OAI Tools

This handout provides an overview of OAI-implementation tools, grouped according to their role in the OAI-implementation process: (1) OAI Data Provider Tools and Scripts, (2) Digital Library Systems, and (3) OAI Validation and Harvesting Systems.

A wide array of technological resources serves the needs of OAI data and service providers, ranging from simple scripts to entire systems. OAI Data Provider Tools and Scripts expose metadata via the OAI-PMH. Digital Library Systems employ tools for creating, editing, managing, and sharing metadata. OAI Validation and Harvesting Systems collect and display shareable metadata.

SEE ALSO: **OAI "Cheat Sheet"** (strategies for converting metadata formats into shareable OAI) and **Summary of OAI Metadata Best Practices** (common metadata formats, quality issues in sharing metadata, and best practices for OAI data and service providers).

OAI Data Provider Tools & Scripts

When selecting tools for managing metadata, consider not only whether they meet minimum OAI requirements (simple or unqualified Dublin Core format) but also whether they include other features for increasing the quality of shareable metadata. For instance, can the tool support multiple metadata formats? Does it support harvesting in sets (as opposed to harvesting the full repository)? Does it allow harvesting to be done in batches using resumption tokens? Is it possible to indicate to harvesters when a record has been deleted?

- **Metadata Migrator Tool** **http://metacluster.library.emory.edu/mosc/upload.php**
  Web-based program for migrating local files (formatted in .csv, .tab, or .dbf) into simple Dublin Core XML-formatted files. Available for free; register to receive login and password. *Allows* harvesting using resumption tokens; *does not yet support* multiple metadata formats, deleted records, or automatic updating; *does not yet allow* set harvesting.

- **OAI Static Repositories & Gateways** **http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm**
  OAI Static Repositories allow small metadata collections to be made available via the OAI-PMH using a single XML file on a web server. Registering the file with a Static Repository Gateway, such as at UIUC, allows the gateway to intermediate all OAI harvesting requests for that particular metadata collection.

- **OAI FileMakerPro Repository Gateway** **http://cicharvest.grainger.uiuc.edu/fmpgateway/**
  This is a service similar to the OAI Static Repository Gateway that allows FileMakerPro databases which are already accessible over the Web to be harvested using the OAI-PMH. It was developed at the University of Illinois at Urbana-Champaign.

- **OCLC OAICat** **http://alcme.oclc.org/oaicat/**
  Open Source Software repository framework easily customized using Java interfaces. A popular data-provider tool on DSpace.

- **OJS** (Open Journal Systems) **http://www.pkp.ubc.ca/ojs**
  Open-source software developed by the Public Knowledge Project and used primarily for managing and publishing scholarly journals. Metadata associated with articles published on this system may be shared and harvested using OAI. *Supports* multiple metadata formats and *allows* set harvesting.

- **Virginia Tech DLRL Projects** **http://www.dlib.vt.edu/projects/OAI**
  Several useful OAI tools and scripts have arisen from Virginia Tech's OAI work with OAI, including the OAI-PMH File-based Data Provider. A commonly used data provider tool, due to its great flexibility:
  http://www.dlib.vt.edu/projects/OAI/software/xmlfile/xmlfile.html.

▪ **XSLT** (eXtensivle Stylesheet Language Transformation)   **http://www.w3.org/TR/xslt**
A programming language used to reformat information in one XML vocabulary into another.

▪ **ZMARCO**   **http://zmarco.sourceforge.net**
Developed by the University of Illinois at Urbana-Champaign as a way of providing OAI-PMH access to MARC records already accessible through Z39.50 gateways. Free, Open Source software, written in Visual Basic and VBScript, and easily modified.

Digital Library Systems

The following are a selection of digital library systems with strong OAI-data provider capabilities. While more vendors are providing Open Source versions of their tools, the focus here is primarily on non-commercial options.

▪ **ContentDM**   **http://www.contentdm.com**
Useful for displaying images and multimedia. Proprietary software, not Open Source: flexible pricing, based on collection size. [1]
Does not support multiple metadata formats.

▪ **CWIS** (Collection Workflow Integration System)   **http://scout.wisc.edu/Projects/CWIS**
Open-source software created to support the National Science Digital Library (NSDL). Available for free. Allows simultaneous querying and updating of records. Well-adapted to working with smaller collections.

▪ **DLXS** (Digital Library eXtension Service)   **http://www.dlxs.org**
DLXS Broker20, the data provider component for DLXS systems, *allows* set harvesting and resumption tokens, but *does not yet support* multiple metadata formats or deleted records. UMHarvester, the Open-Source harvester component of the DLXS tools package, provides simple and controlled harvesting of metadata, and places harvested records into a browseable directory.

▪ **DSpace**   **http://www.dspace.org**
MIT-developed suite of tools for creating online repositories, including the OCLC-developed data-provider tool OAICat. Runs on LINUX and UNIX systems. Software is downloadable at SourceForge (http://www.sourceforge.net).

▪ **Ex Libris DigiTool**   http://www.exlibrisgroup.com/digitool.htm
A suite of software tools for managing digital resources. S*upports* multiple metadata formats, and *allows* set harvesting and the use of resumption tokens.

▪ **Fedora**   **http://www.fedora.info**
Open-source software developed by the University of Virginia and Cornell University for creating online repositories.

▪ **Greenstone**   **http://www.greenstone.org/cgi-bin/library**
Open-source software, available for free, with a graphical interface for editing metadata. Embedded relational database works well in Greenstone, but with additional programming can be incorporated into other systems. Interoperable with DSpace.

OAI Validation & Harvesting Systems

Validation systems may be used to explore metadata records before they are harvested. Minimally, harvesters gather metadata records so they can be indexed, searched, and browsed. The easier the harvester is to use, and the more compatible it is with current systems, the better. Ideally, harvesters should search sets (rather than entire repositories) and use resumption tokens.

▪ **OAITransform**   **http://www.dlxs.org/products/archive-by-CDROM/11a/Lib/src/bin/o/oaister/oaitransform/OAITransform**
Open-source transformation tool, a component of the DLXS tools package. Converts harvested simple Dublin Core metadata into the DLXS "Bibliographic Class" format. Interoperates with the DLXS Broker.

▪ **Perl O-O Harvester**   **http://www.dlib.vt.edu/projects/OAI/software/harvester/harvester.html**
This Virginia Tech harvesting software is highly recommended as a flexible and dependable harvesting tool.

▪ **PKP Open Archives Harvester**   **http://www.pkp.ubc.ca/pkp-harvester/**
Open-source harvester created by the Public Knowledge Project at The University of British Columbia (Canada), and available for free download through their website. *Does not allow* the use of resumption tokens.

▪ **Reap and OAI Harvester Object Library**   *http://sourceforge.net/project/showfiles.php?group_id=47963&package_id=46165*
The OAI Harvester Object Library is a well-documented COM ActiveX DLL for writing OAI harvesting applications on the Microsoft Windows platform. Reap is a full-featured and robust command-line OAI harvester written using this library.

▪ **Repository Explorer**   **http://www.dlib.vt.edu/projects/OAI/software/explorer/explorer.html**
Commonly used software from Virginia Tech. Allows data providers to check how their data is harvested and service providers to check how well their repositories can be searched.

---

[1] ContentDM is the only system in this list that is not Open Source.

REFERENCES

Foulonneau, M., et al. (2004, November 15). *Getting a Leg Up on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)*. Presentation for the National Science Digital Library Conference, Chicago, IL. http://nsdl.comm.nsdl.org/meeting/session_docs/2004/2620_National_Science_Digital_Library_Conference.doc

Kellogg, D. (2004, August 27). Open source OAI metadata harvesting tools. *California Digital Library* http://www.cdlib.org/inside/projects/preservation/recall/harvesting.pdf

Tennant, R. (2004, November 4). Specifications for metadata processing tools. *California Digital Library* http://www.cdlib.org/inside/projects/harvesting/metadata_tools.htm

Tennant, R. (2004, May 14). Bitter harvest: Problems & suggested solutions for OAI-PMH data & service providers. *California Digital Library* http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html