DLF IMLS OAI Best Practices Project: Training Component

Martin Halbert
Emory University
DLF Fall Forum 2005
Tuesday, 8-Nov-2005
Charlottesville, VA

Goals of Emory Project Component

- Plan the training needed to ease the transition of DLF institutions to the provision of OAI metadata records for harvesting
- Create a series of OAI training modules
- Make all training materials publicly available
- Organize a DLF training program in support of project goals

DLF Digital Library Federation

OPEN ARCHIVES

INSTITUTE of MUSEUM and LIBRARY SERVICES

Preparation and Planning

- Analyzed findings from IMLS project partners experiences with metadata harvesting efforts
- Hosting a planning meeting at the Spring 2005 DLF Forum to gather information from interested DLF institutions concerning OAI training
- Incorporated results of Aquifer surveys conducted in Summer of 2005 concerning metadata and DL services

DLF Digital Library Federation

OPEN ARCHIVES

INSTITUTE of MUSEUM and LIBRARY SERVICES

Requests by Institutions considering OAI Implementation

- Administrative: What are the key reasons to consider implementation? What resources will we need to deploy? Staff? Equipment?
- Technical: What are the most effective strategies for implementing OAI data providers?
- Metadata: What are the appropriate standards for metadata that is to be shared?
- Everyone: Give us clear and brief documents, we don't have time for tomes.

DLF Digital Library Federation

OPEN ARCHIVES

INSTITUTE of MUSEUM and LIBRARY SERVICES

Training Handouts

- Brief documents (1-2 pages each)
- DL technical writer (Liz Milewicz) worked hard to produce clear and concise information
- Each handout is a standalone piece focused on one element of the training program, with see also references to link the series together

Training Series

- The Case for OAI
- OAI Implementation: Administrative Planning
- Summary of Metadata Best Practices
- Summary of the DLF Aquifer MODS Profile
- OAI Tools
- OAI “Cheat Sheet”: Rapid Deployment Strategies
- DLF IMLS OAI Project Summary Sheet
- Workshop Curriculum Sheet

DLF Digital Library Federation

OPEN ARCHIVES

INSTITUTE of MUSEUM and LIBRARY SERVICES

OAI Training Program

- DLF OAI Implementers Workshop after Fall 2005 Forum
- Training Series can be used in many ways by local institutions
- Additional training workshops may be held as part of Aquifer project or future DLF projects

DLF Digital Library Federation

OPEN ARCHIVES

INSTITUTE of MUSEUM and LIBRARY SERVICES

The Distributed Library: OAI for Digital Library Aggregation

OAI "Cheat Sheet": A Taxonomy of Rapid OAI Deployment Strategies

This handout provides a taxonomy of strategies that have been frequently deployed in institutions seeking to implement OAI data providers in the context of commonly used metadata formats and digital library systems. Comments are provided to name each of the strategies.

Effort required: While all these strategies certainly require some coordinated work by digital library professionals, the effort entailed for implementation is typically not onerous for most organizations and reaps many rewards through the resulting capability to exchange metadata by means of a standardized protocol. OAI implementation is often misunderstood as requiring major systems development, which is not at all the case. Major overhauls of existing digital library infrastructures are not necessary. OAI implementations is often more akin to a systems patch, mainly requiring the installation of a few programming scripts or a new module onto existing software. Many effective, simple, and quick solutions currently exist for implementing OAI on top of existing infrastructures — solutions requiring only moderate planning and some staff time to deploy. The following taxonomy maps a number of commonly used systems to solutions, and highlights issues to consider when deciding when or whether to employ these strategies.

SEE ALSO: **OAI Implementation: Administrative Planning** (a guide to work and resource planning) **Summary of OAI Metadata Best Practices** (common metadata formats, quality issues in sharing metadata, and best practices for OAI data and service providers) and **OAI Tools** (technology available for generating, converting, managing, and harvesting metadata)

OAI Deployment Taxonomy

Many scripts and software for enabling OAI implementation can be deployed on top of existing digital library infrastructures. Examples of infrastructures commonly found in digital libraries today are listed below (left column) and mapped to OAI solutions that various institutions have found easily and cost-effectively allowed them to implement OAI. Strengths and limitations of each OAI solution are also provided.

Current Infrastructure	→	OAI Solution
XML files in directory structure This is a common situation in many digital library infrastructures, in which XML files maintained in a structured directory hierarchy are indexed by a search engine for public search and display. This is also one of the most flexible infrastructures, and lends itself to many possible OAI implementations. Staff who maintain this variety of digital library infrastructure frequently have developed a strong expertise in one or more preferred programming languages for data wrangling, and often use XML style-sheet transformation tools. The easiest option for proficient programmers is often to write a simple set of scripts that can respond to the six queries defined by the OAI protocol.	→	**Virginia Tech OAI Scripts** (http://www.dlib.vt.edu/projects/OAI) Pros: This frequently deployed toolkit is comprised of free, open source Perl scripts that can be easily configured for many XML formats (for example, works well with TEI headers). The scripts are flexible and simple to adapt to many infrastructures. Cons: The scripts were developed several years ago and are no longer being actively developed or enhanced. **XSL Transforms (or other customized solution using preferred programming language in use at the institution)** Pros: Most flexible option for those who know their own systems well. A set of correctly functioning OAI data provider scripts can usually be written by a competent programmer in a week or less. Cons: Requires a competent programmer.
Z39.50 gateways Many libraries have Z39.50 gateways for systems which hold digital library item records. These systems may include online catalogs or database systems with Z39.50 gateways such as the popular Zebra open source software. Libraries may wish to share MARC records available in these systems via the OAI-PMH.	→	**ZMARC** (http://sunsite.berkeley.edu/~) Pros: This tool was 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. ZMARC is free open source software that is straightforward to implement. The software is written in Visual Basic and VBScript and is easy to modify if desired. Cons: Correctly parsing MARC records occasionally requires some troubleshooting.

Related DLF projects:

DLF Aquifer
Bringing collections to light through the Digital Library Federation

document authorship:

METASCHOLAR
An Emory University Digital Library Research Initiative

last revised: 7 November 2005
Mlawicz & Halbert - v. 11

The Distributed Library: OAI for Digital Libra

The Case for OAI

This handout reviews (1) the background and development of the OAI protocol, (2) past a benefits of increasing the quantity and quality of shared metadata through OAI implement systems.

Responding to scholarly communities' demands for greater access to pre-publication artic Archives Initiative (OAI) developed an interoperable, scalable protocol for sharing metada Protocol for Metadata Harvesting, also simply referred to as the OAI protocol. Scholars' c awareness of institutional collection holdings, combined with their increased use of online materials, points to a greater need for institutions to make their holdings more easily sear the OAI-PMH offers a cost-effective and low-barrier means for institutional metadata shar demonstrate the benefits to researchers and institutions and the areas for further improve

SEE ALSO: **Project Abstract** (more in-depth background on this IMLS-funded OAI proje

Development of the OAI Protocol

Early demands for electronic repositories of scholarly information came from the sciences sharing and accessing pre-print publications. The Santa Fe Convention of the Open Archi agreements for sharing these publications across institutions through its development of t metadata. As the OAI protocol was being successfully implemented in the scholarly scient how they might expand their own services into this realm. In 2000, the Andrew W. Mellon efforts to develop metadata harvesting services in digital libraries; OAI was selected as a project.

Need & Demand

Current popular search engines still cannot locate resources within searchable databases that contains documents and resources highly valuable to learning communities. As onlin researchers (and sometimes the only stop), resource-sharing institutions must find new w accessible. At the same time, the reality of software and hardware costs and the potential reformatting metadata discourages many institutions from implementing new digital-library new systems must consider the long-term value as well as immediate demands: what offe expenditures, and can also be easily integrated with existing systems?

Benefits

The OAI protocol works with multiple systems and users, and its use of Dublin Core (DC) sharing. Many solutions currently exist for implementing OAI on top of current digital-libra open-source software available online for free download. OAI's interoperability makes it a quickly and inexpensively build their collection online, and helps ensure that learning com institutions can access collection metadata. Exposing richer metadata formats in addition metadata and enhances awareness of collections.

related DLF projects:

DLF Aquifer
Bringing collections to light through the Digital Library Federation

METASCHOLAR
An Emory University Digital Library Research Initiative

last revised: 7 November 2005
Mlawicz - v. 8

INSTITUTE of MUSEUM and LIBRARY SERVICES

OAI for Digital Library Aggregation

Metadata Best Practices

ich may be used with OAI, (2) quality issues that currently limit the increasing the quality of shareable metadata.

ds on high-quality metadata. Simple or unqualified Dublin Core (DC) more descriptive metadata formats may be additionally shared via OAI to ality issues that can limit the reliability and usability of records. AI data providers and service providers will alleviate these concerns.

g metadata formats into shareable OAI), **OAI Tools** (technology available a), and **Summary of the DLF Aquifer MODS Profile** (recommendations gital collections of cultural heritage materials).

a providers can help identify and eliminate problems with metadata. The ics Digital Library and the Digital Library Federation) is developing implementation: http://oai-best.comm.nsdl.org/cgi- delines for improving the quality of shareable metadata.

ials and to provide the most appropriate views of the metadata.

system, find out their preferred metadata formats and best practices.

ection. (For example, technical metadata is usually not necessary – users

consider whether it makes sense to expose all of the metadata ook is made available, it usually doesn't make sense to include a s something unique about each page that merits their inclusion. It makes e book as a whole.)

ng described. (For example, don't try to use all of the elements of a only the top level information to the DC record.)

your institution would be able to identify the resources described.

uch easier for service providers to work with 'imperfect' metadata if the documentation of the decisions and standards used for exposed metadata.

a.

records.

interoperability issues.

related DLF projects:

DLF Aquifer
Bringing collections to light through the Digital Library Federation

document authorship:

METASCHOLAR
An Emory University Digital Library Research Initiative

last revised: 7 November 2005
Mlawicz & Shreeves - v. 13