

IMLS National Leadership Grant LG-02-02-0281
"Proposal for IMLS Collection Registry and Metadata Repository"

This white paper is part of the three-year interim project report for the IMLS Digital Collections & Content Project, summarizing major findings October 2002 through September 2005. Project is hosted at the University of Illinois at Urbana-Champaign. Project Director is Timothy W. Cole (t-cole@uiuc.edu). Full report is available at <http://imlsdcc.grainger.uiuc.edu>. The material in this report is based upon work supported by the Institute of Museum and Library Services under IMLS National Leadership Grant Award No. LG-02-02-0281. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the Institute of Museum and Library Services.

Barriers to Metadata Sharing via the OAI Protocol

A White Paper by Sarah L. Shreeves
IMLS Digital Collections and Content Project
University of Illinois at Urbana-Champaign
September 2005

Introduction

A principal goal of the IMLS Digital Collections and Content (DCC) project is to promote interoperability among digital collections funded through the National Leadership Grant (NLG) program. The primary tools used to achieve this goal were a collection registry and an item level metadata aggregation based on the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The IMLS DCC project offered remote support and guidance to NLG projects in setting up OAI data provider implementations as well as advice on preparing metadata for sharing via the OAI protocol.

As of September 2005, the project had harvested metadata from 22% of the NLG projects with digital collections funded between 1998 and 2004. However, fully 52% of the total NLG projects appear to meet the minimal criteria (i.e. having item-level metadata and a maintained web site) for implementing an OAI data provider but have not. The capability of the remainder of the projects is either unknown (13%) or they do not meet the minimal requirements for an OAI data provider (13%). Many of the projects in the latter category have built a digital 'collection' that consists of multimedia exhibits that do not have item level metadata for their individual components (for example, Southern Utah University's "Voices of the Colorado Plateau"¹). Table 1 categorizes all of the NLG funded digital collections according to their OAI capability.

¹ <http://archive.li.suu.edu/voices/>

Table 1 - 1998-2004 NLG Projects Categorized According to OAI Data Provider Capability (as of September 2005)

Category:	Number (%) of Respondents:
Group 1 – Projects with OAI data provider sites for NLG content	32 (22%)
Group 2 – Projects whose institutions have an OAI implementation (not yet being used for NLG content) and projects that have explicitly expressed plans to add OAI functionality	50 (35%)
Group 3 – Projects whose NLG content meet certain technical criteria, e.g. item-level metadata and a maintained web site	24 (17%)
Group 4 – Projects with no item-level metadata, no interest in providing metadata via OAI, or whose grants are inactive	18 (13%)
Unknown (IMLS DCC staff unable to contact project)	19 (13%)
Total	143 (100%)

So why are over half of the NLG funded digital collections not implementing OAI data provider services when it appears that they could? This white paper outlines a range of barriers to the implementation of OAI data provider services and suggests possible solutions. These can be broadly grouped into three main categories: technical infrastructure, metadata, and organizational challenges. We should note that in many cases organizations faced not a single barrier, but a matrix of interrelated barriers (e.g. OAI was not a priority for the organization and (thus) the organization did not have the technical resources to support an OAI data provider).

The barriers and discussion presented here are based on conversations, emails, and other anecdotal experience gathered through interactions with National Leadership Grant recipients and other representative organizations between 2003 and 2005. This is an unscientific case study (i.e. the information was not gathered through valid survey or interview methods), and the conclusions presented here are not generalizable and should be taken only as the experience of the IMLS DCC project.

Barriers Related to the Technical Infrastructure

While the OAI protocol has been promoted as a ‘low barrier’ means to share metadata, it is not without technical hurdles. Even when digital content management systems include a built-in OAI data provider, staff at a minimum need to know how to turn on and configure the OAI data provider, but also should know how to trouble shoot the data provider and to resolve technical issues such as character encoding problems. When a stand alone OAI data provider system is in place, the technical infrastructure needs are more substantial because of customization and ongoing maintenance issues. Outlined below are several barriers related to the technical infrastructure as well as some solutions for overcoming these.

Technical infrastructure, whether computing or staff resources, is not available for implementation and/or ongoing support. The lack of technical resources is a fundamental barrier to the implementation of OAI data provider services particularly for those institutions which do not have a digital content management system with a built in OAI data provider. This was often the case for smaller institutions whose technical resources had to be directed toward their day-to-day operations, and for whom OAI and interoperability was simply not a priority. In

other cases the institution would have liked to participate, but could not because of the lack of infrastructure – they had no programmers to customize an OAI data provider, could not provide the ongoing support needed, and were not familiar with XML and its attendant issues.

The lack of a technical infrastructure is not an intractable barrier to overcome, but there are a limited number of solutions to offer. The IMLS DCC project did offer resources to help install and customize OAI data provider services, but the ongoing maintenance of these systems – making sure that the data provider is communicating with the database or file system where the metadata is housed, resolving character encoding problems, troubleshooting error messages, communicating with harvesters – is not something that the project could take on for the long term. Offering less robust but still quite workable solutions such as the static repository is another possibility, but does depend on how much the collection changes. Other solutions are to encourage resource developers to use digital content management software that has built in OAI support, although this can be problematic as well (see below). However, convincing institutions to reprioritize how they allocate their technical resources so that OAI data provider services are implemented and maintained is another issue altogether and is discussed further below.

Technical infrastructure will be, is, or just was in transition. There were several cases when the institution was in planning for, in the midst of, or cleaning up after a migration to a new digital content management system. Institutions are understandably reluctant to implement OAI provider services in the midst of a migration particularly if the system they are migrating to is going to provide built in support for OAI.

In the best case scenario this sort of barrier is temporary, and the institution will implement OAI after the migration. However, getting familiar with and working out the knots in the systems takes time and resources, and implementation of OAI is likely to take a backseat to getting the local environment running smoothly. In a worst case scenario, an institution might discover after migration that they cannot implement OAI. One NLG project found, for example, that the promised built in OAI data provider did not work.

Proprietary system in use does not have OAI capability or the available OAI data provider is seriously flawed. The popularity of OAI has meant that many proprietary digital content management systems on the market today – particularly those aimed at the library market – have added built in OAI data provider support. However, many proprietary systems do not; this is particularly true for the museum market. If such systems are not sufficiently transparent or do not provide good support for export of metadata fields, it can be difficult to implement a stand alone OAI data provider.

Several of the proprietary systems with OAI data providers have proven to be seriously flawed – in some cases so much so that either the data provider simply does not work or the institution is reluctant to turn on the data provider for performance or compliance reasons. However, feedback from customers to the vendors has proven to be somewhat effective in moving vendors to correct and upgrade their OAI support. This, of course, takes time as the upgrades have to fit into the product cycle. In addition, the Digital Library Federation, in part with IMLS funding, has also issued a draft of best practices for OAI data provider implementations² which should help provide some benchmarks for functionality support for vendors to work towards.

² <http://oai-best.comm.nsd.org/cgi-bin/wiki.pl?DataProviderPractices>

An investment has already made in the technical infrastructure for another means of sharing metadata. OAI is not the only the way that institutions can share metadata or federate access to their content. Communities with specialized content or resources, such as some scientific communities, have developed other means and standards to sharing metadata and resources within their community. Institutions falling in this category are reluctant to make an additional investment in the technical infrastructure for OAI.

It is questionable whether this is really a barrier, in that the institution is actually sharing the metadata via another means. Institutions in this situation may feel that they are already making an effort to make their metadata available to communities that can use it and that their metadata is not going to be useful in a more general portal (particularly when they will need to ‘dumb down’ their metadata to meet the simple Dublin Core requirement).

Barriers Relating to Metadata

As mentioned in the introduction, the most basic barrier in this category is a lack of item level metadata. However, there are two other major barriers related to uptake of OAI and metadata.

Metadata is not in a shareable state. In order for an OAI data provider to be useful, the metadata should be in a shareable state. This means that the metadata should generally be of good quality and should provide the appropriate context, be consistent, and coherent. Metadata should also be technically shareable in that it should refer to appropriate namespaces and not contain character encoding problems. Poor quality metadata that does not meet these standards can be close to useless for service providers and end users. Additionally, the reluctance of several institutions to share their metadata seemed grounded not only in concerns about the ‘shareability’ of their metadata, but also how their metadata will look next to the metadata from other institutions.

The ‘shareability’ of the metadata does not, of course, directly impact whether or not an OAI data provider can be technically implemented. However, it does impact the true interoperability of a federated system. Planning for ‘shareable’ metadata should be incorporated into the beginning of a project, and not, as seemed to often be the case, tagged onto the end of the project. Promotion and use of best practices, like the Best Practices for Shareable Metadata³ (funded in part by IMLS) and RLG’s Descriptive Metadata Guidelines for Cultural Materials⁴ can also help mitigate metadata issues.

Metadata is too complex to be represented well in simple Dublin Core. The OAI protocol does require that data providers expose metadata in at least simple Dublin Core. Some domains like museums, archives, and some scientific communities, as well as others, often create metadata which is used for far more than simple discovery; the metadata may provide contextual and historical information and record relationships between items. Simple Dublin Core does not have the semantic complexity and richness to express much of what these communities need to express.

³ <http://oai-best.comm.nsd.gov/cgi-bin/wiki.pl?PublicTOC>

⁴ http://www.rlg.org/en/page.php?Page_ID=214

The objections to the Dublin Core requirement are valid, and the OAI community occasionally revisits this decision.⁵ Outside of abandoning the simple DC requirement, there are other ways to mitigate this barrier. The protocol does allow multiple metadata formats to be exposed for each item; thus an item could be exposed in both DC and MODS. The only constraint is that the metadata format must have an accessible XML schema so that harvesters can validate the metadata against it. Encouraging use of standard, community specific metadata formats with published XML schemas could be crucial to increasing the usefulness of the services built on aggregated metadata. A parallel approach is to advance the idea that the metadata exposed via OAI – and in particular, the simple Dublin Core metadata – is but one view of the metadata and cannot and should not represent all of the information that may be in the native format, but should support enough to move a user to the appropriate content.

Barriers Relating to the Institution or Project

Barriers relating to the institution or the project itself are more difficult to specify as they are often more closely tied to the culture or social dynamics with an institution or project staff. These barriers are also often in addition to or intertwined with the technical and metadata barriers enumerated above, which adds a level of complexity to addressing them.

Collaborative projects often have to reach agreement among all partners to expose metadata. For large collaborative projects this can be a time consuming particularly if the participants are less familiar with OAI or have fears about exposing their metadata to a wider audience. This is often even more difficult if the project had not originally planned to implement OAI data provider services and had not considered issues of metadata shareability. In these cases, shareability problems are often compounded by the number of partners contributing.

Getting agreement to implement OAI data provider services in the project planning stage can mitigate this barrier. Services and materials that can demonstrate the value of OAI may also help convince partners to implement OAI data provider services. Promoting metadata ‘shareability’ from day one of a project may also help a collaboration be prepared to share their metadata.

The collection is not yet public. NLG projects wish to wait until they unveil their digital collection in their local environment before sharing the metadata via OAI.

This is less a barrier than a delay in implementation; it makes sense that projects will not want to share their metadata with the larger world until their own collection and web site is launched and ready to use.

Interoperability and OAI in particular is not a priority or is unfamiliar. For a number of projects OAI is simply not a priority. This is particularly true for NLG projects whose funding and support had ended and those which had never included OAI in the project plan. Some institutions, particularly smaller and less technically adept ones, are simply unfamiliar with what OAI is or how it works. Some projects/institutions may be waiting to see whether OAI is part of the long term future of libraries before they adapt their systems.

This barrier is directly related to the value that can be shown in building services around

⁵ The last extended discussion on the OAI Implementers listserv was in August 2003.
<http://oaisrv.nsd.l.cornell.edu/pipermail/oai-implementers/2003-August/thread.html>

metadata aggregated via OAI. This in turn may be reliant on the quality of the metadata that can be harvested. Clearly OAI is slowly becoming an integral piece of the digital library environment, but it has yet to be really mainstreamed into the day to day operations and often remains on the side. Building services based on OAI that show a real impact on users is crucial to building buy-in. Hand in hand with that work, however, must come integration of interoperable standards from the very start of a digital project.