

**OAI Static Repositories
(work area F)**

**Serhiy Polyakov
Mark Phillips**

**May 31, 2007
Draft 3**

Table of Contents

1. Introduction	1
2. OAI static repositories	1
2.1. Overview	1
2.2. Restrictions and Conformance Rules	1
2.3. Intermediation with a Static Repository Gateway	2
3. Setting OAI static repositories	2
3.1. Workflow and best practices	2
3.2. Creating XML file	3
3.2.1. Interface for creation of metadata for OAI static repositories	3
4. References	4
Appendix A. Static Repository XML Schema	5
Appendix B. Static Repository Example	6

1. Introduction

Among the objectives of Texas Heritage Digitization Initiative is establishing resource Digital Repositories (THDI, 2005). Any institution in Texas that meets the technical standards and executes an agreement to participate with the THDI can qualify as a Digital Repository. Another goal of THDI is establishing common standards and ensuring that Digital Repositories meet these standards.

Each repository consists of a digital asset management system (DAM). Hosting such a system requires networking infrastructure, including web servers and sufficient support staff. The DAM must provide a method of ingesting and managing digital collections. At a minimum, each DAM must be OAI compliant in order to contribute metadata to the central OAI harvester.

THDI recognizes three levels of interoperability: Minimal, Basic, and Enhanced (THDI, 2006). Minimal level of interoperability presumes that participants will make metadata available in the form of a static repository as defined by the Open Archives Initiative. Participants will create metadata in the oai_dc format and will create a single static repository XML document which will be registered with a static repository gateway as described in the DLF/NSDL *Best Practices for OAI Static Repositories* (DLF/NSDL, 2005).

2. OAI static repositories

2.1. Overview

A Static Repository provides a simple approach for exposing relatively static and small collections of metadata records through the OAI-PMH (DLF/NSDL, 2005). The Static Repository approach is targeted at organizations that:

- Have metadata collections ranging in size between 1 and 5000 records;
- Can make static content available through a network-accessible Web server;
- Need a technically simpler implementation strategy compared to acting as an OAI-PMH Repository, which requires processing OAI-PMH requests.

A Static Repository is an XML file that is made accessible at a persistent HTTP URL. The XML file contains metadata records and repository information.

A Static Repository becomes accessible via OAI-PMH through the intermediation of one Static Repository Gateway. A Static Repository Gateway uses the metadata records and repository information, provided via XML in the Static Repository, to respond to the six OAI-PMH requests for access to that information. Because a Static Repository Gateway maps a unique Static Repository base URL to each such Static Repository, harvesters can access a Static Repository in exactly the same manner as they access any other OAI-PMH Repository. A Static Repository is not a OAI-PMH Repository, because it is a file, not a server that can respond to the six OAI-PMH requests.

An XML file constituting static repository may be created manually with an XML editing tool, or a text processing application. Alternatively, a Static Repository might be generated periodically by a script that extracts information from an existing database.

2.2. Restrictions and Conformance Rules

A Static Repository must only be accessible at a single Static Repository URL. This must be an HTTP URL of the form:

`http://host:port/path/file`

The HTTP URL must not contain a fragment or a query string. The contents of the Static Repository provide metadata records and repository information necessary for intermediation by a Static Repository Gateway. There are a number of restrictions on a Static Repository relative to a standard OAI-PMH Repository. Static Repositories do not support sets, deleted records, response compression, harvesting granularity other than YYYY-MM-DD, or resumptionTokens.

Some of the conformance rules are (all rules can be found in DLF/NSDL, 2005):

- The Static Repository must validate against the XML Schema (see Appendix A).
- Contents of metadata records within the ListRecords elements of the Static Repository must conform to the OAI-PMH record format.
- The Static Repository must be a single XML file, it must not include any resumptionToken elements.

2.3. Intermediation with a Static Repository Gateway

OAI-PMH access to a Static Repository is only possible through the intermediation of a Static Repository Gateway. To initiate this intermediation, the administrator of a Static Repository must select one Static Repository Gateway that will act as the intermediary and must issue an HTTP GET request of the form:

<Static Repository Gateway URL>?initiate=<Static Repository URL>

To terminate intermediation by a Static Repository Gateway, the administrator of a Static Repository must first either remove the Static Repository from its Static Repository URL, or change the baseURL element so that it no longer matches the Static Repository base URL. The administrator must then issue an HTTP GET request of the form:

<Static Repository Gateway URL>?terminate=<Static Repository URL>

3. Setting OAI static repositories

3.1. Workflow and best practices

The following components are required to run OAI static repository:

- standard web-server
- XML file (content of a repository) in a specific metadata format made accessible at a persistent HTTP URL
- OAI Static Repository Gateway, operated by a third party

The data provider needs to contact the OAI Static Repository Gateway owner and provide direction to the HTTP-accessible XML file.

Patrick Hochstenbach and Herbert van de Sompel have written a tutorial (Hochstenbach& Sompel, 2004) on how to develop an OAI Static Repository, using LANL Gateway software. This tutorial includes installation own gateway which is optional because there are some third party gateways that can be used.

The Data Provider needs to follow these steps:

- Create a valid XML metadata file (updated as necessary), using a particular metadata format and UTF-8 character encoding
- Make it available via HTTP
- Create an XML Schema for any custom metadata formats (see Appendix A for standard schema)
- Register at a Static Repository Gateway

Some Static Repository Gateway which can be used are:

- The University of Illinois at Urbana-Champaign (UIUC) <<http://web.library.uiuc.edu/grainger/staff/habing.htm> >
- Los Alamos National Laboratory <<http://libtest.lanl.gov/>>

An example of an OAI Static Repository available here: <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm#SR_example> (see Appendix B)

Static repositories can support multiple metadata formats (DLF/NSDL, 2005). As with all OAI repositories, oai_dc is required, but other formats are encouraged. Make the required additions to the <ListMetadataFormats> section of the static XML file, and add another <ListRecords metadataPrefix="my_format"> section containing all of the records in the alternate metadata format.

In general, the size of a single static repository XML file should not exceed 2.5 million bytes. This is the same recommend maximum size for a single resumptionToken worth of data when using the standard OAI-PMH. This is the maximum size of XML file that can be conveniently handled by some XML parsers.

OAI Static Repositories do not allow sets. Current best practice is to provide multiple static XML files, and therefore multiple Static Repositories, for distinct collections a data provider may have.

3.2. Creating XML file

XML file containing metadata records may be created and maintained manually by means of an XML editor. Alternatively, it might be generated periodically by a script that extracts information from an existing database. Script for extraction of the metadata and creating XML files depends of the specific database, its structure, and frequency of updates.

Typically a given institution or individual should publish the metadata for all its resources in a single static repository. An exception is appropriate when different scripts are used to generate repositories for distinct collections from different databases.

3.2.1. Interface for creation of metadata for OAI static repositories

Static repositories must validate against the Static Repository XML Schema. This schema imports elements from the OAI-PMH v2.0 namespace via an intermediate Restricted OAI-PMH XML Schema which enforces a number of the Static Repository Conformance Rules by restricting OAI-PMH v2.0 type definitions. The schema for validating a static repository is found at:

<<http://www.openarchives.org/OAI/2.0/static-repository.xsd>> (see Appendix A)

Schema driven XML editors may be used to enter metadata and creating XML file for static repository. One of the freely available schema driven XML editors is XAmple XML Editor v2.2 available at <<http://www.felixgolubov.com/XMLEditor/>>

XAmple XML Editor analyzes a given schema and then generates a document-specific graphical user interface. Unlike other XML editors, the XAmple XML editor GUI exposes not just a tree representation of the XML document but rather a logical combination of the XML document and respective XML Schema. The user interface of the XML editor is highly logical and intuitively comprehensible. To be able to prepare valid XML documents of significant complexity, a user is not required to be familiar with XML and XML Schema languages and to have any a-priori knowledge about the documents structural requirements.

Before starting working with the editor

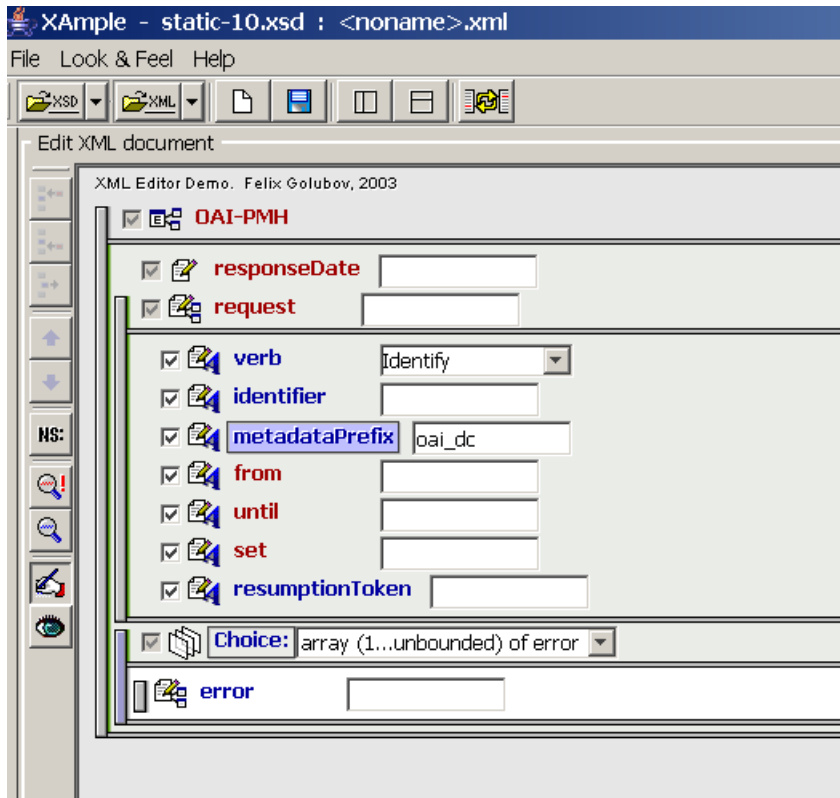
- Save Static Repository XML Schema (see Appendix A) into static.xsd file

To start working with the editor

- Download, unpack XAmple application (see manual for details)

- Start the XML editor with "run.bat". (Make sure that java 1.3+ is installed on your computer and JAVA_HOME system property is set).
- Click "Open XML Schema" button (button #1, marked "XSD") and select an *.xsd file from subdirectory where you store static.xsd file.
- If you want to continue working with the same XML file, click "Open XML Document" button (button #2) and select an *.xml file.
- Edit XML file, add records, save it, etc.

Sample data entry screen is presented below.



XML file with matadata records may be updated later. Every time XML schema has to be loaded into editor first, then target XML file can be opened and edited.

After entering records into XML file with XAmple editor this file is ready to uploading to the web server, making accessible at a persistent HTTP URL, and registering with one of the static repository gateways as described in section 2.3

4. References

THDI (2005). Texas Heritage Digitization Initiative Strategic Plan.

THDI (2006). Texas Heritage Digitization Initiative Standards and Best Practices.

DLF/NSDL. (2005). OAI Best Practices: OAI Static Repositories. <<http://oai-best.comm.nsd.org/cgi-bin/wiki.pl?StaticRepositories>>

Hochstenbach, P. & Sompel, H. (2004). OAI Static Repository & OAI Static Repository Gateway <<http://srepod.sourceforge.net/>>

Appendix A. Static Repository XML Schema

This Schema is available at <http://www.openarchives.org/OAI/2.0/static-repository.xsd>

```
<?xml version="1.0" encoding="utf-8"?>
<schema targetNamespace="http://www.openarchives.org/OAI/2.0/static-repository"
  xmlns="http://www.w3.org/2001/XMLSchema"
  xmlns:sr="http://www.openarchives.org/OAI/2.0/static-repository"
  xmlns:oai="http://www.openarchives.org/OAI/2.0/"
  elementFormDefault="qualified"
  attributeFormDefault="unqualified">

  <import namespace="http://www.openarchives.org/OAI/2.0/"
    schemaLocation="http://www.openarchives.org/OAI/2.0/OAI-PMH-static-repository.xsd"/>

  <annotation>
    <documentation>
      This XML Schema specifies the structure of an OAI-PMH Static Repository.
      A Static Repository is an XML file that is valid according to this XML
      Schema and is described in
      http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm
      A Static Repository is made accessible as an XML file on a standard
      web-server. No special software is required at the end of the organization
      that makes the Static Repository available. A Static Repository becomes
      harvestable via the OAI-PMH through the intermediation of a Static
      Repository Gateway.

      This Static Repository XML Schema by Herbert Van de Sompel and Henry N.
      Jerez (Los Alamos National Laboratory, Research Library, Digital Library
      Research and Prototyping Team; original 2002-10-26), and Simeon Warner
      (Cornell University). Inspired by the Vida work by Steven Bird for OAI-PMH
      v1.0 and for the Open Languages Archives Community; see
      http://www.language-archives.org/docs/implement.html#Vida

      Beta2 release: 2004-03-30
      Release: 2004-04-23
      $Date: 2004/04/23 15:17:46 $
    </documentation>
  </annotation>

  <element name="Repository" type="sr:RepositoryType"/>

  <complexType name="RepositoryType">
    <annotation>
      <documentation>The Repository element has 2 child elements, Identify
      and ListMetadataFormats, that are derived from the OAI-PMH v2.0 XML
      Schema. The third element, ListRecords, is repeatable and is an
      extension of the ListRecords defined in the OAI-PMH v2.0 XML Schema;
      it has an additional attribute indicating the metadataPrefix of the
      included metadata records.
      </documentation>
    </annotation>
    <sequence>
      <element name="Identify" type="oai:IdentifyType"/>
      <element name="ListMetadataFormats" type="oai:ListMetadataFormatsType"/>
      <element name="ListRecords" type="sr:ListRecordsType" maxOccurs="unbounded"/>
    </sequence>
  </complexType>

  <complexType name="ListRecordsType">
    <annotation>
      <documentation>The ListRecords element contains all records with
      metadata expressed in one of the metadata formats supported by the
      Static Repository. The mandatory metadataPrefix attribute specifies the
      metadataPrefix of the included metadata; it must correspond with a value
      of the metadataPrefix element contained in the ListMetadataFormats
```

```

        element.
      </documentation>
    </annotation>
  <complexContent>
    <extension base="oai:ListRecordsType">
      <attribute name="metadataPrefix" type="oai:metadataPrefixType" use="required"/>
    </extension>
  </complexContent>
</complexType>
</schema>

```

Appendix B. Static Repository Example

```

<?xml version="1.0" encoding="UTF-8"?>
<Repository xmlns="http://www.openarchives.org/OAI/2.0/static-repository"
  xmlns:oai="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/static-repository
    http://www.openarchives.org/OAI/2.0/static-repository.xsd">
  <Identify>
    <oai:repositoryName>Demo repository</oai:repositoryName>
    <oai:baseURL>http://gateway.institution.org/oai/an.oai.org/ma/mini.xml</oai:baseURL>
    <oai:protocolVersion>2.0</oai:protocolVersion>
    <oai:adminEmail>jondoe@oai.org</oai:adminEmail>
    <oai:earliestDatestamp>2002-09-19</oai:earliestDatestamp>
    <oai:deletedRecord>no</oai:deletedRecord>
    <oai:granularity>YYYY-MM-DD</oai:granularity>
  </Identify>
  <ListMetadataFormats>
    <oai:metadataFormat>
      <oai:metadataPrefix>oai_dc</oai:metadataPrefix>
      <oai:schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</oai:schema>
      <oai:metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</oai:metadataNamespace>
    </oai:metadataFormat>
    <oai:metadataFormat>
      <oai:metadataPrefix>oai_rfc1807</oai:metadataPrefix>
      <oai:schema>http://www.openarchives.org/OAI/1.1/rfc1807.xsd</oai:schema>
      <oai:metadataNamespace>http://info.internet.isi.edu:80/in-
notes/rfc/files/rfc1807.txt</oai:metadataNamespace>
    </oai:metadataFormat>
  </ListMetadataFormats>
  <ListRecords metadataPrefix="oai_dc">
    <oai:record>
      <oai:header>
        <oai:identifier>oai:arXiv:cs/0112017</oai:identifier>
        <oai:datestamp>2001-12-14</oai:datestamp>
      </oai:header>
      <oai: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>
  </oai_dc:dc>
</oai:metadata>
</oai:record>
<oai:record>
  <oai:header>
    <oai:identifier>oai:perseus:Perseus:text:1999.02.0084</oai:identifier>
    <oai:datestamp>2002-05-01</oai:datestamp>
  </oai:header>
  <oai: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>Germany and its Tribes</dc:title>
      <dc:creator>Tacitus</dc:creator>
      <dc:type>text</dc:type>
      <dc:source>Complete Works of Tacitus. Tacitus. Alfred John Church.
        William Jackson Brodribb. Lisa Cerrato. edited for Perseus.
        New York: Random House, Inc. Random House, Inc. reprinted 1942.
      </dc:source>
      <dc:identifier>http://www.perseus.tufts.edu/cgi-bin/ptext?
        doc=Perseus:text:1999.02.0083</dc:identifier>
    </oai_dc:dc>
  </oai:metadata>
</oai:record>
</ListRecords>
<ListRecords metadataPrefix="oai_rfc1807">
  <oai:record>
    <oai:header>
      <oai:identifier>oai:arXiv:cs/0112017</oai:identifier>
      <oai:datestamp>2001-12-14</oai:datestamp>
    </oai:header>
    <oai:metadata>
      <rfc1807
        xmlns="http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt"
        xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:schemaLocation="http://info.internet.isi.edu:80/in-notes/rfc/files/rfc1807.txt
          http://www.openarchives.org/OAI/1.1/rfc1807.xsd">
        <bib-version>v2</bib-version>
        <id>cs/0112017</id>
        <entry>December 23, 2001</entry>
        <title>Using Structural Metadata to Localize Experience of
          Digital Content</title>
        <author>Naomi Dushay</author>
        <date>December 14, 2001</date>
      </rfc1807>
    </oai:metadata>
    <oai:about>
      <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:publisher>Los Alamos arXiv</dc:publisher>
        <dc:rights>Metadata may be used without restrictions as long as
          the oai identifier remains attached to it.</dc:rights>
      </oai_dc:dc>
    </oai:about>
  </oai:record>
</ListRecords>
</Repository>

```