The MOSC Project: Using the OAI-PMH to Bridge Metadata Cultural Differences across Museums, Archives, and Libraries

Eulalia Roel

The MetaScholar Initiative of Emory University Libraries, in collaboration with the Center for the Study of Southern Culture, the Atlanta History Center, and the Georgia Music Hall of Fame, received an Institute of Museum and Library Services grant to develop a new model for library-museum-archives collaboration. This collaboration will broaden access to resources for learning communities through the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The project, titled Music of Social Change (MOSC), will use OAI-PMH as a tool to bridge the widely varying metadata standards and practices across museums, archives, and libraries. This paper will focus specifically on the unique advantages of the use of OAI-PMH to concurrently maximize the exposure of metadata emergent from varying metadata cultures.

The MetaScholar Initiative of Emory University Libraries, in collaboration with the Center for the Study of Southern Culture, the Atlanta History Center, and the Georgia Music Hall of Fame, received an Institute of Museum and Library Services grant to develop a new model for library-museum-archives collaboration to broaden access to resources for learning communities through the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).[1] The collaborators of the project, entitled Music of Social Change (MOSC), are creating a subject-based virtual collection concerning music and musicians associated with social-change movements such as the civil-rights struggle. This paper will specifically focus on the advantages offered by OAI-PMH in amalgamating and serving metadata from these institutional sources that are significantly different in kind.[2]

There has been a great deal of discussion within the library community as to the possibilities OAI-PMH holds for harvesting, aggregating, and then disseminating research metadata. However, in reality, only a few of institutions (be they museum, archives, or libraries) have actually begun to utilize OAI-PMH to this end. There are some practical, historical barriers to implementing any shared system for distributing metadata across institutions that are, more than in degree, different in kind. One of these significant differences is of metadata cultures and practices.

Eulalia Roel (eulalia.roel@gmail.com) is Coordinator of Information Resources at the Federal Reserve, Atlanta.

Libraries have traditionally incrementally assigned metadata at an item level within their collection(s). The strength of this model is that at least a minimal amount of metadata is assigned to a very high percentage of items within the collection. The challenge of such a system is that for such metadata records to interoperate within a shared database and through a common interface (for example, the traditional union catalog), the metadata fields have been quite rigidly defined compared to those within archival and museum environments. Due to tradition as well as the sheer volume of items collected by libraries, metadata at an item level are not greatly detailed or contextualized. Often, items within library collections lack robust relationary mapping to other items within or outside of the collection, as is done, for example, in archival processing.

Content contextualization is highly valued by archival metadata practices and culture as the central tenet of metadata creation. Items at a subcollection level almost always have metadata derivative from and deferential to that of the collection-level metadata. The great benefit of archival practices in metadata assignment is a contextualization of content that reflects the background, the topographic place in time and space of a given portion of a collection and its organic, emergent relationship to the whole. The weaknesses of this model are a great inconsistency in description details and variables (at the collection and subcollection levels), as well as very disparate levels of granularity within the hierarchy of the structure of a collection at which metadata are assigned. Such disparities among institutional types feed an unnecessary level of misunderstanding by libraries of the metadata culture and aims of archives as well as those of museums.

Museums often have very skeletal documented (as opposed to undocumented) metadata about their collections or objects therein. Often museums are not funded to make metadata on their collections freely available. It is common, in fact, for curatorial staff to view metadata as intellectual property to which they serve as gatekeepers, reflecting a professional value placed upon contextualizing materials for users. This is done on a user-by-user or exhibition-by-exhibition basis, depending on user background or the thesis of a given exhibition. Additionally, museums perceive information on the *aboutness* of their collections to be a class of capital with which they can always potentially cost-recover or generate income. Within the culture of museums, staff have traditionally been disinclined to make their collections available in an unmediated manner. Additionally, there has been resistance to documenting information about collections in a systematic way. There is even greater resistance to adhering to any prescriptions on metadata as would be required for compliance with even the most minimally structured database. Such regulation would discriminate

against the nuanced information required for each and every object within a collection.

Why OAI-PMH to Bridge These Cultures?

OAI-PMH was selected by the MOSC project as a means to bridge some of these substantial disparities. The protocol is often mistakenly assumed to function only with metadata expressed as unqualified Dublin Core (DC). In fact, the protocol functions with any metadata format expressed by Extensible Markup Language (XML); this is the minimal requirement for content to serve metadata through OAI-PMH. This includes those formats that have been well received by institutions other than libraries, such as XML Encoded Archival Description (EAD) as it is used in archives. As per 4.2 of the OAI-PMH *Guidelines for Repository Implementers*,

> Communities are able to develop their own collection description XML schemas for use within description . . . elements. If all that is desired is the ability to include an unstructured textual description, then it is recommended that repositories use the Dublin Core description element. Seven existing schemes are: Dublin Core, Encoded Archival Description (EAD), the eprints schema, RSLP collection description schema, UDDI/WSDL, MARC21, and the branding schema.[3]

The OAI protocol has often been partnered with unqualified DC metadata, as this is the most minimal metadata structure necessary for participation in an OAI harvesting system. Not only are these DC fields unqualified, no fields are actually required. No structure or regulations are codified outside of requiring metadata contributors to adhere to this unqualified metadata schema. Therefore, the OAI protocol requires minimal technology support and resources at any given contributing site (such support varying more widely across institutions than even their metadata practices themselves). This maximizes flexibility in metadata contribution, as well as maximizing interoperability between the collective data pool from which a user can search. Granted, this unregulated framework does come at a cost of inconsistency in metadata detail and quality.

However, the great advantage of such nominal requirements is that they enable contributors with minimal metadata-encoding practices to participate in the metadata collaborative. Following is an example of a record as it may appear in the MOSC collection:

```
<record>
<header>
<identifier>oai:atlantahistorycenter.com:10
```

```
</identifier>
<datestamp>2003-03-31</datestamp>
<setSpec>south:blues</setSpec>
<setSpec>south:mississippi-delta-region
</setSpec>
</header>
<metadata>
<oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/
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>Long Hall Recordings</dc:title>
<dc:creator>Morris, William</dc:creator>
<dc:subject>Blues</dc:subject>
<dc:description>..</dc:description>
<dc:description>Comment: sound amateur recording
</dc:description>
<dc:date>2003-05-16</dc:date>
<dc:type>sound recording</dc:type>
<dc:identifier>http://atlantahistorycenter.com/
porcelain/10</dc:identifier>
</oai_dc:dc>
</metadata>
</record>
```

Additionally, with no fields required by the DC schema, institutions can have absolute discretion as to what metadata are exposed if this is a concern (as may be for privacy considerations for archives or for intellectual-property concerns for museums).

However, one of the great strengths of implementing OAI-PMH is that, while the threshold for regulating metadata is low, the protocol can also handle any metadata format expressed by XML, including data formats significantly more structured than DC; for example, EAD, Text Encoding Initiative (TEI), and TEI Lite-defined documents. Scholars are then able to access these scholarly objects via one point, while still being able to collectively access and utilize all metadata objects available in all collections, from the most to the least robust.

The aim of the MOSC project participants in selecting OAI-PMH is to maximize participation from fairly disparate kinds of organizations, with equally disparate kinds of metadata cultures and practices. In comparison to other, currently available methods of metadata aggregation, OAI-PMH is maximally forgiving of discordant metadata suppliers. Thereby, the hope is, metadata contributions are maximized. Concurrently, the protocol

allows for highly robust metadata formats. As the cost for inclusion in aggregated systems, in some cases metadata objects are stripped down. This need is eliminated when OAI-PMH is utilized. The use of the protocol allows for the inclusion of objects consisting of the most skeletal unqualified Dublin Core elements, while still accommodating the most complicated metadata objects. Optimally, this is a means to achieve a critical mass of contributed resources that will enable end users to utilize the MOSC project as the premier site and a primary resource for information on materials about music and musicians associated with social-change movements.

■ Acknowledgment

References

1.   "MetaScholar: An Emory University Digital Library Research Initiative," Emory University Libraries Web site. Accessed Sept. 1, 2004, http://metascholar.org/; "The Center for Southern Culture," University of Mississippi Web site. Accessed Sept. 1, 2004, www.olemiss.edu/depts/south/; "Atlanta History Center," Atlanta History Center Web site. Accessed Sept. 1, 2004, www.atlantahistorycenter.com/; "Georgia Music Hall of Fame," Georgia Music Hall of Fame Web site. Accessed Sept. 1, 2004, www.gamusichall.com/home.html; "Institute of Museum and Library Services: Library-Museum Collaboration," Institute of Museum and Library Services Web site. Accessed Sept. 1, 2004, www.imls.gov/grants/l-m/index.htm.

2.   "Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting," Open Archives Initiative Web site. Accessed Sept. 1, 2004, www.openarchives.org/ OAI/openarchivesprotocol.html#Introduction.

3.   "4.2 Collection and Set Descriptions," Open Archives Initiative Web site. Accessed Sept. 1, 2004, www.openarchives.org/ OAI/2.0/guidelines-repository.htm#setDescription.