

Cultivating Learning Communities through Metadata Harvesting

A model for implementing collaborative OAI
systems in museums and libraries

November 2004

Version 0.7

Table of Contents

Overview.....	1
Section 1: Benefits of Collaborative OAI Systems.....	2
1.1 Roles and Responsibilities of Museums.....	2
1.2 Roles and Responsibilities of Libraries.....	2
1.3 Museum-Library Collaboration and Benefits to Learning Communities.....	2
1.4 Scenarios: Museum-Library and Museum-Museum Collaborations.....	3
1.4.1 Museum-Library Collaboration.....	3
1.4.2 Museum-Museum Collaboration.....	3
Section 2: Metadata Practices in Libraries and Museums.....	4
2.1 Museum Metadata Practices.....	4
2.2 Library Metadata Practices.....	4
2.3 Bridging Metadata Gaps with OAI-PMH.....	4
2.3.1 Creating Standardized Field Tags with Dublin Core.....	5
2.3.2 Isolating and Deselecting Sensitive Fields.....	5
Section 3: Process Model for Planning Collaborative Projects.....	6
3.1 Contact Phase.....	6
3.2 Information-Gathering Phase.....	6
3.3 Design Phase.....	6
3.4 Implementation Phase.....	6
3.5 Cataloging/Enriching Metadata Phase.....	6
3.6 Harvesting Phase.....	7
3.7 Evaluation Phase.....	7
Section 4: Roles and Responsibilities Key to Successful Collaborations.....	8
4.1 Essential Roles and Responsibilities.....	8
4.1.1 Host.....	8
4.1.2 Cataloger.....	8
4.1.3 Catalog-Content Provider.....	8
4.2 Supplementary Roles and Responsibilities.....	9
4.2.1 Context Provider.....	9
4.2.2 Research Liaison.....	9
4.2.3 On-Line Exhibition Director.....	9
4.2.4 Media Manager.....	10
Section 5: Technical Implementation and Software Resources.....	11
5.1 Migrating, Gardening, and Harvesting Metadata.....	11
5.1.1 Phase 1: Upload Records into Conversion Software.....	11
5.1.2 Phase 2: Transfer Dublin Core XML Records into Cataloging Software.....	11
5.1.3 Phase 3: Enhance the Catalog Records.....	11
5.1.4 Phase 4: Verify the Catalog Records.....	11
5.1.5 Phase 5: Upload the Records for Harvesting.....	11
5.2 Software Resources.....	12
5.2.1 Metadata Migrating Tool.....	12
5.2.2 MySQL.....	Error! Bookmark not defined.
5.2.3 Greenstone.....	12

Overview

Scenario: A small research and educational center in the 1960s, facing antagonism and imminent shutdown for its support of the civil rights movement, sends portions of its historical collection to libraries and archives across the country in order to protect them. Today, this collection is still divided among several institutions, and no catalog is available to researchers that can provide searching of the entire collection. Though the collection will likely remain divided, collaboration among these institutions using the OAI-PMH could not only provide an interoperable public database of the collection but could also make digital copies of some collection materials available online.

Scenario: An American museum, with a strong collection focus on Southeast Asian art, wishes to build a collaborative relationship with a similarly focused non-U.S. museum as part of the International Partnerships Among Museums (IPAM) program. Both museums agree that, as a way of extending their educational missions, they would like to create a publicly accessible catalog of their collections – one which scholars could use to identify items for study, and which the museums themselves could use for planning collaborative exhibitions and cultural exchange programs. However, both museums have already invested considerable time and money into their own collection management systems, which are, unfortunately, not interoperable. A collaborative OAI system could provide a cost-effective, scalable alternative to buying and implementing new software, and ensure that in the future other institutions may easily join this collaborative project.

The following collaborative model details how museums can collaborate with libraries and other institutions, by means of the OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). This model, developed by the Music of Social Change central project staff at Emory University, addresses these aspects of collaboration using OAI systems:

1. scenarios and benefits of implementing collaborative OAI systems in museums and libraries
2. museum metadata practices, and how they relate to libraries and the OAI protocol
3. process model for planning collaborative projects with libraries and other institutions
4. roles and responsibilities key to successful collaborations
5. technical implementation and software resources

Ideally, this model should guide museums through the collaborative process, providing them with the tools and solutions needed for successful partnerships with libraries, as well as with an outline that can be used in the creation of new projects and the construction of project proposals and grant applications. Though the model focuses primarily on “museums” as collaborators, a range of educational institutions (such as historical societies) are implicitly considered in this process and encouraged to place themselves within such collaborative endeavors.

Section 1: Benefits of Collaborative OAI Systems

As the opening scenarios demonstrate, museums stand to benefit considerably from collaborative OAI systems. Collaboration between museums and libraries, in particular, extends the educational missions of both institutions by consolidating their efforts and building a shared, interoperable resource. By establishing a common metadata-sharing system, these institutions also improve their ability to collaborate on future projects.

1.1 Roles and Responsibilities of Museums

Museums' roles in community life are multi-faceted and demanding. Museums develop and maintain collections, create public exhibitions, and coordinate events and programs as part of their general mission to collect, conserve, educate, and exhibit. In some cases, the demands of collection management and preservation may compete with museums' educational missions to affiliated research groups that act in scholarly contexts. Collection management may itself be frustrated by time and money, when the budget provides little room for investing in a comprehensive management system and when time constraints limit staff's ability to routinely update records or verify their accuracy. Privacy concerns and ethical considerations regarding accession-list records often mean that very few individuals, sometimes only the museum's registrar, have access to these records. Yet some metadata in these records are self-evident descriptions of the object itself, such as title, artist/creator, medium, dimensions, and so forth – information that is not sensitive and which could be quite useful to learning communities and to the museum staff who serve them.

An OAI system may be able to simultaneously benefit both practical research activities that come into play for scholars and inventory considerations that come into play for curatorial staff. Through this system, museums can move publicly useful information about collection items to an online database, while keeping the full, sensitive records intact in their own separate and restricted accession lists. As museums come under increasing pressure to remain commercially viable, collaborative alliances and strategies such as the OAI-based model proposed in this project can assist museums in cooperatively working with consortia for the benefit of learning communities, and extending the benefits of work done in one consortium to other groups.

1.2 Roles and Responsibilities of Libraries

While libraries may engage in small-scale exhibits or the collection and preservation of rare materials through affiliated archives, their primary focus is on increasing public access to information. Through the use of standardized cataloging procedures and authority control mechanisms, libraries facilitate user access to their collections and inter-institutional sharing of information and materials. Libraries provide their learning communities with materials on-site or connect them with outside resources. As portals to other sources of information, search tools (like citation indexes, abstract-only databases, research guides, and consortia catalogs) indicate sources of more specific or specialized content, while librarians themselves may use these tools to refer patrons to persons, organizations, or institutions that specialize in a particular interest area.

1.3 Museum-Library Collaboration and Benefits to Learning Communities

Learning communities benefit from collaboration between museums and libraries. Shared access to general information about their collections increases both institutions' abilities to inform their publics and plan collaborative projects. Libraries and museums assist research communities, particularly scholars in fields heavily reliant on primary resources, by sharing aggregated and electronically stored metadata on their collections.

When it is permissible for the collection items to be digitized and placed on-line, more opportunities are presented for supporting scholarly research. Because museum databases will be able to interoperate with library catalogs, automated aggregation of records from disparate archives can be used to create virtual on-line collections. These collections of topically related items held at multiple institutions could be assembled for

ad hoc purposes (i.e., subject portals for targeted learning communities, supplementary on-line exhibits of broader collections, and union catalogs of consortium holdings). Museum-library collaborators will also have freely available software to support the construction of subject-specific virtual collections.

1.4 Collaboration Scenarios

Minimally, implementing an OAI system creates searchable records of an institution's holdings – information that can be shared with other institutions or simply used internally. Maximally, collaborative OAI systems increase the researching capabilities of staff and public alike, and provide a powerful resource for building further collaborative endeavors. The following scenarios highlight both the utility of museum-library collaborations and some of the basic benefits of using OAI systems: migration of records into harvestable metadata, the hosting and sharing of databases, and the interoperability of diverse collection management systems. They also begin to detail the roles and processes involved in establishing collaborative relationships.

1.4.1 Museum-Library Collaboration

A library uses its server space to host a small historical museum's OAI-formatted records in its online catalog. Searches of this shared database retrieve primary-source records (of museum holdings) along with secondary- and tertiary-source records (of library holdings), thus highlighting contextual and conceptual links between the collections. Educational programs within each institution (i.e., school tours of the museum's Civil War collection, and library-led classes on genealogical research) lead to cross-fertilization and extended use of resources (i.e., the school group discovers biographies and auto-biographies of Civil War heroes in the library, and students in the library class explore the museum's military records for their Southern ancestors).

1.4.2 Museum-Museum Collaboration

A museum consortium uses an OAI system to convert their existing database records into a sharable and harvestable format. Each museum in the consortium creates publicly available and searchable metadata on its collections by individually converting the accession-list records into Dublin Core XML records. At the same time, sensitive information from each record is suppressed, and each museum continues to maintain its own internal accession-list. The now harvestable metadata descriptions of the collection, uploaded into a shared database, are used by the consortium to plan loans, exhibitions, and educational programs. Scholars and school groups can refer to the catalog when planning educational trips, and visitors can use it to locate items of interest in all museums.

Section 2: Metadata Practices in Libraries and Museums

An initial challenge to museum-library information sharing lies in these institutions' metadata practices. Though libraries and museums both document and describe items in their collections, they approach this task with different objectives in mind, and often with different results. OAI-PMH makes possible inter-institutional sharing of metadata, without compromising each institution's practices.

2.1 Museum Metadata Practices

Museums very often possess unique objects, for which no previously recorded metadata information is available, let alone downloadable. Additionally, because these objects rarely circulate, metadata descriptions of museum items (or accession lists) are not created with a searching-and-retrieving public in mind. Instead accession lists function internally, as descriptions that contextualize the object's relationship to the collection and that document sensitive donor and provenance information. Often collection management is the responsibility of a single registrar, who maintains the accuracy and security of the records. Where library records are typically uniform and brief, museum records may be quite extensive and unique, documenting much information that other museum staff may lack the ability to interpret. Some museums, such as art museums, can rely on standard thesauri for uniform descriptors; in many cases, though, the more specialized the collection's focus, the less likely it is that the museum can rely on a standard tool for describing the collection. Thus, museum metadata practices – particularly in small museums – tend to emphasize uniqueness and context in order to provide local curators with a record of the collection's scope, depth, and value, and tend to highly restrict access to records. Consequently, inter-institutional sharing of records is rare, if it occurs at all.

2.2 Library Metadata Practices

Cataloging in libraries, on the other hand, is geared towards future retrieval and use of the object by the public or the library's patron population. To this end, libraries create metadata records of items in their collections and make those records publicly accessible. Both libraries and library archives employ standard cataloging practices for recording metadata about an object, whether that object is mass-produced and widely available or unique and rare. The culture of these institutions emphasizes uniformity, brevity, and standardization, in the service of maximizing the public's search and retrieval of metadata records and, ultimately, the objects they define. Yet inter-institutional sharing of information is still difficult: cataloging formats and practices differ between libraries and archives, and databases and technical systems differ between institutions.

2.3 Bridging Metadata Gaps with OAI-PMH

Despite these generic differences in metadata practice, museums and libraries are historically committed to their educational mission of serving learning communities. Museums provide breadth and experiential depth in focused subject areas, through their expansive supply of primary resources and their experience in presenting those resources in a comprehensive and contextual way. Though libraries may also possess some primary resources, their strength typically lies in collecting, cataloging, and providing public access to secondary and tertiary resources, and in helping scholars and researchers navigate diverse metadata collections (such as databases and other library catalogs) to find the information they need. As the collaboration scenarios above help illustrate, these specialized and complementary services can be interwoven to maximize collaborative institutions' potential to educate the public and support scholarship and to expand the reach of both museum and library collections and services. And importantly, the security of sensitive accession-list information will not be sacrificed in the process. One of the first steps in that collaborative process is bridging the gaps in metadata practices through the use of OAI protocol.

2.3.1 Creating Standardized Field Tags with Dublin Core

The migration of local database records into Dublin Core XML records is a pivotal step in bridging metadata gaps. Dublin Core tags provide a standard and flexible format for organizing metadata. Replacing locally specific field labels with relevant Dublin Core tags makes it possible for the information to be reliably searched and harvested online. Migration software, such as Emory University's Metadata Migrating Tool, allows users to map their current field labels to one of sixteen metadata elements. Because the OAI protocol is application-independent, institutions may use it in conjunction with their current automated systems.

2.3.2 Isolating and Deselecting Sensitive Fields

Because a primary concern for museums is the protection of certain accession-list metadata, an important step in metadata sharing is the isolation and selection of which metadata to display and which to suppress. The Metadata Migrating Tool, developed and available through Emory University, allows registrars and other collection-management personnel to deselect specific fields from the conversion process. The rest are mapped to Dublin Core tags, then uploaded online where they can be searched and harvested by OAI search engines. The museum's original accession-list records remain intact within the museum's collection-management system, while the converted records are saved as separate files that can be further edited in-house or transferred to a collaborator.

Section 3: Process Model for Planning Collaborative Projects

The following process model defines procedures for establishing collaborative partnerships. This model may serve as an organizing tool for institutions as they design and implement an OAI-compliant gateway to their collections. It may also provide a conceptual assembly line or stream that can be easily understood by all involved parties. This process model includes seven phases: 1) Contact, 2) Information Gathering, 3) Design, 4) Implementation, 5) Cataloging/Enriching Metadata, 6) Harvesting, and 7) Evaluation.

3.1 Contact Phase

- A central project staff member makes an initial contact with an institution to provide them with information about the collaborative project, the project's timeline, and the role of a collaborating institution.

3.2 Information-Gathering Phase

- Several members of the central project staff make an initial site visit to gather information from the collaborating institution about their collections, applications, infrastructure, and native metadata formats that will enable metadata harvesting using the OAI protocol.
- A central project staff member helps the potential collaborator to determine whether or not the institution holds relevant materials in a system structure that can be made OAI compliant.
- If the institution expresses interest in sharing metadata, but does not have adequate records based in an OAI compliant system, the central project staff potentially offers the institution cataloguing services, depending on available resources.

3.3 Design Phase

- The central project staff works with the institution to develop a description and workflow of the current process for metadata creation at the collaborating institution.
- The central project staff assists the collaborating institution with the design of software to facilitate serving of existing metadata via an OAI protocol, based on findings in the information-gathering phase.

3.4 Implementation Phase

- The central project staff assists the collaborating institution with the implementation of OAI-compliant gateways to local systems (where necessary).
- The collaborating institution migrates its existing metadata into OAI-formatted (Dublin Core XML) records, deselecting any metadata information that is non-public or sensitive prior to conversion.

3.5 Cataloging/Enriching Metadata Phase

- If the institution's process for metadata creation produces records that do not contain enough information to be of use to researchers when mapped into Dublin Core record formats, a cataloger develops a process model for enriching the metadata records held by the collaborating institution.

- The cataloger and the collaborating institution work together to enrich the record (or “garden metadata”), sharing information until both the institution and the central project staff members are satisfied with the quality of information that will be served through an OAI protocol.
- The collaborating institution verifies their metadata’s representation in the system.

3.6 Harvesting Phase

- The OAI system harvests the collaborating institution’s cataloged/enriched metadata into a central repository and makes it available for harvest by other systems.

3.7 Evaluation Phase

- The collaborating institution completes an on-line survey to assess the impact of the project.

Section 4: Collaboration Roles and Responsibilities

Because libraries and museums bring unique educational abilities and resources to the table, their collaborative roles and responsibilities should reflect these assets and put them to constructive and mutually beneficial use. The following list identifies important collaborative roles and the responsibilities they entail, and suggests which institution may best serve those needs. Though geared primarily toward library-museum collaborations, these roles and responsibilities may be applied to other collaborative relationships: for instance, a consortium of small museums might assign a separate role to each cooperating institution.

4.1 Essential Roles and Responsibilities

4.1.1 Host

Smaller institutions like subject-specific museums may require server support for their converted metadata records. In such instances, libraries may provide server space for literally “hosting” their information. Systems departments or technology support in these libraries may also assist in the maintenance of metadata collections.

Ideally, the Host should:

- Provide server space for and maintain the shared database of library and museum metadata records, in addition to any on-line exhibitions and digital collection materials.
- Routinely harvest new records, migrating into Dublin Core XML if necessary.
- Work with the Catalog-Content Provider to ensure software compatibility and provide technical support.

4.1.2 Cataloger

A library or larger institution’s cataloging resources may assist in the creation or reconstruction of metadata records. It is particularly important for the Cataloger to work closely with the collaborating institution, to ensure that records provide accurate descriptions and sufficient information to be of value to researchers. At the same time, the Cataloger acts as an additional check on the information, to ensure that such internally significant and private information (for instance, the donor field) is kept out of the metadata record.

Ideally, the Cataloger should:

- Edit metadata records before uploading to the database.
- Work with the Catalog-Content Provider to clarify the type of metadata information needed to adequately describe the item.

4.1.3 Catalog-Content Provider

The Catalog-Content Provider is the member of the museum staff responsible for creating the shared metadata records. This person provides the necessary item descriptions for metadata records, supplying additional information to the Cataloger when necessary to make the record useful to researchers. This person works with the Cataloger to determine which fields are important to researchers and which fields can or should be excluded from the catalog.

Ideally, the Catalog-Content Provider should:

- Work with the Host to create metadata records within OAI-compliant systems which can be uploaded to the server.
- Work with the Cataloger to clarify the type of metadata information needed to adequately describe collection items.

4.2 Supplementary Roles and Responsibilities

For collaborative projects involving digitization of collection materials (i.e., scanning photographs or documents), the following additional roles and responsibilities will ensure that the on-line collection meets the demands of quality, contextuality, and accessibility.

4.2.1 Context Provider

The Context Provider, likely a museum curator, supplies general information about collection items and the collection as a whole. This person provides descriptions that place the collection within a larger historical and topical context and/or that indicate how items within the collection relate to each other. The Context Provider also works with the Research Liaison to collect brief scholarly articles that further extend this contextual base.

Ideally, the Context Provider should:

- Provide contextual descriptions of the collection that can be used by the Research Liaison, for promotional publications, and within the on-line portal that houses the shared database collection.
- Work with the Research Liaison to collect scholarship that contextualizes the collection and/or individual collection items.

4.2.2 Research Liaison

The primary goal of library-museum collaboration is to better serve the learning community and thus further the educational missions of both institutions. A Research Liaison provides a consistent contact for this community, someone who can field information requests and schedule museum visits. If the museum provides researchers with access to materials, this liaison may also serve as an intermediary, who works to meet the scholar's research needs while helping the museum to preserve and protect its collection. Because the liaison works with researchers at the museum, this role will likely fall to a member of the museum staff.

Ideally, the Research Liaison should:

- Handle requests from and provide information to the learning community.
- Balance researcher requests with the museum's security and preservation concerns.
- Work with the Context Provider to collect scholarship that contextualizes the collection and/or individual collection items.

4.2.3 On-Line Exhibition Director

A major asset of museums is their ability to organize and present collection materials in meaningful ways. Museums can increase the educational reach of these exhibitions and minimize handling of collection items by creating on-line digital displays. The On-line Exhibition Director selects which collection materials to digitally reproduce then arranges those materials into on-line displays. Once constructed, these displays may be archived on the site (as part of the "permanent" on-line collection) or rotated out, and materials may be easily re-organized to demonstrate other contextual links.

Ideally, the On-Line Exhibition Director should:

- Work with the Media Manager to create high-quality digital copies of collection materials.
- Work with the Media Manager and the Host to create on-line exhibitions.
- Provide contextualization for on-line exhibition materials.
- Work with the Research Liaison to develop exhibitions that meet the needs of the learning community.

4.2.4 Media Manager

Digitizing museum materials helps to maximize exhibition potential and minimize handling of the physical objects. Digital copies must be of sufficient quality to be of benefit to researchers. Additionally, useful on-line exhibitions of these materials must attend to typical download times and the software needed to access these materials, among other concerns. The Media Manager works with the On-Line Exhibition Director and the Host to ensure that on-line exhibitions are accessible to the public and supported by the server.

Ideally, the Media Manager should:

- Work with the On-Line Exhibition Director to create high-quality digital copies of collection materials.
- Work with the On-Line Exhibition Director and the Host to create on-line exhibitions.

Section 5: Technical Implementation and Software

This section provides an overview of how the OAI system is implemented, including specific instructions on how to use the Metadata Migrating Tool, a free, open-source software program that moves local metadata into harvestable Dublin Core records. It also provides information and links to software programs used in this process.

5.1 Migrating, Gardening, and Harvesting Metadata

Converting accession-list records into uploadable OAI-compliant records is typically a one-step process, provided that the metadata records are already in a machine-readable form. If the museum or archive accession lists are already available as Dublin Core XML records, they may be harvested directly by the Greenstone cataloging software (phase 2). Additional phases allow records to be further checked and modified before being uploaded into a shared database.

5.1.1 Phase 1: Upload Records into Conversion Software

The Metadata Migrating Tool, developed at Emory University, converts accession-list records into XML records that use Dublin Core tags, and saves them as a separate file. It also provides the option of de-selecting sensitive fields, which are then excluded from the conversion. This tool can convert any of the following types of file formats: comma-separated value (csv), tab delimited (tab) and dBASE (dbf). The user simply chooses which file to convert, indicates which Dublin Core element should be mapped onto each original field label, and then selects which fields to convert. Records are converted into harvestable Dublin Core XML files, which can be further modified using cataloging software.

5.1.2 Phase 2: Transfer Dublin Core XML Records into Cataloging Software

The Dublin Core XML records must be transferred into a cataloging software program and enhanced before they can be searched through a shared catalog or database. Additional checks and changes may be needed to ensure that the desired metadata is presented in the record. Greenstone software is recommended for amending the records into a more local-catalog-friendly format.

5.1.3 Phase 3: Enhance the Catalog Records

During this phase records are checked and modified to provide sufficient information for researchers. A “Librarian Interface” is available for use with Greenstone cataloging software, as a separate application run from the desktop. Records may be edited in batches from a shared file.

5.1.4 Phase 4: Verify the Catalog Records

Catalogers and collaborating institutions check the cataloged records to ensure that they contain the desired metadata.

5.1.5 Phase 5: Upload the Records for Harvesting

Verified records are uploaded into a shared database, where they can be “harvested” by users.

5.2 Software Resources

5.2.1 Metadata Migrating Tool

Developed at Emory University, the Metadata Migrating Tool is free, open-source software for converting accession-list records into Dublin Core XML records. For more information and to access the software, visit <http://metacluster.library.emory.edu/mosc/upload.php>.

5.2.3 Greenstone

Greenstone Digital Library Software supports the development and publication of metadata records on-line. A suite of open-source software products addresses the particular needs of collaborating institutions. For more information, visit <http://www.greenstone.org/cgi-bin/library>.