



Sharing with Other Organizations

Grace Agnew
Rutgers University Libraries

One of the reasons libraries, archives and museums create digital collections is to participate in a digital information commons, where the organization's digital collection joins with others to create a mega-collection with breadth and depth. Creating a digital collection remains a complex undertaking, with many unknowns and opportunities to discover better workflows or otherwise contribute to expanding arena of knowledge that comprises digital librarianship.

Sharing your collection with others is both a technical and a social/political endeavor. Each dimension of sharing is discussed in turn.

Technical Aspects of Sharing with Other Organizations

Recognizing that no library, museum or archive can own everything, considerable effort has always been placed on making the broad range of information resources available to users, regardless of whether the organization owns it. "We don't have it, but I can get it for you," has been a common statement of service within libraries, on par with, "I don't know the answer, but I can find out." Interlibrary loan has been a major library service since the early days of the lending library. The importance of lending

from one library or archive to another is explicitly recognized in Section 108 of the U.S. Copyright Code.

From the beginnings of digital library automation, considerable focus has been placed on the development of digital standards that would enable organizations to share information, to achieve economies of scale in acquiring and cataloging resources and to make it possible for their users to have any information they needed, regardless of who owns or provides it.

Standards are very important for the following reasons:

- They provide consistency for the tasks of creating or sharing information.
- They provide semantic consistency for the audience, who may be viewing resources, or metadata, from a wide range of contributing organizations.
- They provide economies of scale and automated workflow for activities such as sending, merging and searching metadata, as well as retrieving digital objects.
- They codify best practices in an emerging field so that everyone can create and share digital resources in an effective and persistent manner and so that issues such as format migration and emerging formats can leverage the expertise of many contributors to the benefit of all digital library practitioners.

True standards are distinguished from common practice in a variety of ways. A standard should be:

- Rigorous in its design, to insure consistency of application across different types of organizations and across space and time.
- Maintained by a standards body, who commits to an accepted practice of design, testing, public comments, review, documentation and versioning.
- Offering documentation, training and a body of practitioners who can share best practices.
- Offering some level of community customization that supports interoperability for the core standard with extensibility for local or community needs.

- Offering both a forward and backward migration path for new versions of the standard, particularly those that incorporate a significant upgrade or redesign.
- International in scope, or able to integrate with international standards. A standard should be developed with support for internationalization elements, such as Unicode support and multi-linguality, or at the very least able to be re-engineered for internationalization.

There are several important categories of standards that an organization wishing to share its information with others in a seamless, automated way will wish to consider and implement:

1. **Content creation standards.** Content creation standards are used to create digital objects and metadata. These standards are intended to create persistent digital resources that will be useful for many generations of students and scholars and also to create a consistent body of work that can be migrated to new formats or utilize new display technologies, without requiring that hardware and software vendors support an unlimited array of formats. Metadata standards provide consistency of creation, with the ability to use shared tools, such as name and subject authority files, consistency of indexing across shared portals and search applications, and consistency of interpretation for end users, regardless of which organization owns and cataloged the resource.

An organization creating digital resources should digitize resources to prevailing digital standards and best practices. A good place to start is to examine the minimum and recommended standards for digitizing resources for inclusion in a collaborative portal that your collection would be likely to join. In many cases, this will be a statewide or regional digital collections initiative, such as the *New Jersey Digital Highway*. These initiatives will publish minimum and recommended digitization standards on their websites, under sections for content contributors or librarians, archivists and curators. National consortia, such as the American Library Association, Research Libraries Group, and Society of American Archivists, will also publish recommended standards, often in a “best practices” document. Standards bodies, such as the National Information Standards Organization (NISO) will also provide best practices recommendations, such as the *Framework of Guidance for Creating Good Digital Collections*, which references standards for digital master files. Granting agencies may provide guidance or refer you to someone who can provide guidance, as part of the grant preparation process. The comments from reviewers of an unsuccessful grant proposal can also be an excellent way of obtaining guidance on your technology design for your project.

Examples of digital object creation standards include, TIF, WAV, MPEG-2, M-JPEG 2000, etc.

One or more metadata schemas should be selected for describing the resources in your collection, rather than designing your own schema. A metadata schema should be very robustly designed. Among other things, an XML representation or “binding” of the metadata schema should be available, so that records created can be validated against the schema, so considerable knowledge of XML is required to create your own schema. Your own schema must also be robustly designed, with controlled vocabularies and formatting principles governing every data element, except the free text data elements. Comprehensive documentation for users creating metadata with your schema is also required. Finally, your schema must map readily to standard schemas, as a data display and export option, for sharing data with others. It is much safer, in terms of internal workflow, consistency for end users, and sharing with other organizations, to utilize an existing schema, rather than committing to the level of effort and expertise required to design your own. Again, the first place to begin is with the consortium that might represent your first place to share your metadata and objects. In many cases, the consortium will provide a metadata schema and tools for creating metadata to share with the consortium. The New Jersey Digital Highway offers a web-based cataloging utility, the Workflow Management System or WMS, for use by participating organizations in the New Jersey Digital Highway. In a future release of the WMS, participating organizations will be able to export records in a variety of schemas and formats for local re-use, so that the WMS becomes a fully functioning cataloging utility for creating metadata about digital objects.

Common metadata standards include ISBD within MARC, Dublin Core, MODS and PREMIS.

2. Container and transport standards for data

Another important category of standard are container and transport standards which (at least in theory!) enable data to be packaged and transported from one repository or digital library application to another. Some common container and transport standards include:

- **METS** (Metadata Encoding and Transmission Standard). METS is a metadata envelope or wrapper that enables an organization to utilize multiple metadata schemas to provide descriptive, technical, and rights metadata, as well as a structure map for navigating complex objects. METS also supports associating behaviors (programming scripts or applications) with digital objects and binds all digital versions of the object with the metadata, structure map and behaviors. The

METS document, consisting of all these component parts, is a standardized envelope that can be transported from one METS-compliant repository or digital library application to another. The Fedora open repository architecture forms the technical infrastructure for *New Jersey Digital Highway*. Fedora is a METS-compliant repository architecture, so NJDH's event-based data model is implemented and expressed as a METS document.

More information about METS:

<http://www.loc.gov/standards/mets/>

- **Dublin Core.** Dublin Core is a metadata schema but also a mapping utility that enables metadata schemas to be mapped to a common "core" standard for search, retrieval and display of metadata information. Dublin Core (ISO 15836) consists of 15 optional, repeatable data elements. Dublin Core is required for support of the OAI-PMH metadata harvesting standard and is often required for many information portals, such as the NSF-funded *National Science Digital Library*. To insure metadata interoperability, an organization should be able to export its data in Dublin Core for sharing with a wide variety of distributed or consortial initiatives.

More information about Dublin Core: <http://dublincore.org/>

- **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)**

The OAI-PMH protocol provides a fairly simple way to harvest large amounts of metadata in an automated manner from a metadata catalog. OAI-PMH uses simple http commands to search for and retrieve metadata based on some core searching attributes: owning organization, set (or collection) and date of creation or update for the metadata record. Export of metadata records in Dublin Core is required for OAI-PMH. Profiles for other metadata schemas, such as MARC, also exist.

The New Jersey Digital Highway's Fedora repository architecture includes an OAI-PMH gateway for sharing NJDH records, or the metadata records of any participating institution, with other digital initiatives, repositories and collections

More information about OAI-PMH:

<http://www.openarchives.org/OAI/openarchivesprotocol.html>

- **OpenURL Framework for Context-Sensitive Services**

OpenURL (ANSI/NISO Z39.88-2004) is a standard for providing links between an information resource and a linking service, which provides access to the most appropriate version of a resource, often a locally held full-text database that requires a subscription for access. OpenURL can have increasing utility in the digital library arena as different digital versions of a title or artifact begin to appear.

Use of OpenURL can enable a digital collection to participate in Google Scholar, and other ways of surfacing information to a wide range of users via the Web.

More information about OpenURL:

http://www.niso.org/standards/standard_detail.cfm?std_id=783

- **NISO Metasearch Initiative**

The NISO Metasearch Initiative was convened to develop a standard or suite of standards to enable interoperable metasearching, or search and retrieval across multiple databases. Three task groups addressed the three components of a suite of tools needed to accomplish interoperable, standards-based metasearching: TG1 – Access Management; TG2 – Collection and Service Description; and TG3 – Search and Retrieval. The collection description is based on Dublin Core Collection Application Profile, the Service Description is based on SRU/SRW Explain. Search and Retrieval, the heart of the standards suite, requires support for SRU (Search/Retrieve Operation) and SRW (Search/Retrieve Web Service). SRU/SRW are essentially Z39.50 distributed search and retrieval expressed in XML in a web-friendly syntax. The NISO Metasearch Initiative is bringing standardization to distributed searching, retrieval and display for metadata and web documents, across many catalogs and other digital environments.

More information about NISO-MI:

http://www.lib.ncsu.edu/niso-mi/index.php/Main_Page

- **Registries**

Registries provide a standardized way of documenting the characteristics of information, whether the data elements that comprise a metadata schema or the digital objects that comprise a collection. A registry provides valuable information about standardized formatting for information that can be used by a computer application to retrieve and interpret information. The ISO/IEC 11179 registry standard provides the basic elements for describing information, including the registration authority, controlled

vocabularies, etc. You should document your use of a metadata schema, with the controlled vocabularies you select, in a registry.

A current, useful registry for digital library initiatives is the Registry of Digital Masters—a trusted service for the communication, coordination and discovery of information about digital masters, jointly offered by the Digital Library Federation and OCLC

DLF/RLG Registry of Digital Masters:

<http://purl.oclc.org/DLF/collections/req/OCLCservice>

When designing a digital project or collection, it is critical to think about issues of digital interoperability. Standards that you implement should be based on, or interoperate smoothly with, other digital collection providers with whom you are likely to collaborate or that also serve your key target audience.

Social Aspects of Interoperability

The social aspects of interoperability are as critical as the technical aspects. A key part of the planning process is to identify similar initiatives, with respect to type of project or collection and also with respect to collection content. This serves the organization, since you can often benefit from decisions and workflows that have already been developed by a similar organization, and also the user, because you incorporate the principle of seamless access to information into your design, and because you actively identify and solicit collaboration with digital collections that will be useful to your target audience base.

Social interoperability can be as simple as searching for similar projects and collections that can serve as models or collaborators with your organization and making contact to schedule a site visit or a conference call to discuss the project. Other ways to establish and maintain contact are to reference each other's projects on the respective websites, through a webliography or simple link exchange, attending conferences, joining email discussion lists and sharing information via blogs, articles and presentations.

It is important to remember that a key role for your project or collection will be to serve as a gateway or springboard to other, similar collections for your target audiences.

Created: August 28, 2006

Last Updated: August 28, 2006