Enforcing Interoperability with the Open Archives Initiative Repository Explorer

Hussein Suleman, hussein@vt.edu
Digital Library Research Laboratory
Virginia Tech
Overview of the OAI

- What is the Open Archives Initiative?
- What is the Metadata Harvesting Protocol?
- What is an Open Archive?
- History of the OAI
- Development of the OAI Protocol
- Brief description of the OAI Protocol
Repository Explorer

🌟 Technical Testing Tool
🌟 Simulates use of an archive (aka a Repository) by a typical service provider
🌟 Allows browsing of an archive’s contents using only the OAI protocol (aka Exploring)
🌟 Checks syntax and encoding of data
🌟 Interactive and Non-interactive modes of testing
Motivation

- Chicken-and-egg problem with protocol
- Lack of service providers
  - Need for “stub” routines
- Ambiguity in interpretation of protocol
  - Need for validation of implementations
- Data quality assurances
Validation Procedure

- HTTP – check for server errors
- Schema – check for conformance to XML Schema Description
- Parser – check for validity of XML structure and encoding
- Protocol – check protocol-specific requirements
Interactive Browsing

Open Archives Initiative - Repository Explorer

explorer version - 1.1 : protocol version - 1.0 : April 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 and Netscape Navigator is recommended

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] ]
[ Digital Cultural Heritage Community [UIUC] ]
[ Perseus [Open Language Archives] ]
[ RDN [Resource Discovery Network] ]

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

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify</td>
<td>from (YYYY-MM-DD) :</td>
</tr>
<tr>
<td>List Metadata Formats</td>
<td>until (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</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Online schemata</td>
</tr>
</tbody>
</table>

*Send all comments to hussein@vt.edu --- Digital Libraries Research Laboratory@Virginia Tech*
## Browsing

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>Virginia Tech Electronic Thesis and Dissertation Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base URL</td>
<td><a href="http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTED.pl">http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTED.pl</a></td>
</tr>
<tr>
<td>Protocol Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Admin Email</td>
<td><a href="mailto:webmaster@scholar.lib.vt.edu">mailto:webmaster@scholar.lib.vt.edu</a></td>
</tr>
<tr>
<td>Other Information</td>
<td>description:</td>
</tr>
<tr>
<td></td>
<td>oai-identifier:</td>
</tr>
<tr>
<td></td>
<td>scheme: oai</td>
</tr>
<tr>
<td></td>
<td>repositoryIdentifier: VTED</td>
</tr>
<tr>
<td></td>
<td>delimiter: :</td>
</tr>
<tr>
<td></td>
<td>sampleIdentifier: oai:VTED:etd-171110282975860</td>
</tr>
<tr>
<td></td>
<td>description:</td>
</tr>
<tr>
<td></td>
<td>eprints:</td>
</tr>
<tr>
<td></td>
<td>content:</td>
</tr>
<tr>
<td></td>
<td>text: Theses and Dissertations produced by students at Virginia</td>
</tr>
<tr>
<td></td>
<td>metadataPolicy:</td>
</tr>
<tr>
<td></td>
<td>text: Metadata may be used by commercial and non-commercial users</td>
</tr>
<tr>
<td></td>
<td>dataPolicy:</td>
</tr>
<tr>
<td></td>
<td>text: Full texts are individually tagged and the rights statement</td>
</tr>
</tbody>
</table>
Open Archives Initiative - Repository Explorer

http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTED.pl?verb=ListSets

List of Sets

Click on the link to list the contents

All theses and dissertations

Request URL: http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTED.pl?verb=ListSets
Response Date: 2001-06-14T20:27:00-05:00

<table>
<thead>
<tr>
<th>Verbs</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>from (YYYY-MM-DD)</td>
<td></td>
</tr>
</tbody>
</table>
Browsing

Open Archives Initiative - Repository Explorer

`http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=ListIdentifiers&set=All`

List of Record Identifiers

Select a link to view more information

- oai:VTETD:etd-3345131939761081 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-171110282975860 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-05012000-14030054 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-3621112139711101 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-133422039701091 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-23261533974920 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-123322282975860 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-255314202974780 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-335713312971890 [display record in Dublin Core] [display metadata format]
- oai:VTETD:etd-104722369631841 [display record in Dublin Core] [display metadata format]
Browsing

Open Archives Initiative - Repository Explorer

explorer version - 1.1 : protocol version - 1.0 : April 2001

http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTED.pl?verb=ListMetadataFormats&identifier=oai:VTE

List of Metadata Formats

Click on the link to view schema

Prefix=[oai_rfc1807]
Schema=[http://oai.dlib.vt.edu/OAI/rfc1807.xsd]

[display record]

Prefix=[oai_marc]
NameSpace=[http://www.openarchives.org/OAI/oai_marc]
Schema=[http://oai.dlib.vt.edu/OAI/oai_marc.xsd]
Browsing

http://scholar.lib.vt.edu/theses/OAI/cgi-bin/VTETD.pl?verb=GetRecord&metadataPrefix=oai_etdms&id=3345131939761081

List of Fields

definition : oai:VTETD:etd-3345131939761081
datestamp : 1997-03-31

metadata :
thesis :
title : Conceptual Development and Empirical Testing of an Outdoor Recreation Experience Preference (REP) Scale
creator : Walker, Gordon James
publisher : Virginia Polytechnic Institute and State University
subject : outdoor recreation
subject : recreation experience preference scales
subject : recreation experience matrix
subject : recreation opportunity spectrum
description : This dissertation examines four issues, including: (a) whether outdoor recreationists prefer the Recreation Experience Preference (REP) scales exist; (b) whether these experiences can be measured using a framework called the Recreation Experience Matrix (REM); (c) how well the Recreation Opportunity Spectrum (ROS) variables of activity, setting, and expertise explain the types of experiences outdoor recreationists receive; and (d) how well two new variables--primary mode and mode dependency--predict the types of experiences outdoor recreationists receive. In order to address these issues, the questionnaire was distributed to 410 people at Mount Rogers National Recreation Area in Virginia during October and November. A total of 336 people completed this questionnaire. Of these, 336 provided useable answers. The answers were analyzed using multiple regression analysis, and the results are presented herein.
Multiple views of data

Raw XML Output

```xml
<?xml version="1.0" encoding="UTF-8"?>
<ListSets xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListSets" xmlns:
    "http://www.openarchives.org/OAI/1.0/OAI_setStatus"
AMILY xmlns
   ="http://www.openarchives.org/OAI/1.0/OAI_Res

<responseDate>2001-06-14T20:32:33-05:00</responseDate>

<set>
  <setSpec>All</setSpec>
  <setName>All theses and dissertations</setName>
</set>

</ListSets>
```
Multilingual Support

Open Archives Initiative - 揣紓坪霜

Harvesting 坍匈OAI ??紫偶腔賜醲 [ 剪殼載巋僼渀 薎傾躍鬱 ]

剉繡猛厥JavaScript 膨枚朧-sn Netscape 霜擬

貂礦：巋旋轍 HTTP 渣吻， 脹善善熙瑞煩縊膱utory綺ostringstream

懷 善OAI賜醲腔URL(? 坡郾搞 ) 漕植瑩桶制度善顧熙瑞藉珂隔疊腔紫偶

American Memory [LoC]
Digital Cultural Heritage Community [UIUC]
Perseus [Open Language Archives]
RDN [Resource Discovery Network]

[ 脹艘紫偶蜃侖 ][ 聳淞甜 preca 璃.UltraWin紫偶善遝桶篇 ]
Automatic Test Suite

Open Archives Initiative :: Metadata Harvesting Protocol v1.0
Protocol Tester v1.1 :: Virginia Tech DLRL :: April 2001

Testing: Identify
URL: http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=Identify
Test Result: OK
---- [ Repository Name = Virginia Tech Electronic Thesis and Dissertation Collection ]
---- [ Protocol Version = 1.0 ]
---- [ Base URL = http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTD/VTETD.pl ]
---- [ Admin Email = mailto:hussein@vt.edu ]

Testing: ListMetadataFormats
URL: http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=ListMetadataFormats
Test Result: OK
---- [ Sample Metadata Format = oai_rfc1807 ]

Testing: ListSets
URL: http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTD/VTETD.pl?verb=ListSets
Test Result: OK
---- [ Sample Set Spec: All ]
## Error in Response

### Archive Self-Description

<table>
<thead>
<tr>
<th>Repository Name</th>
<th>Virginia Tech Electronic Thesis and Dissertation Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base URL</td>
<td><a href="http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTDErr1/VTED.pl">http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTDErr1/VTED.pl</a></td>
</tr>
<tr>
<td>Protocol Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Error: Missing field:</td>
<td><code>&lt;Identify&gt;</code>/<code>&lt;adminEmail&gt;</code></td>
</tr>
</tbody>
</table>

### Other Information

```xml
<description>
  oai-identifier:
    scheme: oai
    repositoryIdentifier: VTED
    delimiter: : 
    sampleIdentifier: oai:VTED:etd-171110282975860
  description:
    eprints:
      content:
        text: Theses and Dissertations produced by students at Virginia
    metadataPolicy:
      text: Metadata may be used by commercial and non-commercial users
    dataPolicy:
      text: Full texts are individually tagged and the rights statement
```
Error in XML

XSD Schema/Instance Validation Error!

Errors in XML instance

<?xml version='1.0'?>
<xsv docElt='(http://www.openarchives.org/OAI/1.0/OAI_Identify)Identify' instanceAsses
<importAttempt URI='http://oai.dlib.vt.edu/OAI/1.0/OAI_Identify.xsd' namespace='http://w
<importAttempt URI='http://oai.dlib.vt.edu/OAI/oai-identifier.xsd' namespace='http://w
<importAttempt URI='http://oai.dlib.vt.edu/OAI/eprints.xsd' namespace='http://www.open
<invalid char='4' code='cvc-complex-type.1.2.4' line='11' resource='file:///tmp/file2V

</node>
<node id='1'>
<edge dest='2' label='(http://www.openarchives.org/OAI/1.0/OAI_Identify):responseDate'
</node>
<node id='2'>
<edge dest='3' label='(http://www.openarchives.org/OAI/1.0/OAI_Identify):requestURL'/
</node>
<node id='3'>
Conclusions and Future Work

- Compliance testing is a critical part of protocol development and support in interoperability efforts

- Support future versions of OAI protocol

- Generalize to other protocols
That’s All Folks!