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
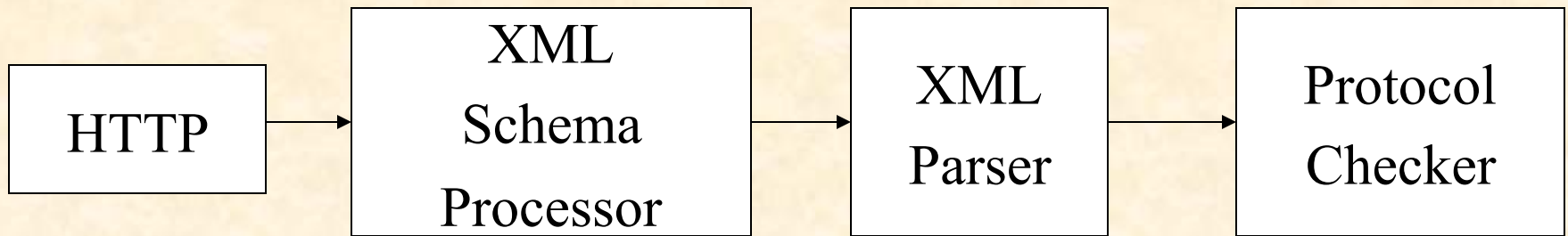* Brief description of the OAI Protocol

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

```
┌─────────┐     ┌─────────────┐     ┌─────────┐     ┌───────────┐
│         │     │    XML      │     │   XML   │     │ Protocol  │
│  HTTP   │ ──▶ │   Schema    │ ──▶ │ Parser  │ ──▶ │  Checker  │
│         │     │  Processor  │     │         │     │           │
└─────────┘     └─────────────┘     └─────────┘     └───────────┘
```

✳ 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

Parameter Testing

Browsing

Archive Self-Description

Repository Name	Virginia Tech Electronic Thesis and Dissertation Collection
Base URL	http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl
Protocol Version	1.0
Admin Email	mailto:webmaster@scholar.lib.vt.edu

Other Information	description:
	    oai-identifier:
	        scheme: oai
	        repositoryIdentifier: VTETD
	        delimiter: :
	        sampleIdentifier: oai:VTETD: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 user
	        dataPolicy:
	            text: Full texts are individually tagged and the rights statemer

Browsing

Browsing

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/VTETD.pl?verb=ListMetadataFormats&identifier=oai:VTE

List of Metadata Formats

Click on the link to view schema

Prefix=[oai_rfc1807]
NameSpace=[http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt]
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&i

List of Fields

```
header:
  identifier : oai:VTETD:etd-3345131939761081
  datestamp : 1997-03-31

metadata:
  thesis:
    title: Conceptual Development and Empirical Testing of an Outdoor Recreation Exp
    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 outd
the Recreation Experience Preference (REP) scales exist; (b) whether these experiences
using a framework called the Recreation Experience Matrix (REM); (c) how well the Recr
Spectrum (ROS) variables of activity, setting, and expertise explain the types of expe
recreationists receive; and (d) how well two new variables--primary mode and mode depe
types of experiences outdoor recreationists receive. In order to address these issues,
was distributed at Mount Rogers National Recreation Area in Virginia during October an
A total of 410 people completed this questionnaire. Of these, 336 provided useable add
```

Multiple views of data

1-3345131939761081

	Raw XML Output

	`<?xml version="1.0" encoding="UTF-8"?>`
	
	 `<ListSets xmlns="http://www.openarchives.org/OAI/1.0/OAI_ListSets" xmlns`
	
	 `<responseDate>2001-06-14T20:32:33-05:00</responseDate>`
	 `<requestURL>http://scholar.lib.vt.edu:80/theses/OAI/cgi-bin/VTETD.pl?ver`
	
	 `<set>`
	 `  <setSpec>All</setSpec>`
	 `  <setName>All theses and dissertations</setName>`
	 `</set>`
	
	 `</ListSets>`

31939761081	

Multilingual Support

Automatic Test Suite

Open Archives Initiative - Repository Explorer

explorer version 1.1 : protocol version 1.0 : April 2001

```
Open Archives Initiative :: Metadata Harvesting Protocolv1.0
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
```

Error in Response

Archive Self-Description

Repository Name	Virginia Tech Electronic Thesis and Dissertation Collection
Base URL	http://oai.dlib.vt.edu:80/~hussein/cgi-bin/NDLTDErr1/VTETD.pl
Protocol Version	1.0
Error: Missing field : <Identify>/<adminEmail>	

Other Information	description:
	    oai-identifier:
	        scheme: oai
	        repositoryIdentifier: VTETD
	        delimiter: :
	        sampleIdentifier: oai:VTETD: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 user
	        dataPolicy:
	            text: Full texts are individually tagged and the rights statemer

Error in XML

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

http://oai.dlib.vt.edu/~hussein/cgi-bin/NDLTDErr1/VTETD.pl?verb=Identify
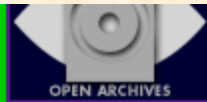
XSD Schema/Instance Validation Error !

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:/
<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
<fsm>
<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 !