on harvesting
2.3. Data and Service Providers

• Data Providers refer to entities who possess data/metadata and are willing to share this with others (internally or externally) via well-defined OAI protocols (e.g. database servers)

• Service Providers are entities who harvest data from Data Providers in order to provide higher-level services to users (e.g. search engines)

• In networking terms, the data provider is a network server and a service provider connects to the server as a client.
2.4. HTTP and XML

• Protocol for Metadata Harvesting is an almost stateless request/response protocol
• Requests and responses are sent via the HTTP protocol
• Requests are encoded as GET/POST operations
• Responses are well-formed XML documents
2.5. What is a record?

- A record refers to an independent XML structure that may be associated with digital or physical objects.
- Records are usually associated with metadata, not data.
- OAI advocates harvesting of records, which contain metadata and additional fields to support the harvesting operation.
# 2.6. As Compared to Z39.50

<table>
<thead>
<tr>
<th></th>
<th>Z39.50</th>
<th>OAI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content (Objects)</strong></td>
<td>Distributed</td>
<td>Distributed</td>
</tr>
<tr>
<td><strong>World View</strong></td>
<td>Bibliographic</td>
<td>Bibliographic</td>
</tr>
<tr>
<td><strong>Object Presentation</strong></td>
<td>Data provider</td>
<td>Data provider</td>
</tr>
<tr>
<td><strong>Searching is</strong></td>
<td>Distributed</td>
<td>Centralized</td>
</tr>
<tr>
<td><strong>Search done by</strong></td>
<td>Data provider</td>
<td>Service provider</td>
</tr>
<tr>
<td><strong>Metadata searched is</strong></td>
<td>Up to date</td>
<td>Stale</td>
</tr>
<tr>
<td><strong>Semantic Mapping</strong></td>
<td>When searching</td>
<td>Metadata delivery</td>
</tr>
</tbody>
</table>
2.7. What OAI Is Not

- Not search
- Not database
- Not metadata
- Not OAIS
2.8. What OAI is good for

- Where content is widely distributed, in different kinds of non-Z39.50 enabled locations
  - Metadata provider more lightweight than Z39.50
  - Metadata provider scales well
    Service provider scales according to search capability
- Metadata is sufficient for services desired
- Normalization, de-duping, augmentation desired

Not mutually exclusive
- Portals can use both Z39.50 & OAI
2.9. Sample OAI Record

<record>
  <header>
    <identifier>oai:sigir:ws3</identifier>
    <datestamp>2001-08-13</datestamp>
  </header>
  <metadata>
    <dc>
      <title>OAI Workshop at SIGIR</title>
      <creator>Hussein Suleman</creator>
      <language>English</language>
    </dc>
  </metadata>
  <about>
    <metadataID>oai:sigir:ws3md</metadataID>
  </about>
</record>
2.10. Multiplicity of Metadata

• Multiple formats of metadata allowed
• Dublin Core is mandatory
• Any other format allowed as long as it has an XML encoding
• E.g. MARC (Libraries), IMS (Education), ETDMS (Theses/Dissertations), RFC1807 (Bibliographies)
2.11. Sets

- Protocol mechanism to allow for harvesting of sub-collections
- No well-defined semantics – depends completely on local data providers
- May be defined by arrangement between data providers and service providers
- E.g. Subject areas, years, author names, search queries
2.12. Datestamps & Harvesting

• Each record needs a datestamp that indicates its date of creation or modification

• Dates are used to allow for harvesting by date range, thus allowing incremental and continuous transfer of metadata from a data provider to a service provider
2.13. Flow Control

- HTTP “retry-after” mechanism can be leveraged to support server-side delaying of a client’s request.
- Resumption Tokens can be used to return partial results – the client is issued with a token which may be presented to the server to receive more results.
2.14. How OAI Works

OAI “VERBS”

Identify
ListSets
ListMetadataFormats
ListIdentifiers
GetRecord
ListRecords

Service Provider                Metadata Provider

HTTP Request
(OAI Verb)

HTTP Response
(Valid XML)
2.15. The baseURL

• Requests are sent by HTTP to baseURLs, with parameters appended, e.g.
  – http://www.test.org/oai.pl?verb=Identify

• Responses are the documents that are returned by the server

• The baseURL is the point of contact to communicate with a component!
3. Protocol for Metadata Harvesting

- Service Requests
  - Identify
  - ListSets
  - ListMetadataFormats
  - ListIdentifiers
  - GetRecord
  - ListRecords

- Metadata Multiplicity

- Date Ranges

- Resumption Tokens
3.1. Identify

• Purpose
  – Return general information about the archive and its policies

• Parameters
  – None

• Sample URL
  – http://www.anarchive.org/cgi-bin/OAI?verb=Identify
3.2. ListSets

- **Purpose**
  - Provide a hierarchical listing of sets in which records may be organized

- **Parameters**
  - None

- **Sample URL**
  - http://www.anarchive.org/cgi-bin/OAI?verb=ListSets
3.3. ListMetadataFormats

• **Purpose**
  – List metadata formats supported by the archive as well as their schema locations and namespaces

• **Parameters**
  – identifier – for a specific record (O)

• **Sample URL**
  – http://www.anarchive.org/cgi-bin/OAI?
    verb=ListMetadataFormats
3.4. ListIdentifiers

- **Purpose**
  - List headers for all items corresponding to the specified parameters

- **Parameters**
  - from – start date (O)
  - until – end date (O)
  - set – set to harvest from (O)
  - metadataPrefix – metadata format to list identifiers for (R)
  - resumptionToken – flow control mechanism (X)

- **Sample URL**
3.5. GetRecord

• Purpose
  – Returns the metadata for a single identifier in the form of an OAI record

• Parameters
  – identifier – unique id for record (R)
  – metadataPrefix – metadata format (R)

• Sample URL
3.6. ListRecords

- **Purpose**
  - Retrieves metadata for multiple records

- **Parameters**
  - from – start date (O)
  - until – end date (O)
  - set – set to harvest from (O)
  - resumptionToken – flow control mechanism (X)
  - metadataPrefix – metadata format (R)

- **Sample URL**
3.7. Protocol Details

- OAI Transaction == An OAI request (HTTP) & corresponding OAI response (XML)
  - Optional: use resumptionToken & other flow control mechanisms to manage service load

- Item Identifiers – Persistence & Uniqueness

- Item Datestamps – Date of last metadata change; supports selective harvesting
3.8. Examples of OAI Requests

http://www.language-archives.org/cgi-bin/olaca3.pl?verb=Identify

http://publications.uu.se/portal/OAI?verb=ListSets

http://www.language-archives.org/cgi-bin/olaca3.pl?
  verb=ListMetadataFormats

http://www.language-archives.org/cgi-bin/olaca3.pl?
  verb=ListIdentifiers&metadataPrefix=oai_dc&from=2002-12-01

http://www.language-archives.org/cgi-bin/olaca3.pl?
  verb=GetRecord&metadataPrefix=oai_dc&
  identifier=oai%3Aacl.sr.language-archives.org%3AA00-1006
3.9. An OAI Response

```xml
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns=… xmlns:xsi=… xsi:schemaLocation=…>
  <responseDate>2002-05-01T19:20:30Z</responseDate>
  <request verb="GetRecord"
    identifier="oai:arXiv:hep-th/9901001"
    metadataPrefix="oai_dc">
    http://an.oa.org/OAI-script</request>
  <GetRecord>
    <record>
      ...
    </record>
  </GetRecord>
</OAI-PMH>
```
3.10. An OAI Record

<header>
  <identifier>oai:arXiv:cs/0112017</identifier>
  <datestamp>2002-02-28</datestamp>
  <setSpec>cs</setSpec>
</header>

<metadata>
    <dc:title>Using Structural Metadata…</dc:title>
  </oai_dc:dc>
</metadata>

<about>
  <provenance xmlns="http://purl.org/net/mediawiki/provenance/"/>
  ....
</about>
3.11. Unique Identifiers

- Each item must have a unique identifier
- Identifiers must follow rules for valid URIs
- Example:
  - oai:<archiveId>:<recordId>
  - oai:etd.vt.edu:etd-1234567890
- Each identifier must resolve to a single item and always to the same item
  - Can’t reuse OAI item identifiers
3.12. Datestamps

• Needed for every OAI record to support incremental harvesting
• Must be updated when addition or modification or deletion made in order to ensure changes are correctly propagated to harvesters
• Different from dates within the metadata – OAI datestamp is used only for harvesting
• Can be either YYYY-MM-DD or YYYY-MM-DDThh:mm:ssZ (must be GMT timezone)
3.13. OAI Provider Architectures

**Descriptive Metadata**
- DBMS
- XML
- HTML

**OAI Administrative Metadata**
- DBMS

**OAI Application (CGI, ASP, PHP, etc.)**

**Webserver - HTTP**

**OAI Harvesters**
3.14. Repository Explorer

Open Archives Initiative - Repository Explorer

explorer version - 1.45a : protocol version - 1.0/1.1/2.0 : March 2003

This site presents an interface to interactively test archives for compliance with the OAI Protocol for Metadata Harvesting [Click here for details]

JavaScript is required

Note: To avoid HTTP errors, please wait for each page to finish loading before clicking on any link.

Please enter the URL to the OAI interface (everything before the ?) or choose a predefined archive from the table
Then click on a verb from the list below to test that function (entering parameters as necessary)

URL :

A Celebration of Women Writers
AIM25 - Archives in London
Alaska Native Language Center Archives
American Memory [LoC]

[ View Archive Website ][ Test and Add an archive to this list ]
# 3.15. RE Parameter Testing

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify</td>
<td>from (eg., YYYY-MM-DD):</td>
</tr>
<tr>
<td>List Metadata Formats</td>
<td>until (eg., YYYY-MM-DD):</td>
</tr>
<tr>
<td>List Sets</td>
<td>metadataPrefix:</td>
</tr>
<tr>
<td>List Identifiers</td>
<td>identifier:</td>
</tr>
<tr>
<td>List Records</td>
<td>set:</td>
</tr>
<tr>
<td>Get Record</td>
<td>resumptionToken:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language</th>
<th>Display</th>
<th>Schema Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Parsed</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Raw XML</td>
<td>Local mirror of schemata (Xerces)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Online schemata (Xerces)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Local mirror of schemata (XSV)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online schemata (XSV)</td>
</tr>
</tbody>
</table>

Send all comments to hussein@vt.edu --- Digital Library Research Laboratory@VirginiaTech

OAI & ODL - CS6604
3.16. RE Formatted View of Data

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>jcdl Picture Archive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base URL</td>
<td><a href="http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl">http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl</a></td>
</tr>
<tr>
<td>Protocol Version</td>
<td>2.0</td>
</tr>
<tr>
<td>Admin Email</td>
<td><a href="mailto:jcdlpix@rocky.dlib.vt.edu">jcdlpix@rocky.dlib.vt.edu</a></td>
</tr>
<tr>
<td>Earliest Datestamp</td>
<td>1970-01-01T00:00:00Z</td>
</tr>
<tr>
<td>Deleted Record Handling</td>
<td>no</td>
</tr>
<tr>
<td>Granularity</td>
<td>YYYY-MM-DD</td>
</tr>
</tbody>
</table>

Other Information:

description:

toolkit:

title: VTOAI Perl Data Provider
author:
    name: Hussein Suleman
    email: hussein@vt.edu
institution: Virginia Tech
version: 3.05
URL: http://www.dlib.vt.edu/projects/OAI/
3.17. RE Raw XML views of data

```xml
<?xml version="1.0" encoding="UTF-8"?>

<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <responseDate>2003-05-20T22:10:19Z</responseDate>
  <request verb="Identify">http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl</request>

  <Identify>
    <repositoryName>jcdl Picture Archive</repositoryName>
    <baseURL>http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>jcdlpix@rocky.dlib.vt.edu</adminEmail>
    <earliestDatestamp>1970-01-01T00:00:00Z</earliestDatestamp>
    <deletedRecord>no</deletedRecord>
    <granularity>YYYY-MM-DD</granularity>
    <description>
      <toolkit xmlns="http://oai.dlib.vt.edu/OAI/metadata/toolkit" xsi:schemaLocation="">
        <title>VTOAI Perl Data Provider</title>
        <author>
          <name>Hussein Suleman</name>
          <email>hussein@vt.edu</email>
          <institution>Virginia Tech</institution>
        </author>
        <version>3.05</version>
        <URL>http://www.dlib.vt.edu/projects/OAI/</URL>
      </toolkit>
    </description>
  </Identify>
</OAI-PMH>
```
3.18. RE Automatic Test Suite

Open Archives Initiative - Repository Explorer
explorer version - 1.45a : protocol version - 2.0 : March 2003

Open Archives Initiative :: Protocol for Metadata Harvesting v2.0
RE Protocol Tester 1.45a :: Virginia Tech DLR :: March 2003

(1) Testing : Identify
URL : http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl?verb=Id
Test Result : OK
----- [ Repository Name = jcdl Picture Archive ]
----- [ Protocol Version = 2.0 ]
----- [ Base URL = http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai ]
----- [ Admin Email = jcdlpix@rocky.dlib.vt.edu ]
----- [ Granularity = YYYY-MM-DD ]

(2) Testing : Identify (illegal_parameter)
URL : http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl?verb=Id
Test Result : OK

(3) Testing : ListMetadataFormats
URL : http://rocky.dlib.vt.edu/~jcdlpix/cgi-bin/OAI2.0/hspics/jcdl/oai.pl?verb=Lis
Test Result : OK
----- [ Only oai_dc supported ]
3.19. RE Error in XML

XML Schema Validation Error!

Errors in XML instance

```xml
<?xml version='1.0'?>
<oai xmlns='http://www.openarchives.org/OAI/1.0/OAI_Identifier'>
<import xmlns='http://www.openarchives.org/OAI/1.0/OAI_Identifier' namespace='http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTDErr1/VTETD.pl?verb=Identify'>
  <invalid char='4' code='cvc-complex-type.1.2.4' line='11' resource='file://tmp/...'
  <node id='1'>
    <edge dest='2' label='http://www.openarchives.org/OAI/1.0/OAI_Identifier':responseD...</node>
  <node id='2'>
    <edge dest='3' label='http://www.openarchives.org/OAI/1.0/OAI_Identifier':requestUR...</node>
  </node>
</oai>
```
4. Introduction to ODL

• Open Digital Libraries
  – Framework for componentized Digital Libraries
  – Design principles for components
  – Protocols for inter-component communications
  – Built upon OAI-PMH v1.1
4.1. Users and Objects

users

digital objects

OAI & ODL - CS6604
4.2. Digital Library

Monolithic and/or Custom-built web-based application

digital library
4.3. Componentized DL

componentized digital library
4.4. How about OAI-PMH?

• Metadata transfer among digital libraries “is almost =” metadata exchange among components

• Need a few changes to support inter-component communication, including:
  – Support for additional information in responses
  – Support for adding records as well (PutRecord)
4.5. Open Digital Library

open digital library

OAI & ODL - CS6604 49
Protocol for Metadata Harvesting

Extended OAI-PMH

Open Digital Library Protocol
Open Digital Library Component

Extended OPEN ARCHIVE

OPEN ARCHIVE
4.8. Open Digital Library

• Network of Extended Open Archives where each node acts as either a provider of data, services or both.

• Component = Node
• Protocol = Arc
4.9. Example Open Digital Library

Students and researchers

ETD Digital Library

ETD collections

OAI & ODL - CS6604
4.10. Prototype - FrontPage

Electronic Thesis/Dissertation
OAI Union Catalog

Some Recent Additions to our Collection


- The Ontology of Persistence, Love, Shanon, Virginia Polytechnic Institute and State University, 2001-06-25 [More Info]

- Virginia Save Our Streams (SCS): Volunteers' Motivations for Participation and Suggestions for Program Improvement, Haas, Steven Christopher, Virginia Polytechnic Institute and State University, 2000-08-03 [More Info]

Quick Search
Query: 

Quick Browse
Institution: All
Sort By: Default
Year: All

OAI & ODL - CS6604
4.11. Prototype - Search

Electronic Thesis/Dissertation
OAI Union Catalog

Search Results

1. Characterizing Web Response Time
   
   Liu, Baozhang M.S.
   
   Abstract: It is critical to understand WWW latency in order to design better HTTP protocols. In this study we characterize Web response time and examine the effects of proxy caching, network bandwidth, traffic load, persistent connections for a page, and periodicity. Based on studies with four workloads, we show that at least a quarter of the total elapsed t...

   Date: 1998-05-07

   [More Info] [Go To Document] [Find Similar Documents]

2. Chemical Interferences on the Atomization Yield of High Reduction Potential Elements - Signal Suppression in the Plasma Source Spectrometry
   
   Liao, Jian, jianieliu78@yahoo.com, 1960-11-04, Changzhou, CHINA
   
   Date: 2001-03-08

   [More Info] [Go To Document] [Find Similar Documents]
4.12. Prototype - Browse

Browse ETDs

<table>
<thead>
<tr>
<th>Institution</th>
<th>Sort By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aarhus University, Chemistry Department</td>
<td>All</td>
</tr>
<tr>
<td>California Institute of Technology</td>
<td>All</td>
</tr>
<tr>
<td>Caltech Library System</td>
<td>All</td>
</tr>
<tr>
<td>Forskningscenter Risoe</td>
<td>All</td>
</tr>
<tr>
<td>Gerhard-Mercator-Universität Duisburg</td>
<td>All</td>
</tr>
<tr>
<td>Humboldt-Universität zu Berlin</td>
<td>All</td>
</tr>
<tr>
<td>Massachusetts Institute of Technology</td>
<td>All</td>
</tr>
<tr>
<td>Other</td>
<td>All</td>
</tr>
<tr>
<td>Rice National Laboratory</td>
<td>All</td>
</tr>
<tr>
<td>Technische Universität Dresden</td>
<td>All</td>
</tr>
<tr>
<td>Universität Stuttgart</td>
<td>All</td>
</tr>
<tr>
<td>Virginia Tech</td>
<td>All</td>
</tr>
</tbody>
</table>

1. Test
   Bossert, Sven
   - **Abstract**: The solidification microstructure in wedge-shaped castings of Cu-Ni-Ti-Zr glass forming alloys is investigated, while the composition was systematically varied. Near the critical thickness for glass formation, a spatially inhomogeneous dispersion of nanocrystals is observed, where spherical regions contain a much higher density of nanocrystals than...
   - **Date**: 2001-07-13
   - [More Info] [Go To Document] [Find Similar Documents]
3. All-Optical Logic Circuits based on the Polarization Properties of Non-Degenerate Four-Wave Mixing

OAI & ODL - CS6604
4.13. ODL Component Requirements

- **Search**
  - Retrieve a list of items
  - Index new items

- **Annotate**
  - Add annotation to item
  - Retrieve a list of annotations for an item
4.14. Layer 1 : OAI PMH

- Protocol for Metadata Harvesting
  - Transfer stream of metadata from one archive or component to another
- Service Requests
  - Identify, ListSets, ListMetadataFormats
  - ListIdentifiers, GetRecord, ListRecords
4.15. Layer 2 : Extended OAI-PMH

• OAI-PMH + extensions for general-purpose inter-component communication
  – Added in generic containers in every response for additional information
  – Added “PutRecord” to submit a record
  – Increased granularity to support times as well as dates (same as OAI-PMH v2.0)
  – Ignored DC requirement
4.16. Layer 3 : ODL Protocols

- Specialized protocol semantics for different components, e.g.:
  - Search component uses ODLSearch protocol
    - ListRecords and ListIdentifiers embed query terms in “set” parameter
  - Annotation component uses ODLAnnotate protocol
    - ListRecords and ListIdentifiers specify the item for which annotations are requested in the “set” parameter
    - PutRecord adds an annotation to an item
4.17. Case Study: ETD ODL Prototype

- Electronic Thesis and Dissertation Open Digital Library

```
Virginia Tech
PhysNet
Humboldt
Duisburg
CalTech
Dresden
MIT Filter
MIT
```

User Interface

Browse
Search
Recent

Union Archive

User Interface

OAI/ODL component

OAI/ODL protocol
4.18. Ultimate Goal

• Package different configurations into instant DL systems
• DL building = component configuration
• All DLs speak the same language(s)
• Basic services are trivial to provide so more effort is spent on advanced capabilities of DLs
5. OAI and ODL components

• No one needs to start from scratch!
• OAI Components create OAI data providers from existing systems or collections
  – XMLFile, ETD-db extensions, etc.
• ODL Components implement basic digital library services and communicate using ODL and OAI protocols
  – Search, Browse, Annotate, etc.
5.1. Basic Model
5.2. Simple Searching

![Diagram showing the components of a simple searching system: OAI Data Provider, Search Engine Component, Search Engine WWW Interface, IRDB, IRDB user interface, OAI-PMH, and ODLSearch.]
5.3. Software to be installed

• XML-File
  – create Open Archive from collection of XML files

• Harvester
  – test harvesting of data from OAI archive

• IRDB
  – simple search engine

• IRDB user interface
5.4. Steps in building it

- Install XMLFile
  - Test XMLFile
- Install IRDB
  - Connect to XMLFile’s baseURL
  - Test IRDB
- Install user interface
  - Connect to IRDB’s baseURL
  - Test user interface
5.5. Testing: Repository Explorer

- The Repository Explorer is a tool for testing Open Archives.
- You can issue individual commands and validate the results (using XML Schema)
- You can also perform a sequence of automatic tests

- [http://purl.org/net/oai_explorer](http://purl.org/net/oai_explorer)
5.6. Wrap up and discussion

- We will build a simple digital library from components!

http://dlbox.nudl.org/docs/tutorial/odl_cc_instructions_ming.htm
6.1. Final Thoughts

• OAI-PMH is a simple protocol for exporting and importing metadata
• ODL Components based on OAI can be used to build modular systems
• Lots of tools available now!
• Lots of interest from other people already, even publishers!
Links

• Open Archives Initiative
  – http://www.openarchives.org

• OAI Metadata Harvesting Protocol
  – http://www.openarchives.org/OAI/openarchivesprotocol.htm

• Virginia Tech DLRL OAI Projects
  – http://www.dlib.vt.edu/projects/OAI/

• Repository Explorer
  – http://purl.org/net/oai_explorer

• CITIDEL
  – http://www.citidel.org/
More Links

• NDLTD
  – http://www.ndltd.org

• ARC Cross-Archive Search Service
  – http://arc.cs.odu.edu/

• XML Schema Validator
  – http://www.w3.org/2001/03/webdata/xsv

• Dublin Core Metadata Initiative
  – http://www.dublincore.org

• E-Prints DL-in-a-box
  – http://www.eprints.org

• XML Tools at W3C
  – http://www.w3.org/XML/#software
That’s All, Folks!

Questions?