DALNET Information Hub Development Committee Metadata Task Force
Report and Recommendations
Draft, March 2002.

Information Hub Development Committee Metadata Task Force Members: M. Barash (WSU); Willy Cromwell-Kessler (DPL; Chair); Annaclare Evans (DALNET); Fran Krempasky (WSU); Daniel Golodner (Reuther); Kris McLonis (UDM); Adrienne Aluzzo (DPL). *Past Members*: Mary Galvin (DIA); Donna Roe (UDM)

CONTENTS

1. INTRODUCTION.

The Information Hub Development Committee Metadata Task Force (IHDC MTF) was charged in early summer 2000 to investigate metadata issues and "develop an information architecture that incorporates and integrates the diverse resources of the DALNET Im@gine Information Hub."(1) A number of factors influenced the work of the Task Force:

- A standards based approach to metadata is highly desirable.
- It is not yet clear how the corresponding Information Hub information technical infrastructure will develop.
- It is assumed that the nature of the Hub's content and its potential contributors will most likely be diverse —making a flexible approach highly desirable.
- The library search interface is currently limited in access to the MARC-based Horizon catalog, but it is likely and desirable that library systems will evolve toward a multi-standard environment.
- Application of MARC metadata can be complex, time-consuming and expensive; and not always appropriate to the materials described; it may require more resources than are available for some digitization projects.
- The digital environment is complex and multi-layered; consequently, digital metadata must meet more complex needs than are currently reflected in MARC-based catalog development. (2)
- Metadata systems alternative to MARC, although numerous, are still in the development stage although some firm directions are beginning to emerge.

- Systems vendors are involved in work with several standards and their decisions about implementation will be crucial in determining the direction of the standards environment.

These factors, combined with a survey of the metadata landscape as it looked in 2000/2001, led the MTF members to articulate the following provisions that govern the recommendations formulated in this document:

- Recommendations and metadata models formed at this early stage will need to be revisited regularly and will continue to evolve.
- While the MTF endorses a standards-based approach, they feel that it is still premature to lock DALNET into any particular development path.
- Recommended metadata models must be widely applicable and can not be "one size fits all" in nature.
- While DALNET data standards groups can provide basic guidelines and assist with issues arising in individual projects, DALNET institutions undertaking digitization projects must make metadata decisions based on the specific nature of their projects.
- In a diverse data environment, the role of DALNET metadata standards groups must be to insure the overall compatibility of metadata deployed by individual projects with the development of the overall HUB information architecture.
- Experimentation with different metadata systems and structures is desirable if the Hub is to grow and develop optimally.
- Metadata decisions must be made with a view to an evolving rather than a static information architecture.
- Metadata may be multi-layered in nature (e.g., archival collection level records may be linked to a database of EAD finding aids which are, in turn, linked to digital surrogates that carry structural information in a file header).

2. METADATA INFORMATION MODELS:

The group considered metadata for three different information models:

MODEL 1.

All metadata will reside in a separate MARC-based Horizon database. WebPAC or iPAC indexes and displays will be considered in the selection and application of data elements. The database will be selected from a menu display in the iPAC or WebPAC display.

EXAMPLE: Smithsonian Library Catalog (www.siris.si.edu)

PROS:
- Does not require additional software or database interface
- Permits DALNET institutions to leverage current technical services skills to create digital metadata databases
- Can custom-create MARC or non-MARC templates to assist with data entry

CONS:
- MARC is a complex data system requiring trained staff for correct application.
- MARC does not effectively accommodate all data structures that may be desirable for some types of digital representation.
- May be difficult to tailor to the needs of specific projects.

MODEL 2.

Metadata may reside in either a MARC-based Horizon database or in stand-alone databases that may be accessed via a link in a Horizon library catalog bibliographic record (1st level metadata) for the database, from a menu choice on a IHDC web gateway, and/or through Metadata embedded in a Website for the database. Second level metadata in the stand-alone database may conform to any standard or format or indexing specifications without regard to compatibility with the MARC based Horizon catalog-- although first-level metadata in the library catalog will conform to conventional cataloging specifications.

EXAMPLE: William Blake Archive (www.blakearchive.org) This database is one of several active research projects developed and maintained by the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia. It can be accessed from the IATH active projects Webpage where it is listed in a menu of other digital image archives. It is also represented by a bibliographic record in the University of Virginia library catalog, Virgo. Metadata embedded into the Homepage for the archive may also lead Web-searchers to the archive. The database itself, contains special indexes to digital surrogates of images and text that can be searched by means of a separate, specialized vocabulary that has little relation to other similar archives or to the library catalog.

PROS
- Does not require additional software or database interface on the DALNET level although individual institutions will be responsible for developing/acquiring software for any non-Horizon stand-alone databases.
- Flexible metadata design and deployment
- Metadata for each project can be most appropriate for the nature of the digital objects represented.
- May accommodate inexpensive metadata capture methods.

CONS:
- Access and control becomes more difficult as the number of databases grow.
- Data integration is at the highest level only.

MODEL 3.

Metadata may be stored in separate Horizon, MARC databases or in stand-alone databases using various data systems, all or some of which are accessed via a special

search engine that will permit integrated access and display of the contents of the various databases. Data systems selected must be compatible with the integrated search engine and the selection of elements must take uniform indexing needs into consideration. Databases may also be accessed separately via a Hub gateway and links may exist between the library bibliographic catalog records and specific databases if desirable.

PROS:
- Permits both integrated and individual access to metadata for digital projects.
- Provides a flexible approach where multiple pathways lead to resources.
- Permits use of diverse metadata systems without necessarily sacrificing some degree of integration.
- Library systems are evolving in this direction (e.g., Webfeet, EnCompass, etc.)
- May accommodate inexpensive methods for metadata capture.

CONS:
- Requires additional searching software.
- Requires careful monitoring of emerging standards.

4. DISCUSSION OF INFORMATION MODELS:

DALNET institutions have already make use of models one and two (e.g., UDM Shipping Database and DALNET Health Calendar respectively). It is most likely that these two models will inform efforts in the immediate future. However, the Horizon system vendor, epixtech, is now developing an integrated search engine for Horizon that would bring together databases using several emerging data systems; other vendors have been moving in this direction for some time. It is a possible and highly desirable future direction for DALNET for these reasons:
- Model 3 would offer the best access to DALNET digital resources.
- It may harness data deployed according to the other two models—which will provide intermediate strategies for digital metadata.

The recommendations of the Metadata Task Force presuppose that DALNET will eventually deploy metadata in a fashion compatible with Model 3, but will make use of Models 1 and 2 exclusively during an interim period. Consequently, the MTF believes that metadata choices when making use of model 1 or 2 should be carefully considered so as not to hinder integration of resources as DALNET moves toward Model 3.

In order to insure that databases created now can be effectively integrated later, the MTF work focused on specifying a core semantic data dictionary comprised of elements that are likely to be used in DALNET projects (App. A). To insure compatibility with future standards developments as well as the ability to accommodate a wide variety of content they developed a mapping of these elements to the following standards (App. B):

- MARC21 core record elements (www.loc.gov/marc/)
- Dublin Core Element set (dublincore.org/)
- Encoded Archival Description (EAD) (www.loc.gov/ead/)

- Categories for the Description of Works of Art (CDWA)
  (www.getty.edu/research/institute/standards/cdwa/index.html)
- Visual Resources Association Core Categories (VRA Core)
  (www.vraweb.org/vracore3.htm)

Data elements cross-referenced against these metadata systems can be expected with some confidence to be compatible with future search engine developments as well as to provide some degree of interoperability across systems. They also provide for a core level of description for print, visual, manuscript materials which seem to be likely candidates for initial inclusion in the Hub. This mapping can be expanded to other systems as necessary and the data set itself can be extended as necessary.

The DALNET data element set borrows freely from the Dublin Core and from the Getty Reseach Institute's Categories for the Description of Works of Art (CDWA); it is currently limited to access and description, and it is assumed that metadata will be carried separately from the digital file it references. Future extensions should include elements for preservation, other administrative and structural information as well as more subject and format specific data elements. Additionally, decisions must be made about where structural metadata will be stored—for instance, in some cases it might be feasible to capture such metadata during the creation of the file and retain it in a file header.

5. DATA ENTRY TEMPLATE

The MTF was also charged with developing data entry templates to facilitate creation of metadata for digital projects. For those using Model 1, such templates can be created at need in the Horizon system. However, the MTF have also contributed to the development of a Web-based data entry template based on an elaboration of the proposed DALNET core data set. With very minimal training, this template can be used by staff unfamiliar with MARC or other data element sets.

Metadata involves not only semantic elements for describing and accessing digital objects, but also involves a standard for data transmission. Data entered through this interface can be stored in MARC tagged files for uploading into Horizon or in the Resource Description Framework (RDF) –XML format for uploading into other databases (see http://dublincore.org/documents/2001/11/30/dcq-rdf-xml/) .

The choice of the MARC standard for data transmission was obvious given what are likely to be ongoing DALNET realities. XML has been widely embraced by implementers and seems to be the surest future direction; it is relatively easy to apply, and can accommodate the full range of digital metadata needs. Numerous XML software applications and standards developments such as the Metadata Encoding and Transmission Standard (METS) (www.loc.gov/standards/mets/) lead the MTF to believe that the XML is equally obvious as an alternative data transmission standard.

6. RESOURCE WEBPAGE.

The MTF developed a Task Force Website which includes a page linking to relevant metadata resources. This page has links to general articles about metadata, links to Websites dedicated to special metadata systems, vendors and developers and other resources likely to be of interest to those concerned with selecting metadata for digitization projects. This Webpage will be moved to the DALNET Web site for greater ease of access by those working on Metadata projects.

7. SUPPORTING RECOMMENDATIONS.

General
- Metadata proposals developed by DALNET Hub participants for their projects will be vetted by the appropriate DALNET data standards group.
- The Metadata Resource page will be maintained and updated by the Dalnet Data Standards Group.

Access and Description
- DALNET Digital projects should select descriptive and access metadata compatible with the DALNET data element set whenever possible (App. A.).
- Required data elements in the DALNET data element set should always be present except under special circumstances that are be approved by the Data Standards Committee.
- Extensions to the DALNET Data Element set should be approved by the Data Standards Committee, included in the Data Dictionary and, if possible, should be mapped across relevant data systems.

Authority Control
- Formal name, corporate body and uniform title access points will be verified against the DALNET@imagine bibliographic database; headings will conform to entry forms found there; if access points are not found there, they will conform to the name found on materials in the digital project.
- When subject access is based on a recognized thesaurus that source will be identified (using the conventions of the data system employed). Uncontrolled access will be appropriately tagged or otherwise identified (based on the conventions of data system employed).

Future Directions
- Oversight of DALNET digital metadata will be an ongoing concern and in order to assure proper continuity of action, should fall under the guidance of the DALNET Data Standards Committee or other on-going body with oversight responsibility
- Tasks Forces should be charged by the Data Standards Committee to address the following issues
  - Further develop file-naming conventions to ensure that links between metadata and images are consistent and can be easily maintained, with attention to hierarchy and granularity of interrelated files (to insure that links between metadata and images are consistent and can be easily maintained).

- Extension of the DALNET Data Element set and mapping to provide for preservation and administrative data needs.
- Extension of the DALNET Data Element set for structural metadata; in particular evaluation of the Metadata Encoding and Transmission Standard (METS) for structural metadata (www.loc.gov/standards/mets/).
- Extension of the DALNET Data Element set and mappings to provide more extensively for sound and moving image files.
- Explore issues of authority control and indexing over diverse databases.
- Monitor the developing DALNET information infrastructure so that metadata activities can develop and change in tandem.

7. REFERENCES

(1) IHDC Metadata Task Force: Charge.
(http://home.msen.com/~kessler//MTF/Charge.html

(2) For a thorough discussion of the roles of metadata in the digital environment see the first section of Anne J. Gilliand-Swetland's "Setting the Stage," in *Introduction to Metadata: Pathways to Digital Information*, ed. Murtha Baca. 2nd version., Gettty Research Institute, 2000.
(http://www.getty.edu/research/institute/standards/intrometadata/index.html