

Introduction to the OAI Metadata Harvesting Protocol



*Hussein Suleman, hussein@vt.edu
Digital Library Research Laboratory
Virginia Tech*





1. Introduction

-
- What is the OAI-MHP?
 - General System Strategy
 - Case study: NDLTD



1.1. What is the OAI-MHP ?

- What is the Metadata Harvesting Protocol?
 - ◆ Protocol to transfer metadata from a source archive to a destination archive
 - Any metadata
 - In a continuous stream
 - As simply as possible

1.2. General System Strategy



Services

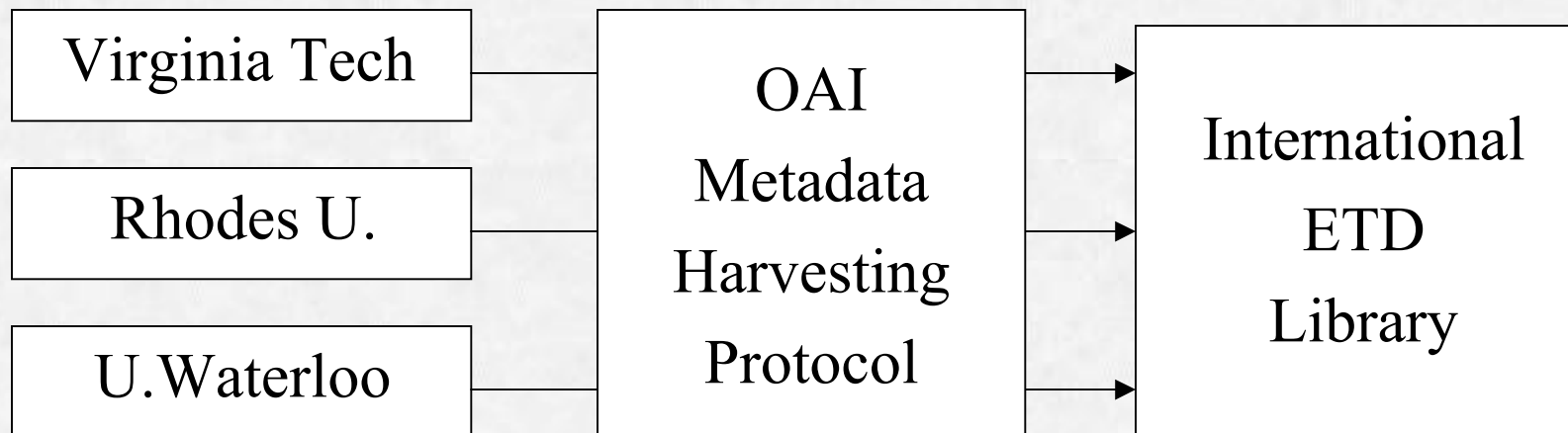
Metadata Harvesting

Document Model

1.3. Case Study: NDLTD



- Networked Digital Library of Theses and Dissertations
- Multiple independent university-based collections of electronic documents



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. Federated 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)
 - OAI uses these denotations for its client/server model (data=server, service=client)

2.4. HTTP and XML



-
- Metadata Harvesting Protocol 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. 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>
```

2.7. 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)

2.8. 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

2.9. 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

2.10. 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

3. Metadata Harvesting Protocol



-
- Service Requests
 - ◆ Identify
 - ◆ ListMetadataFormats
 - ◆ ListSets
 - ◆ GetRecord
 - ◆ ListIdentifiers
 - ◆ 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. Identify - Response

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Identify xmlns="http://www.openarchives.org/OAI/1.0/OAI_Identify"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_Identify
  http://www.openarchives.org/OAI/1.0/OAI_Identify.xsd">
  <responseDate>2001-06-14T15:09:40-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
  verb=Identify</requestURL>
  <repositoryName>Virginia Tech Electronic Thesis and Dissertation
  Collection</repositoryName>
  <baseUrl>http://oai.dlib.vt.edu:80/~hussein/cgi-
  bin/NDLTD/VTETD.pl</baseUrl>
  <protocolVersion>1.0</protocolVersion>
  <adminEmail>mailto:hussein@vt.edu</adminEmail>
- <description>
  - <oai-identifier xmlns="http://www.openarchives.org/OAI/oai-identifier"
    xsi:schemaLocation="http://www.openarchives.org/OAI/oai-identifier
    http://www.openarchives.org/OAI/oai-identifier.xsd">
    <scheme>oai</scheme>
    <repositoryIdentifier>VTETD</repositoryIdentifier>
    <delimiter>:</delimiter>
    <sampleIdentifier>oai:VTETD:etd-171110282975860</sampleIdentifier>
    </oai-identifier>
  </description>
- <description>
  <eprints xmlns="http://www.openarchives.org/OAI/eprints"
    xsi:schemaLocation="http://www.openarchives.org/OAI/eprints
    http://www.openarchives.org/OAI/eprints.xsd">
  </description>
</Identify>
```



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>

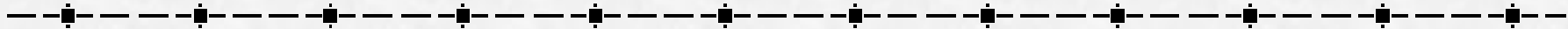
3.4. ListMetadataFormats - Response



```
<?xml version="1.0" encoding="UTF-8" ?>
- <ListMetadataFormats
  xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListMetadataFormats"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_ListMetadataFormats
  http://www.openarchives.org/OAI/1.0/OAI_ListMetadataFormats.xsd">
  <responseDate>2001-06-14T15:12:53-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
  verb=ListMetadataFormats</requestURL>
- <metadataFormat>
  <metadataPrefix>oai_rfc1807</metadataPrefix>
  <schema>http://www.openarchives.org/OAI/rfc1807.xsd</schema>
  <metadataNamespace>http://info.internet.isi.edu:80/in-
  notes/rfc/files/rfc1807.txt</metadataNamespace>
  </metadataFormat>
- <metadataFormat>
  <metadataPrefix>oai_dc</metadataPrefix>
  <schema>http://www.openarchives.org/OAI/dc.xsd</schema>
  <metadataNamespace>http://purl.org/dc/elements/1.1/</metadataNamespace>
  </metadataFormat>
  </metadataFormat>
+ <metadataFormat>
</ListMetadataFormats>
```



3.5. 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>

3.6. ListSets – Response



```
<?xml version="1.0" encoding="UTF-8" ?>
- <ListSets xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListSets"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_ListSets
  http://www.openarchives.org/OAI/1.0/OAI_ListSets.xsd">
  <responseDate>2001-06-14T15:14:30-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=ListSets</requestURL>
  - <set>
    <setSpec>All</setSpec>
    <setName>All theses and dissertations</setName>
  </set>
</ListSets>
```



3.7. 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
 - ◆ http://www.anarchive.org/cgi-bin/OAI?verb=GetRecord&identifier=oai:test:123&metadataPrefix=oai_dc

3.8. GetRecord - Response



```
<?xml version="1.0" encoding="UTF-8" ?>
- <GetRecord xmlns="http://www.openarchives.org/OAI/1.0/OAI_GetRecord"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_GetRecord
  http://www.openarchives.org/OAI/1.0/OAI_GetRecord.xsd">
  <responseDate>2001-06-14T15:16:09-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=GetRecord&identifier=oai:VTETD:etd-
    3123162539751141&metadataPrefix=oai_dc</requestURL>
- <record>
  - <header>
    <identifier>oai:VTETD:etd-3123162539751141</identifier>
    <timestamp>1997-04-22</timestamp>
  </header>
  - <metadata>
    - <dc xmlns="http://purl.org/dc/elements/1.1/"
      xsi:schemaLocation="http://purl.org/dc/elements/1.1/
      http://www.openarchives.org/OAI/dc.xsd">
      <title>SMA-Induced Deformations In general Unsymmetric
        Laminates</title>
      <creator>Dano, Marie-Laure</creator>
      <subject>Engineering Science and Mechanics</subject>
      <description>General unsymmetric laminates exhibit large natural
        curvatures at room temperature. Additionally, inherent to most
        unsymmetric laminates is the presence of two stable configurations.
        Multiple configurations and stability issues arise because of the
        geometric nonlinearities associated with the large curvatures. The
        laminate can be changed from one stable configuration to the other by
```



3.9. ListIdentifiers

-
- Purpose
 - ◆ List all unique identifiers corresponding to records in the repository
 - Parameters
 - ◆ from – start date (O)
 - ◆ until – end date (O)
 - ◆ set – set to harvest from (O)
 - ◆ resumptionToken – flow control mechanism (X)
 - Sample URL
 - ◆ <http://www.anarchive.org/cgi-bin/OAI?verb=ListIdentifiers&set=All>

3.10. ListIdentifiers - Response



```
<?xml version="1.0" encoding="UTF-8" ?>
- <ListIdentifiers xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers
  http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers.xsd">
  <responseDate>2001-06-14T15:17:32-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=ListIdentifiers</requestURL>
  <identifier>oai:VTETD:etd-3345131939761081</identifier>
  <identifier>oai:VTETD:etd-171110282975860</identifier>
  <identifier>oai:VTETD:etd-05012000-14030054</identifier>
  <identifier>oai:VTETD:etd-3621112139711101</identifier>
  <identifier>oai:VTETD:etd-133422039701091</identifier>
  <identifier>oai:VTETD:etd-23281533974920</identifier>
  <identifier>oai:VTETD:etd-123322282975860</identifier>
  <identifier>oai:VTETD:etd-255314202974780</identifier>
  <identifier>oai:VTETD:etd-335713312971890</identifier>
  <identifier>oai:VTETD:etd-104722369631841</identifier>
  <identifier>oai:VTETD:etd-831102339731121</identifier>
  <identifier>oai:VTETD:etd-454016449701231</identifier>
  <identifier>oai:VTETD:etd-3034112939721181</identifier>
  <identifier>oai:VTETD:etd-522014589642481</identifier>
  <identifier>oai:VTETD:etd-274210359611541</identifier>
  <identifier>oai:VTETD:etd-3210192049721391</identifier>
  <identifier>oai:VTETD:etd-0521318109613220</identifier>
  <identifier>oai:VTETD:etd-310141259631631</identifier>
  <identifier>oai:VTETD:etd-12164379662151</identifier>
```



3.11. 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
 - ◆ http://www.anarchive.org/cgi-bin/OAI?verb=ListRecord&metadataprefix=oai_dc&from=2001-01-01

3.12. ListRecords - Response



```
<?xml version="1.0" encoding="UTF-8" ?>
- <ListRecords xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListRecords"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_ListRecords
  http://www.openarchives.org/OAI/1.0/OAI_ListRecords.xsd">
  <responseDate>2001-06-14T15:19:19-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=ListRecords&metadataPrefix=oai_dc</requestURL>
+ <record>
+ <record>
- <record>
  - <header>
    <identifier>oai:VTETD:etd-05012000-14030054</identifier>
    <timestamp>2000-05-01</timestamp>
  </header>
  - <metadata>
    - <dc xmlns="http://purl.org/dc/elements/1.1/"
      xsi:schemaLocation="http://purl.org/dc/elements/1.1/
      http://www.openarchives.org/OAI/dc.xsd">
      <title>An Examination of Race and Recurrent Substance Problems in the
        United States</title>
      <creator>Bell, Tannisha D.</creator>
      <subject>Sociology</subject>
      <description>Several studies show that African-Americans are less likely
        than whites to use alcohol or drugs. However, if African-Americans use
        drugs then they are more likely to become heavy and persistent users.
        African-Americans are also more likely to have a current substance
        abuse disorder. There is not much in the literature to explain this
```

3.13. Metadata Multiplicity



```
<?xml version="1.0" encoding="UTF-8" ?>
- <GetRecord xmlns="http://www.openarchives.org/OAI/1.0/OAI_GetRecord"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_GetRecord
  http://www.openarchives.org/OAI/1.0/OAI_GetRecord.xsd">
  <responseDate>2001-06-14T15:20:41-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=GetRecord&identifier=oai:VTETD:etd-
    3123162539751141&metadataPrefix=oai_rfc1807</requestURL>
- <record>
  - <header>
    <identifier>oai:VTETD:etd-3123162539751141</identifier>
    <timestamp>1997-04-22</timestamp>
  </header>
  - <metadata>
    - <rfc1807 xmlns="http://info.internet.isi.edu:80/in-
      notes/rfc/files/rfc1807.txt"
      xsi:schemaLocation="http://info.internet.isi.edu:80/in-
      notes/rfc/files/rfc1807.txt
      http://www.openarchives.org/OAI/rfc1807.xsd">
      <bib-version>1</bib-version>
      <id>etd-3123162539751141</id>
      <entry>1997-04-22</entry>
      <organization>Virginia Polytechnic Institute and State
        University</organization>
      <title>SMA-Induced Deformations In general Unsymmetric
        Laminates</title>
      <tvne>Thesis/Dissertation</tvne>
```

3.14. Date Ranges

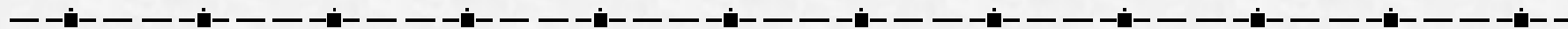


```
<?xml version="1.0" encoding="UTF-8" ?>
- <ListIdentifiers xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers"
  xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers
  http://www.openarchives.org/OAI/1.0/OAI_ListIdentifiers.xsd">
  <responseDate>2001-06-14T15:21:37-05:00</responseDate>
  <requestURL>http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl?
    verb=ListIdentifiers&from=2000-11-24&until=2000-12-01</requestURL>
  <identifier>oai:VTETD:etd-11212000-155513</identifier>
  <identifier>oai:VTETD:etd-11242000-130040</identifier>
  <identifier>oai:VTETD:etd-11272000-115149</identifier>
  <identifier>oai:VTETD:etd-11162000-19160016</identifier>
  <identifier>oai:VTETD:etd-11222000-095443</identifier>
  <identifier>oai:VTETD:etd-11142000-16540027</identifier>
  <identifier>oai:VTETD:etd-11282000-110022</identifier>
  <identifier>oai:VTETD:etd-11152000-13100048</identifier>
  <identifier>oai:VTETD:etd-11272000-114011</identifier>
  <identifier>oai:VTETD:etd-11182000-10350010</identifier>
  <identifier>oai:VTETD:etd-11272000-214847</identifier>
  <identifier>oai:VTETD:etd-11182000-16030010</identifier>
</ListIdentifiers>
```

3.15. Resumption Token

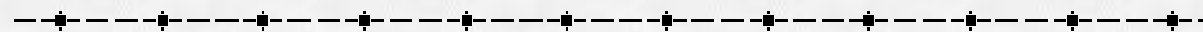


```
<identifier>oai:VTETD:etd-254122839711171</identifier>
<identifier>oai:VTETD:etd-4524171049761291</identifier>
<identifier>oai:VTETD:etd-3156151139751001</identifier>
<identifier>oai:VTETD:etd-424817300974290</identifier>
<identifier>oai:VTETD:etd-13514459731541</identifier>
<identifier>oai:VTETD:etd-2047101569611961</identifier>
<identifier>oai:VTETD:etd-5414132139711101</identifier>
<identifier>oai:VTETD:etd-3132141279612241</identifier>
<identifier>oai:VTETD:etd-3123162539751141</identifier>
<identifier>oai:VTETD:etd-556181169641921</identifier>
<identifier>oai:VTETD:etd-342482139711101</identifier>
<identifier>oai:VTETD:etd-1913943975930</identifier>
<identifier>oai:VTETD:etd-402515359721531</identifier>
<identifier>oai:VTETD:etd-2025212339731121</identifier>
<identifier>oai:VTETD:etd-3331171059721601</identifier>
<identifier>oai:VTETD:etd-18409759651581</identifier>
<identifier>oai:VTETD:etd-34521672975650</identifier>
<identifier>oai:VTETD:etd-120142139711101</identifier>
<identifier>oai:VTETD:etd-4019122049721391</identifier>
<identifier>oai:VTETD:etd-487142639761151</identifier>
<resumptionToken>!!!!100</resumptionToken>
</ListIdentifiers>
```

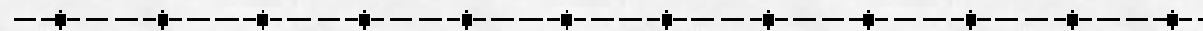



That's All Folks !

The OAI Metadata Harvesting Protocol - Communities and Services



*Hussein Suleman, hussein@vt.edu
Digital Library Research Laboratory
Virginia Tech*

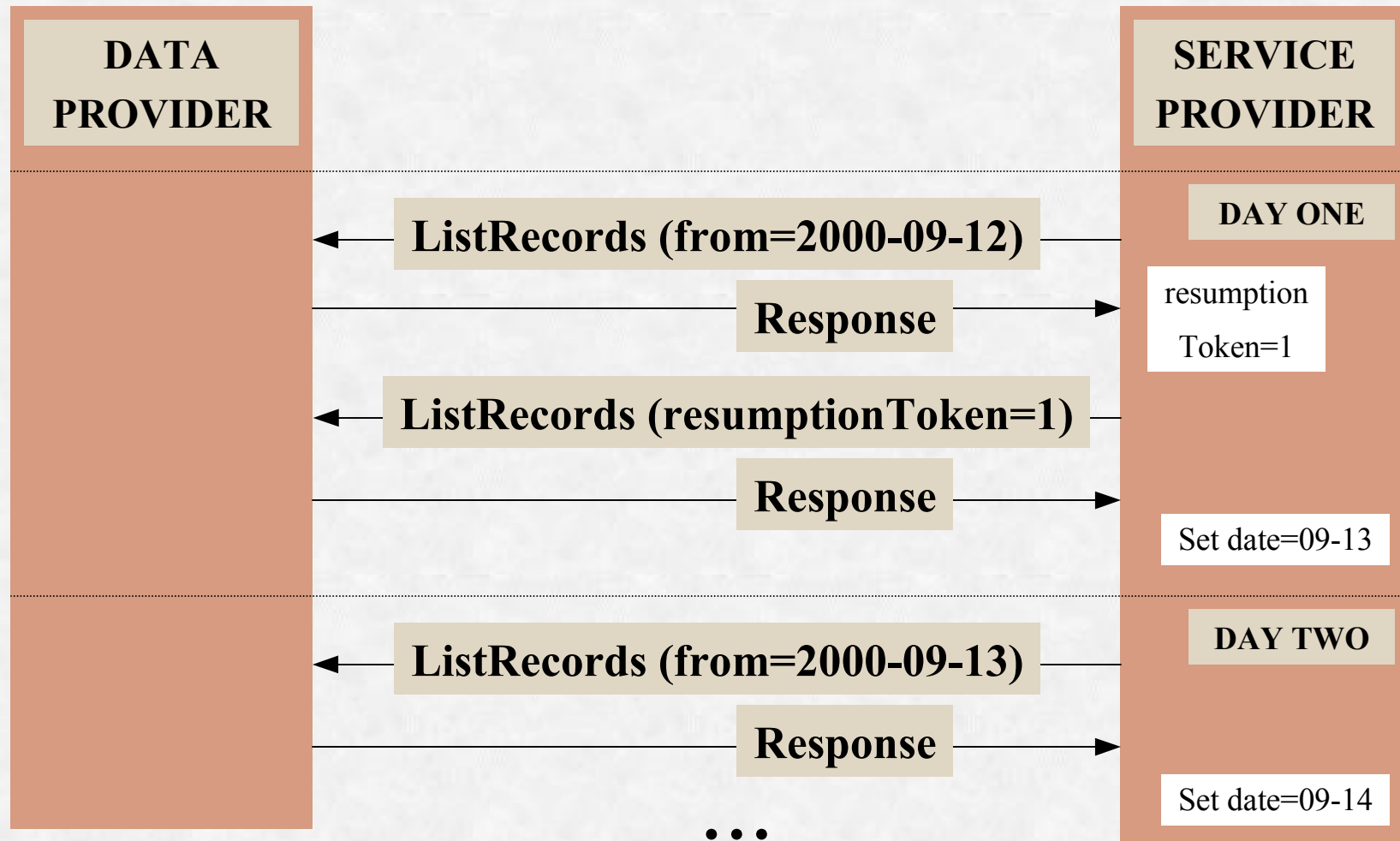




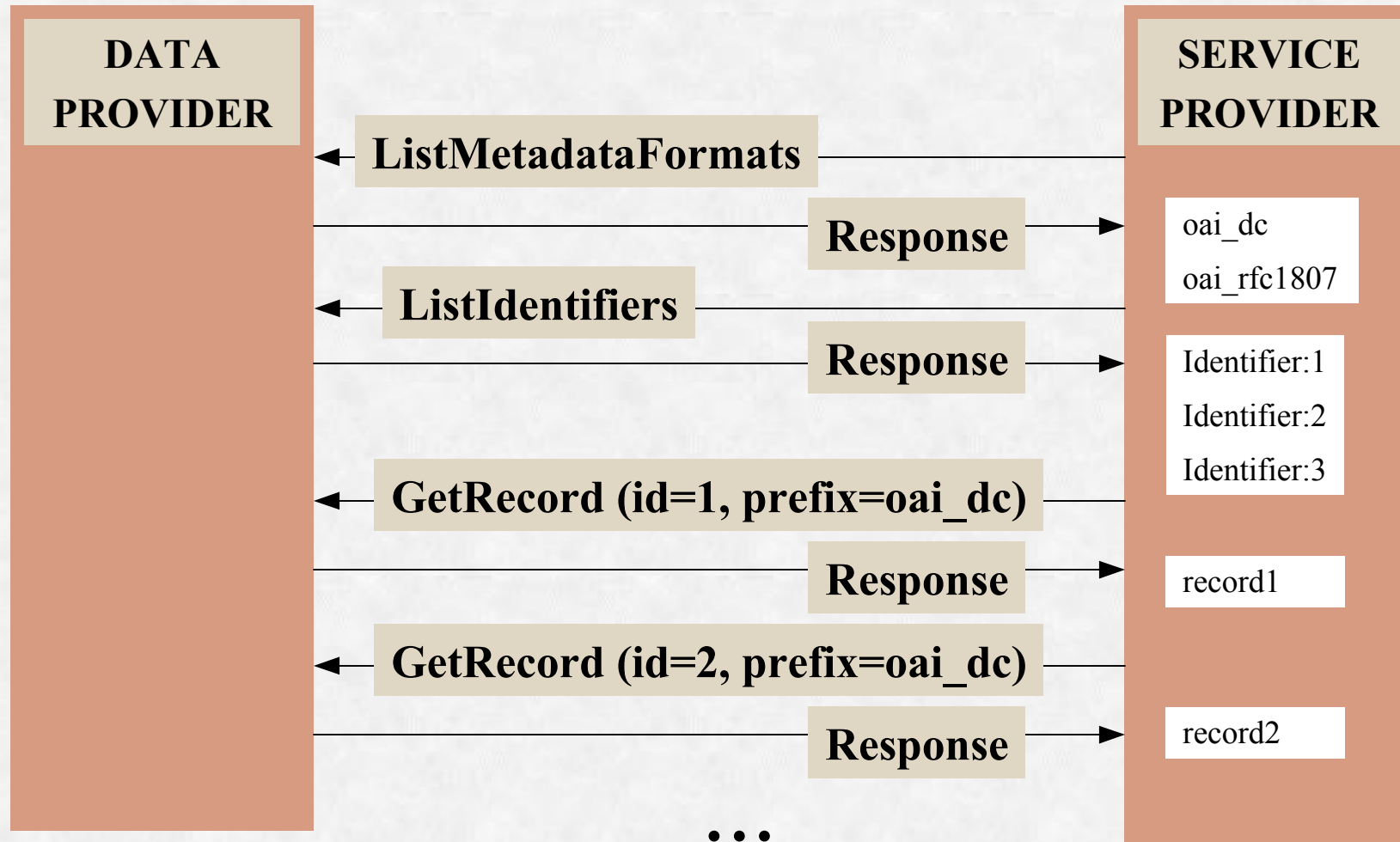
4. Service Providers

- Harvesting 101/102/103
- Scheduling
- Tools
- Repository Explorer
- Case Study: ARC
- Case Study: NDLTD
- VTLS Virtua

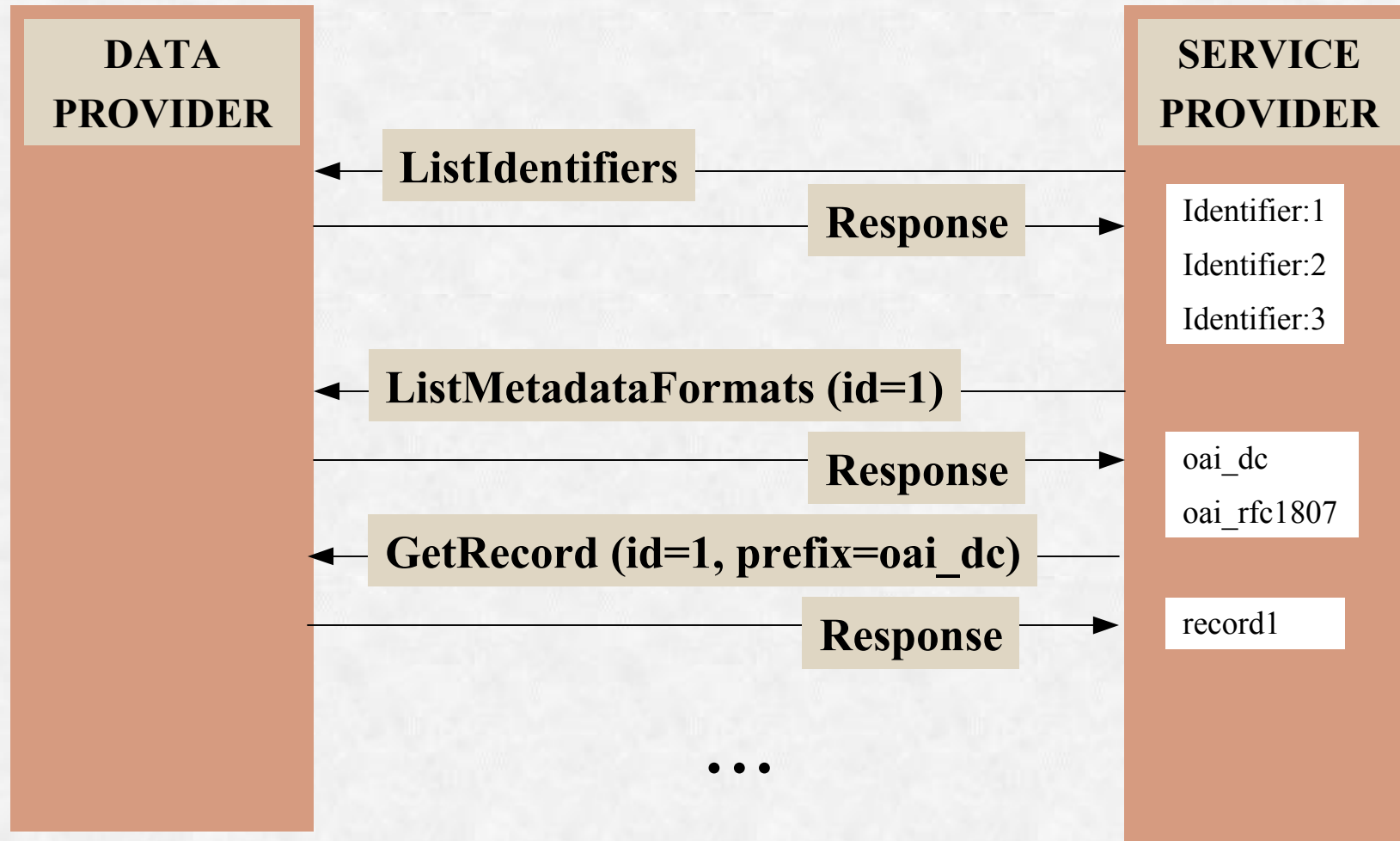
4.1. Harvesting 101



4.2. Harvesting 102



4.3. Harvesting 103



4.4. Scheduling



- Problems:
 - ◆ Granularity is coarse
 - ◆ Timezones are local for each site
- Solutions:
 - ◆ Overlap one day to compensate for granularity
 - ◆ Overlap one day or use remote times to compensate for timezones

4.5. Tools



-
- Check OAI website for sample code
 - XML parsers – depending on platform – check W3C
 - XML Schema validators
 - ◆ Very few available – the reference version works but may not be easy to install
 - ◆ Ignore validation if you can trust the source
 - Sample data providers – check the OAI website for a list of conformant public archives

4.6. Repository Explorer



Open Archives Initiative - Repository Explorer

explorer version - 1.3 : protocol version - 1.0/1.1 : August 2001

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

American Memory [LoC]	▲
arXiv	
BioMed Central	
Computer Science Teaching Center	▼

[[View Archive Website](#)] [[Test and Add an archive to this list](#)]

Verbs	Parameters
Identify	from (YYYY-MM-DD) : <input type="text"/>
List Metadata Formats	until (YYYY-MM-DD) : <input type="text"/>
List Sets	metadataPrefix : <input type="text"/>

4.7. Case Study: ARC

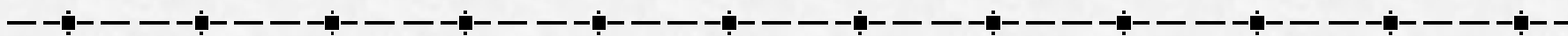


The screenshot shows a Netscape browser window titled "ARC - Netscape". The address bar is empty. The main content area displays the "arc" logo and the text "Cross Archive Searching Service". Below this is a navigation bar with "Simple search", "Advanced Search", and "Help" links. A message states "This is page 1, hits (1--10) of total 66 hits." and "Results Pages: 1 2 3 4 5 6 7". A "SEARCH RESULTS" section follows, listing search results with fields for Title, Creators, Description, Archive, and Document ID. The first result is "Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data" by Nowell, Lucille Terry. The description mentions a user interface design for a digital library. The archive is NDLTD and the document ID is oai:VTETD:etd-111897-163723. At the bottom, a footer note states: "This prototype is based on the UPS project and the NCSTRL+ based digital library developed by Old Dominion University".

Title	<i>Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size to Convey Nominal and Quantitative Data</i>
Creators	Nowell, Lucille Terry
Description	centerH2Graphical Encoding for Information Visualization: Using Icon Color, Shape, and Size To Convey Nominal and Quantitative Data/h2/center centerH3Lucille Terry Nowell/H3/centerBCenterABSTRACT/Center/BpIn producing a user interface design to visualize search results for a digital library called En
Archive	NDLTD
Document ID	oai:VTETD:etd-111897-163723

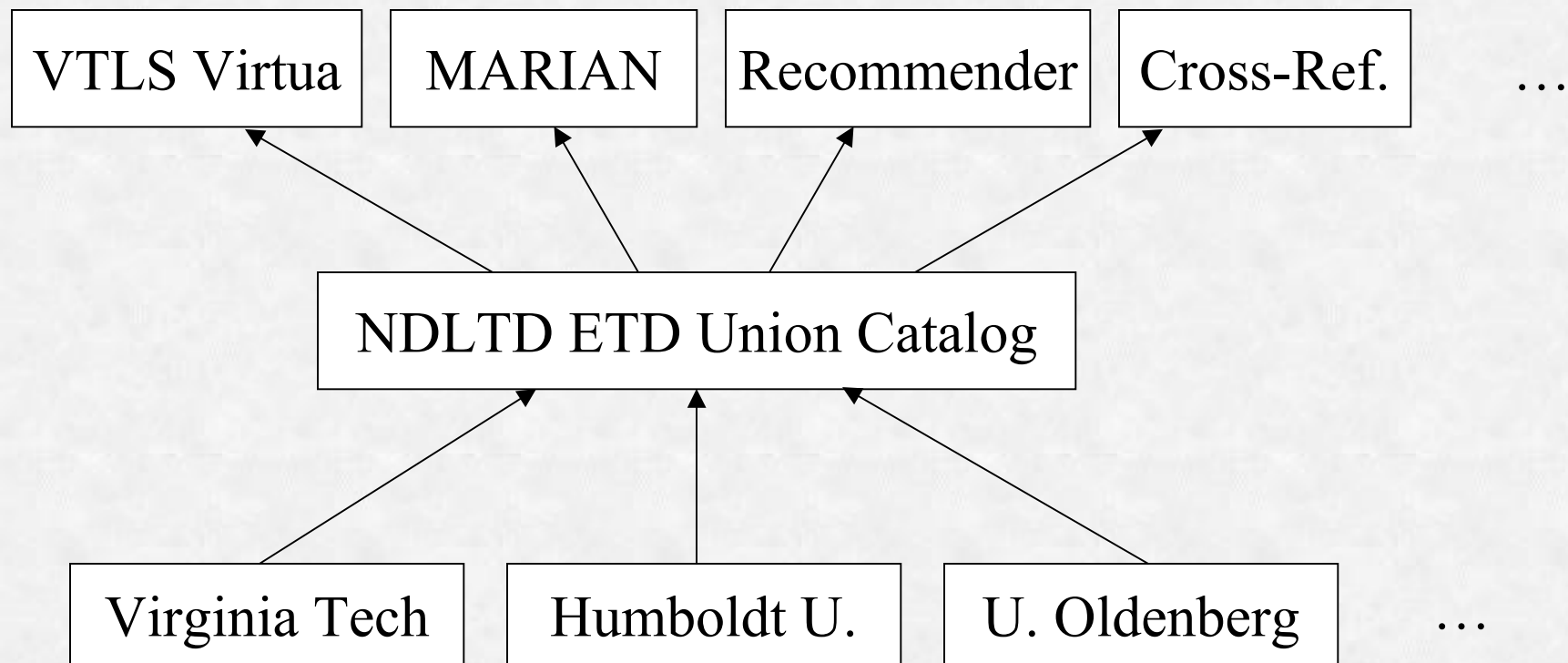
This prototype is based on the UPS project and the NCSTRL+ based digital library developed by Old Dominion University

4.8. Case Study: NDLTD



Search/Browse Engines

Other Services



4.9. VTLS Virtua



NETWORKED DIGITAL LIBRARY
OF THESES AND DISSERTATIONS

SEARCHES [Browse](#) | [Keyword](#) | [Expert](#)

English
Spanish
Korean

[New Session](#) | [Save Session](#) | [Cart](#) | [History](#) | [Help](#)

To begin a search...

Enter your search terms.

Select a search type.

Choose a target
database.

NDLTD Union Catalog

Click the **Search**
button.

Welcome to the Networked Digital Library of Theses and Dissertations Union Catalog. This Union Catalog will serve as a repository of rich graduate educational material contributed by a number of member institutions worldwide. Our hope is that this project will increase the availability of student research for scholars, empower students to convey a richer message through the use of multimedia and hypermedia technologies and advance digital library technology worldwide.

This project is a joint development with NDLTD and VTLS Inc. [Read the press release here.](#)

Browse the catalog by theses and dissertation author, subject, title and library call number. Click on "Browse Search" in the toolbar to initiate your search.

Search for words or exact phrases anywhere in the theses and dissertation catalog record. Specify a keyword field such as author, title or subject to narrow that search. Click on "Keyword Search" in the toolbar to begin.

The Expert search function allows you to use Boolean operators in a command mode. Click on "Expert Search" to get started and follow the helpful search instructions below the data entry box.



Copyright © 2001 VTLS Inc. All rights reserved. 1701 Kraft Drive, Blacksburg, VA USA 24060 | <http://www.vtls.com>



5. OAI Communities

-
- Shared Metadata Formats
 - Shared semantics
 - Layering over OAI
 - Closed OAI networks
 - OAI within the DL

5.1. Shared Metadata Formats



- Use metadata formats accepted within a community to convey more specific information
- Examples
 - ◆ E-Print format (under development)
 - ◆ ETD-MS for theses and dissertations
 - ◆ VRA Core for multimedia
 - ◆ IMS Metadata for educational material



5.2. Shared Semantics

- Develop a shared understanding for the meanings of fields
- Examples
 - ◆ Developing controlled vocabularies for fields
 - ◆ Using specific fields for external links (OAI recommends using identifier in DC for this)
 - ◆ Choosing from among existing standards (like language names)



5.3. Layering over OAI

-
- Convert OAI records into more standard formats like MARC communications format
 - Collapse multiple requests into one to make harvesting easier
 - Name authority system (developed at OCLC) piggybacks name resolution over the OAI protocol

5.4. Closed OAI networks



-
- Data providers need not go public !
 - Within an organization, OAI can be used for data transfer among heterogeneous systems
 - More control over use, making global optimizations possible (like harvesting schedules and choice of metadata formats)



5.5. OAI within the DL

-
- Use the OAI protocol as the basis for components to communicate
 - Examples
 - ◆ Search Engines could use dynamic sets to correspond to search results
 - ◆ Browsing can be directed by sets
 - ◆ Reviews and Annotations can each be independent OAI data providers

6. Now What ?



-
- Reality Check
 - Links
 - More Links



6.1. Reality Check

-
- DO I REALLY WANT TO DO THIS?
 - Can I satisfy the requirements to be a data provider?
 - Do I want to be a service provider ?
 - Do I want to adopt and support this within my community ?



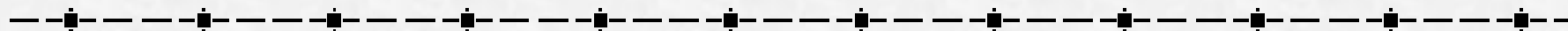
6.2. Links

- 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
- NDLTD
 - ◆ <http://www.ndltd.org>



6.3. More Links

- 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 !