METADATA STANDARDS

This chapter lists the major metadata standards in use or under development. Standards are subdivided into six areas: general, transportation models, education, media-specific, preservation, and rights. Each standard lists its official URL, sponsoring agency, community of use, purpose and goals, description, potential for information organizations, and key projects.

General metadata standards

General metadata standards are the most common, well-known, and universally accepted schemas to date. Some are general and meant to be used as universal or common-denominator standards; others are for specific communities with specific information resources. Most of the metadata standards listed here have emerged as practical applications in use, or will probably emerge as the most commonly applied standards, due to broad international support, history and development, or industry application and acceptance.

Dublin Core Metadata Initiative (DCMI)

http://dublincore.org

The Dublin Core Metadata Initiative (DCMI) is managed by an international board of trustees, but most of the direction and maintenance of the standard has been led by the Online Computer Library Center (OCLC) in Dublin, Ohio.

Community of use

Librarians, Web content providers, Web resource creators, metadata creators, and general public.

Purpose and goals

Dublin Core assists in the discovery and description of Web and electronic resources. It is designed to provide a simple descriptive metadata standard extensible to Web resources of any format or subject domain.

Description

Dublin Core is a set of 15 core elements that assist in simple description and discovery of electronic resources. The standard's basic principles are the reasons for its success as a viable common-denominator metadata standard for electronic resources. These principles are: all elements are optional, all elements are repeatable, all elements can be displayed in any order, the standard is extensible (built to be a starting point for richer description and variety), and the standard is international.

Dublin Core allows different communities to both refine the elements through the addition of qualifiers (the ability to further refine and limit a metadata element, such as specifying whether a subject is Library of Congress Subject Headings (LCSH) or Medical Subject Headings (MESH)) and develop complementary element sets that are discipline-specific or focused on specific types of metadata (administrative, rights management, and so on). In addition, the standard has been developed by an international and multidisciplinary constitu-

ency, providing a strong support base that addresses the concerns and needs of a variety of communities.

Three standards levels are within Dublin Core: Unqualified Dublin Core, Qualified Dublin Core, and Community Dublin Cores. Unqualified Dublin Core is the basic core element set that most other metadata standards either try to emulate or at least maintain interoperability with. This 15-core-element set is divided into three areas:

Unqualified Dublin Core, http://dublincore.org/documents/dces

Qualified Dublin Core, http://dublincore.org/documents/dcmes-qualifiers

Instantiation is the concrete representation of something abstract.

Content	Intellectual Property	Instantiation
Title	Creator	Date
Subject	Publisher	Type
Description	Contributor	Format
Source	Rights	Identifier
Relation		
Coverage		
Language		

Content elements address descriptive metadata, intellectual property elements focus on rights and copyright metadata, and instantiation elements concern time and identification metadata.

Qualified Dublin Core, along with the current library application profile being developed by the DCMI Library working group, is basically a metadata equivalent to the traditional MARC format used by most library cataloging departments. Qualified Dublin Core was developed in conjunction with interested parties, but especially the library community, to provide the necessary interoperability requirements necessary to build the crosswalk between Dublin Core and MARC.

The importance of interoperability is especially apparent in projects such as the Cooperative Online Resource Catalog (CORC) experiment. CORC has now become the new metadata desktop cataloging interface known as Connexion offered by OCLC in July 2002. The various Community Dublin Core standards available are not listed on the official Web site, but some are represented by the working groups assembled to define these elements. Some of the working groups are: Education, Government, and Libraries. Other interest groups that maintain e-mail lists with DCMI include Business and Environment. Many other metadata standards also incorporate Dublin Core as the basic element set in their more complex and discipline-specific metadata standards.

The Dublin Core metadata element set has been approved by the American National Standards Institute (ANSI) and the National Information Standards Organization (NISO) as ANSI/NISO Z39.85-2001.

Potential for information organizations

DC-Library Application Profile (DC-Lib), http://dublincore.org/documents/library-application-profile

Librarians have numerous opportunities for input into the development of the DCMI. Many working groups and interest groups exist, most of which actively work between the regularly scheduled conferences. Of particular interest to libraries is the Libraries working group, which has just issued a DC-Library Application Profile (DC-Lib) that clarifies the use of the Dublin Core in libraries and library-related applications and projects.

Many other metadata standards incorporate or plan to incorporate Dublin Core in some way as the common metadata denominator for interoperability and reliability of data across diverse schemas. Due to its history and organizational

structure, Dublin Core has emerged as one of the most viable and resilient metadata standards for the future.

Key projects

See an extensive list of current digital projects incorporating Dublin Core metadata at www.dublincore.org/projects. Many of these projects are located in Australia and in Scandinavian countries, where early Dublin Core initiators have extensive experience using the standard or have incorporated Dublin Core into their government archival policies and procedures.

Publishing Requirements for Industry Standard Metadata (PRISM), a major application of the Dublin Core, www.idealliance.org/prism, www.oasis-open.org/cover/prism.html, www.idealliance.org/prism/techdev/prismspec11.asp.

Data Documentation Initiative (DDI)

www.icpsr.umich.edu/DDI

A project of the social sciences community, sponsored by the Inter-university Consortium for Political and Social Research (ICPSR).

Community of use

Academic social scientists and teachers.

Purpose and goals

The Data Documentation Initiative (DDI) establishes an international criterion and methodology for the content, presentation, transport, and preservation of metadata about data sets in the social and behavioral sciences. The basic goal early on was to replace the obsolete OSIRIS Codebook/Dictionary format with a modern format. These codebooks can now be created in a uniform, highly structured format easily and precisely searchable on the Web, which also lends itself well to simultaneous use of multiple data sets and that significantly improves the content and usability of metadata. This effort also may have far-reaching implications for the improvement of the entire process of data collection, data dissemination, and data analysis.

Description

DDI is an XML document type definition (DTD) that assists in the digital presentation and preservation of social sciences data sets. The bibliographic portion of this metadata standard is linked and crosswalked to Dublin Core. A DDI-compliant document has the following sections: a Document Description (similar to a TEI header), a Study Description, a File Description, a Variable Description, and an Other Materials section. DDI is on Version 1.02.1 as of February 2002.

Key projects

A list of current projects incorporating the DDI DTD is maintained at www.icpsr.umich.edu/DDI/ORG/index.html.

Encoded Archival Description (EAD)

http://lcweb.loc.gov/ead

This standard is maintained in the Network Development and MARC Standards Office of the Library of Congress in partnership with the Society of American Archivists (SAA).

Community of use

Archivists, librarians, museum curators, manuscript repositories.

Purpose and goals

The Encoded Archival Description (EAD) metadata standard was developed as an electronic equivalent to a finding aid. It was used to investigate the desirability and feasibility of developing a nonproprietary encoding standard for machine-readable finding aids such as inventories, registers, indexes, and other documents created by archives, libraries, museums, and manuscript repositories to support the use of their holdings.

The goals of the project are:

- To present extensive and interrelated descriptive information found in archival finding aids
- To preserve the hierarchical relationships existing between levels of description
- To represent descriptive information that is inherited by one hierarchical level from another
- To move within a hierarchical informational structure
- To support element-specific indexing and retrieval

Description

The EAD metadata standard is an SGML/XML-based document type definition (DTD) that museums, archives, and some libraries are using to create, store, and distribute digital descriptions of their collections in the form of finding aids. Finding aids are the equivalent of card catalogs within archives, except that descriptions are usually based on the collection level rather than the item level, given the context of archival and museum collections. Until EAD, finding aids were only available in print format, meaning that a potential researcher could only examine the card catalog of archives and museums in person. Since EAD is only meant to display and format finding aids online, other metadata standards must sometimes be incorporated to display and format items below the collection level.

EAD-encoded finding aids serve the same general purpose for archival materials as the MARC format does for library materials. EAD provides an online catalog for researchers to search, similar to the functionality available in most library OPACs. Researchers can now query collection-level finding aids from remote locations over the Internet, and are usually able to link to item-level records and even digitized versions of some items. EAD is closely related to MARC and is formatted to take advantage of the many interrelationships and cross-functional applications for interoperability. Crosswalks between EAD and many other metadata standards are available at the official Web site.

Potential for information organizations

Most archival, museum, and manuscript repositories have or are in the process of converting their finding aids into EAD. The conversion of finding aids is the main function of EAD.

EAD guidelines for application of the standard, tools, and helper files, are available at the official Web site. Anyone can join the site's listserv.

Key projects

A large list of projects and implementers using the EAD DTD is available and maintained at http://jefferson.village.virginia.edu/ead/sitesann.html.

An EAD cookbook developed by Michael Fox at the Minnesota Historical Society, http://jefferson.village.virginia.edu/ead/cookbookhelp.html

More EAD-related software and tools, http://jefferson.village.virginia.edu/ead/products.html

Federal Geographic Data Committee (FGDC)—Content Standards for Digital Geospatial Metadata (CSDGM)

www.fgdc.gov/metadata/contstan.html

This site is maintained by the Federal Geographic Data Committee (FGDC), part of the National Spatial Data Infrastructure (NSDI).

Community of use

Geospatial data collectors—federal or state government, private, or academic.

Purpose and goals

Often referred to as the FGDC metadata standard, the Content Standards for Digital Geospatial Metadata has emerged as the primary geospatial metadata standard. Its purpose is to provide a common set of definitions and terminology for documentation related to digital geospatial metadata.

Description

The CSDGM standard was developed in response to Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," that was signed April 11, 1994, by President Bill Clinton. Section 3, Development of a National Geospatial Data Clearinghouse, paragraph (b) states: "Standardized Documentation of Data, ... each agency shall document all new geospatial data it collects or produces, either directly or indirectly, using the standard under development by the FGDC, and make that standardized documentation electronically accessible to the Clearinghouse network."

All federal agencies are required by this order to produce and collect geospatial data using the CSDGM and to deposit the metadata sets in a FDGC. Government agencies of the state and other levels, as well as the private sector, have adopted this standard.

The documentation for the CSDGM comprises 334 different metadata elements, organized into seven major sections and three minor sections. The major sections are:

- **Identification information:** data set title, area covered, keywords, purpose, abstract, and access and use restrictions
- **Data quality information:** horizontal and vertical accuracy assessment, and data set completeness and lineage
- **Spatial data organization information:** raster, vector, or an indirect (such as address) link to location
- **Spatial reference information:** latitude/longitude, coordinate system, or map projection
- **Entity and attribute information:** definitions of the attributes of the data set
- **Distribution information:** distributor, file format of data, offline media types, online link to data, and fees
- **Metadata reference information:** who created the metadata and when

The three floating minor sections are:

- **Citation information:** originator, title, publication date, publisher
- **Time period information:** single date, multiple dates, range of dates

- **Contact information:** contact person or organization, address, phone, e-mail

This metadata standard only defines content, not format. Delivery, storage, creation methods, and interoperability of information are left up to the implementers for incorporating other metadata standards or transportation models. The Spatial Data Transfer Standard (SDTS) was developed by the federal government to help transfer geospatial data among various software editions. Various metadata tools and toolkits are available throughout the geospatial community.

The CSDGM document is in Version 2, published in 1998. Various profiles or extensions to this document are available for biological data, shoreline data, referenced cultural and demographic data, and remote sensing.

Potential for information organizations

Libraries, archives, and museums with extensive map collections should be familiar with this standard. Universities with Geographic Information Systems (GIS) projects, data, and equipment also should learn to work with it. Tremendous applications are available with CSDGM, given the amount of data it collects and stores, and the detail the standard requires. The challenge is to establish interoperability structures among other metadata standards.

Key projects

The official FGDC Geospatial Clearinghouse (www.fgdc.gov/clearinghouse/clearinghouse.html) is a collection of more than 250 spatial data servers that have digital geographic data primarily for use in GIS, image processing systems, and other modeling software.

Many other geospatial clearinghouses exist. Although the official Web site doesn't list any major projects, see a geospatial metadata primer at www.lic.wisc.edu/metadata/metaprim.htm. This primer is an excellent resource, with many links to other clearinghouses as well as major projects, tools, software, FAQs, and other resources.

Of special interest is the NSDI Metadata and WWW Mapping page at www.blm.gov/gis/nsdi.html, and the Oklahoma State University Geospatial Metadata page at www.seic.okstate.edu/gis/metadata.html.

Global Information Locator Service (GILS)

www.gils.net

The formally constituted body for maintaining the GILS Profile is the Open Systems Environment Implementers Workshop/Special Interest Group on GILS (OIW/SIG-GILS). Version 2 is the approved GILS Profile, and the U.S. Geological Survey is the maintenance agency for this standard as designated by the National Institute for Standards and Technology (NIST).

Eliot Christian of the U.S. geological survey was the early initiator of the GILS concept. A partial list of other important contributors can be viewed at www.gils.net/acks.html.

Community of use

Those involved in government information at all levels (federal, state, local).

Purpose and goals

The goal of the Global Information Locator Service (GILS) is to help people to find information of all kinds, in all media, in all languages, and over time.

Description

GILS has become the major metadata standard for information in the U.S.

federal government and has been adopted by many state governments as well. As part of the U.S. government's role in the national information infrastructure, GILS identifies and describes information resources throughout the federal government and provides assistance in obtaining the specific information. GILS supplements other government and commercial information dissemination mechanisms and employs international standards for information search and retrieval so that information can be retrieved in many ways.

The GILS profile specifically mandates using ANSI/NISO Z39.50. The GILS Attribute Set is a superset of the Bib-1 Attribute Set. GILS servers support the following three record syntaxes: USMARC, the Generic Record Syntax (GRS-1), and the Simple Unstructured Text Record Syntax (SUTRS). GILS Schema have two general schema classes: primitive and constructed.

The 28 GILS Core Elements can be more narrowly defined through the use of qualifiers. GILS is mapped to many other metadata standards, including MARC21, EAD, and Dublin Core.

GILS begins with a bibliographic foundation but embraces the diversity found among information communities. Searchers, information providers, and intermediaries have different reasons for supporting GILS.

For the full explanation of this set, view ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt.

GILS assists the federal government in its mandate to produce all information electronically, as well as provide the search mechanisms and address the archival needs of long-term storage. Many state governments have followed the lead of the U.S. government, and adopted GILS. To learn how GILS is important from the perspectives of content providers, intermediaries, searchers, and software companies, see www.gils.net/overview.html#perspectives.

Potential for information organizations

The GILS standard has been adopted as a metadata standard by the federal government and a majority of state and local governments. Since GILS is based on the Z39.50 specification, electronic access to GILS systems and data has been relatively seamless and transparent to most information organizations. The ease of transcription into MARC and other metadata standards means that GILS information can be presented to the user through bibliographic databases and online library catalogs. The ease of use of this metadata standard will eventually result in more information being available from governmental sources, which are typically considered black holes of information by most librarians and archivists.

Key projects

An extensive list of GILS implementers is available at www.gils.net/examples.html and www.gils.net/implement.html.

The vol. 18, no. 3, 2001 issue of *Government Information Quarterly* is titled "Metadata: A Networked Information Strategy To Improve Access to and Management of Government Information," and is devoted to specific examples of GILS implementations.

A similar attempt by the United Kingdom to adopt metadata standards for its governmental information has been posted at www.go.vic.gov.au/pdfs/e-Govt_Metadata_Standard_v0_2.pdf. This particular draft standard uses the Dublin Core as its basis.

Machine-Readable Cataloging (MARC) 21

www.loc.gov/marc

Maintained by the Library of Congress Network Development and MARC Standards Office and the Standards and Support Office at the National Library of Canada, working with the ALCTS/LITA/RUSA Machine-Readable Bibliographic Information (MARBI) Committee.

Community of use

All information organizations, especially libraries.

Purpose and goals

The MARC 21 formats are standards for the representation and communication of bibliographic and related information in machine-readable form. MARC 21 formats are sets of codes and content designators defined for encoding machine-readable records, that is, in moving print card catalog cards into the computer environment. Formats are defined for five types of data: bibliographic, holdings, authority, classification, and community information.

Description

The Machine-Readable Cataloging (MARC) format was specifically designed as a communication format to assist in the exchange of bibliographic and related information among computer systems. Of the five types of data standardized in the MARC 21 format, the most current manifestation of MARC, the bibliographic standard, is discussed here. In conjunction with the *Anglo-American Cataloguing Rules* (*AACR*), in its second edition, 1998 revision (*AACR2R*), the MARC format has been the library cataloging standard for more than 30 years.

A MARC record involves three elements: the record structure, the content designation, and the data content of the record. The record structure of MARC is an implementation of national and international standards, specifically the Information Interchange Format (ANSI Z39.2) and the Format for Information Exchange (ISO 2709). The content designation or data elements of the MARC record are defined by the MARC 21 standard. Finally, the content of the data elements is defined outside the MARC 21 standard by *AACR2R*.

A MARC data record consists of three main sections: the leader, the directory, and the variable fields. The leader is the first 24 characters of the record, designed to assist the computer in translating the information in the directory and variable fields. The directory tells the computer what tags are in the record and where they are placed. Variable fields can be divided into variable control fields and variable data fields. The data in a MARC record is organized into a three-character field known as a tag. The bibliographic format tags are grouped into the following structure:

See the booklet *Understanding MARC—Bibliographic*, available online at www.loc.gov/marc/umb for more information.

0XX = Control information, numbers, codes

1XX = Main entry

2XX = Titles, edition, imprint

3XX = Physical description

4XX = Series statements

5XX = Notes

6XX = Subject access fields

7XX = Name, added entries or series; linking

8XX = Series added entries; holdings and locations

9XX = Reserved for local implementation

The last two characters of the tag are given the following meanings:

X00 = Personal names

X10 = Corporate names

X11 = Meeting names

X30 = Uniform titles

X40 = Bibliographic titles

X50 = Topical terms

X51 = Geographic names

Potential for information organizations

Given the tremendous richness of display and content available with MARC 21, information organizations must be vigilant on issues of interoperability among metadata standards and MARC 21. Many metadata standards have been developed in isolation of MARC 21 and the library community, and although they may serve the needs of their own internal information, they have not been developed to interact or even translate properly outside their particular venues.

The Library of Congress has taken a lead by maintaining many metadata crosswalks between MARC 21 and various other metadata standards. In addition, the Library of Congress is an active participant in developing other major metadata standards, including Dublin Core, Text-Encoding Initiative (TEI), Online Information Exchange (ONIX), and EAD. Participation is important for the continued representation of information organizations and their needs as these other standards are developed and used.

Key projects

The Online Computer Library Center (OCLC), the world's largest bibliographic database, is at this time entirely composed of bibliographic records based on MARC format, www.oclc.org/home.

Library of Congress standards page, http://lcweb.loc.gov/standards.

Library of Congress MARC mappings, www.loc.gov/marc/marcdocz.html.

Metadata Object Description Schema (MODS)

www.loc.gov/standards/mods

MODS is being developed by the Network Development and MARC Standards Office at the Library of Congress.

Community of use

Libraries and information organizations.

Purpose and goals

MODS is developing a schema for a bibliographic element set used for many purposes, particularly for library applications. MODS will carry selected data from existing MARC 21 records and enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than numeric ones, but still maintains the MARC 21 semantics, making the standard more compatible with existing library data.

Description

The Metadata Object Description Schema (MODS) is a new development within the Library of Congress. Some potential uses of this standard include:

- A Z39.50 next generation specified format to represent metadata for harvesting
- Representing a simplified MARC record in XML, as an extension schema to the Metadata Encoding and Transmission Standard (METS)
- Metadata in XML that may be packaged with an electronic resource

- Original resource description in XML syntax

Some advantages of MODS are that its element set would be richer than Unqualified Dublin Core. The schema also is more end-user oriented than the full Open Archives Initiative (OAI) MARC XML schema; it will be easier to implement. In addition, the element set is more compatible with library data than the ONline Information EXchange (ONIX International) standard and simpler than the full MARC format.

Potential for information organizations

If MODS is successful, its interoperability with MARC records and to other metadata standards could be of great importance. The Network Development and MARC Standards Office at the Library of Congress is intensely involved in many of the metadata standards listed here, and MODS development appears to be focused specifically on providing a medium-level metadata standard. Whether MODS is similar to Qualified Dublin Core, also constructed in relation to the MARC format, has yet to be seen.

Key projects

Thus far, only sample MODS documents are available for a test record, book, Web document, serial, and map at the official Web site.

The Open Archives Initiative (OAI)

www.openarchives.org

The Open Archives Initiative (OAI) is composed of three committees: a steering committee, which manages the general policy direction; an executive committee, comprised of the two Cornell professors responsible for the initiative and direction of the standard—Carl Lagoze and Herbert Van de Sompel—and a technical committee, in the process of being constituted. OAI also is supported by the Digital Library Federation (DLF) and the Coalition for Networked Information (CNI).

Community of use

Anyone involved in metadata creation and interoperability of standards, digital preservation, Web content providers.

Purpose and goals

The Open Archives Initiative (OAI) has developed a mission statement of the organization's purpose and goals:

"The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program.

"The fundamental technological framework and standards that are developing to support this work are, however, independent of the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials. As a result, the Open Archives Initiative is an organization and an effort explicitly in transition, and it is committed to exploring and enabling this new and broader range of applications."

As the library industry learns more about the scope of applications for metadata technology and standards and understands the structure and

Open Archives Initiative, www.openarchives.org/organization/index.html

culture of the various adopter communities, the OAI's mission and organization will evolve.

Description

OAI has emerged as one of most invigorating and exciting developments in metadata interoperability. OAI provides a simple, low-barrier approach to interoperability. OAI has produced a mechanism—the Protocol for Metadata Harvesting (PMH)—that enables data providers to expose their metadata. PMH is being deployed in many new services and system architectures for a diverse set of communities.

Even though OAI has the word *archives* in its name, it no longer has much to do with archiving or archives. The phrase *open archives* stems from the idea of authors depositing preprints of their articles into accessible databases, providing free access to their scholarship. Although OAI grew from this effort, OAI has now gone beyond the preprint and e-print community to focus on urgent needs related to interoperability, automatic metadata harvesting, and the development of metadata repositories that this particular standard seems uniquely positioned to address. The name has been retained due to its popularity and recognition.

The PMH is an interface that allows metadata about digital objects to be harvested by external software applications that wish to collect the metadata. PMH uses a simple HTTP-based request-response transaction framework for communication between a harvester and a repository. To accomplish this communication, multiple metadata standards are supported, but the protocol requires that all servers offer Unqualified Dublin Core metadata encoded in XML as the common format. Two applications are enabled by PMH: repository synchronization, which allows the user to compare metadata from two or more repositories and decide what objects and accompanying metadata should be copied from one repository to another; and federated searching, which allows the user to collect metadata from many sites, normalize it into readable data, dedupe the results, and offer search services against the resulting database. PMH provides a powerful mechanism for building union-catalog-type databases through automatic harvesting of digital objects contained in multiple servers and repositories.

Although OAI and the PMH (in theory) solve the interoperability problem of multiple metadata standards and repositories, they also raise many issues that OAI recognizes and invites discussion on. These issues include:

- The acceptable use and intellectual property issues related to metadata
- Questions related about where to harvest metadata, including selection, registries, and trust questions
- Granularity of objects exposed by PMH
- The future of multiple metadata standards
- The pros and cons of manually produced versus computer-generated metadata and the push toward community-based metadata schemas.

OAI is meant to be simple to understand and implement. An OAI record has three parts: a header, metadata, and an about container. The header contains the unique identifier for the record and a datestamp indicating the date of creation and other time-sensitive information. The metadata part must be a minimum in XML Unqualified Dublin Core. The about container is optional and holds rights information about the metadata, including terms and conditions for use. The internal structure of this optional container is up to individual

Granularity: Levels of complexity or description related to metadata standards (for example, Unqualified Dublin Core is considered low granularity, but Qualified Dublin Core is considered high granularity, based on the complexity of the tags and coding)

communities to develop.

The PMH searching mechanism consists of six requests or verbs, carried within the HTTP POST or GET PROTOCOL. This simple structure helps data providers configure their OAI-conformant repositories by using readily available Web tools. These six requests are: GetRecord, Identify, ListIdentifier, ListMetadataFormats, ListRecords, and ListSets. Testing and refinement of the PMH is being carried out in multiple communities, and results are issued through the OAI listserv.

Version 2.0 of the OAI-PMH Version 2.0 was made available in June 2002.

Potential for information organizations

OAI looks at the problem of multiple metadata standards and interoperability from a different angle than other initiatives. Instead of focusing on collaboration, revision, and reduction of the number of metadata standards, and the construction of crosswalks between metadata schemas, OAI provides a quick, clean interoperability of information. The requirement of using Unqualified Dublin Core as the common metadata standard doesn't preclude the use of other schema, as long as the schema can be crosswalked to Dublin Core.

OAI focuses on the exchange of information, not the information itself. OAI lets each community compile its content in whatever metadata standard the community wants, provided that an Unqualified Dublin Core version is compiled as well. OAI then provides the mechanism through the OAI-PMH software to efficiently offer this content to the world. Information organizations are experimenting with the OAI-PMH software, and the OAI is addressing the interoperability questions regarding digital objects and digital content.

Key projects

The Andrew Mellon Foundation is funding seven OAI metadata harvesting projects. Descriptions of these projects are available at www.arl.org/newsltr/217/waters.html. View the work of the Illinois OAI Protocol Metadata Harvesting Project at http://oai.grainger.uiuc.edu. The work of the University of Michigan is available at http://oaister.umdl.umich.edu.

A list of registered OAI Metadata Providers is available at http://oaisrv.nsdl.cornell.edu/Register/BrowseSites.pl.

OAI Metadata Provider Tools and Resources, www.openarchives.org/tools/tools.html.

A tutorial for exposing and harvesting metadata using the OAI-PMH is available at http://library.cern.ch/HEPLW/4/papers/3.

An announcement on the release of OAI-PMH Version 2.0 can be seen at www.openarchives.org/news/OAI-2.0A-PR.html, along with a link to the official e-mail describing the changes in the software and links to other OAI-conformant tools.

An interesting application of OAI in development is taking place in the music information organization community, related to sheet music. Information on their early discussions is available at www.dlib.indiana.edu/oaimeeting.

Another interesting OAI personalization freeware software tool that has been developed is called my.OAI at www.myoai.com.

Online Information Exchange (ONIX)

www.editeur.org/onix.html

The ONline Information EXchange (ONIX) metadata standard is maintained by the European Group for Electronic Commerce in the Book and Serials Sector (EDItEUR) jointly with Book Industry Communication and the Book Industry Study Group, and with user groups in France, Germany, and Latin America.

Community of use

Book and serials publishers, libraries, and publishing companies.

Purpose and goals

ONIX has become the international standard for representing and communicating book industry product information in electronic form. ONIX was designed to solve three problems:

- The need for richer book data online
- The lack of consistency and standards in data exchange formats in use by book retailers and wholesalers
- The need for a universal, international format in which all publishers could exchange information

Description

ONIX is a rich metadata schema comprised of 236 elements of information that fall into 24 categories. The set includes administrative, structural, and descriptive metadata elements.

ONIX allows for two levels of implementation. Level 1, or ONIX-Lite, is designed to support publishers who have not developed a database for information management. ONIX does not require reference to the XML document type definition (DTD) (Release 1.2). Level 2 contains all the elements of Level 1 plus elements that provide for more in-depth description. ONIX-Lite has 82 elements, and ONIX Level 2 has 236 elements. ONIX can transmit information about print books and e-books, and efforts are underway to include information on videos and serials.

Thirty different administrative data elements are necessary for the internal control purposes of publishers, but they are of little use to the library community. Access data elements, such as names and subjects/genres, are the most problematic area of ONIX. Main entry is essentially ignored in ONIX, which causes problems with crosswalk matching and interoperability. ONIX allows for numeric classification and topical subject headings. Descriptive data elements are fairly consistent with the MARC format, making them similar to bibliographic records used in library catalogs.

Users of ONIX International include online booksellers such as Amazon.com, BN.com (Barnes & Noble), Borders, and Fatbrain. Publishers implementing and using ONIX include Cambridge University Press, HarperCollins, McGraw-Hill, Random House, and Yale University Press. Book distributors and wholesalers using ONIX include Baker & Taylor, Follett, and Ingram.

Potential for information organizations

OCLC is interested in the ONIX standard. Librarians have always been interested in more cooperation with publishers, especially in the standardization of information used in bibliographic description. As of September 2001, OCLC wrote a software program that converted ONIX data into MARC. This program is not installed, though, since demand is lacking. As OCLC moves toward a new database structure using Oracle, both Dublin Core and ONIX have been identified as high-priority metadata standards.

The Library of Congress is developing project prototypes incorporating ONIX through its cataloging directorate, Bibliographic Enrichment Advisement Team (BEAT). Three ONIX-related initiatives exist: a text-capture-and-conversion project that harvests information from ONIX records, especially table of contents information; an experiment with NewBooks exploring records export; and electronic Cataloging-in-Publication (CIP) information exchange.

Key projects

Seven PowerPoint presentations, prepared in 2000, give examples of how ONIX is being implemented by various publishers. Presentations by Amazon, Whitaker, BookData, Cambridge University Press, Harper Collins, and The British Library, as well as an introduction to ONIX presentation, are available at www.bic.org.uk/onixsem.html.

Contact information and project details of current implementers of ONIX International is available at www.bic.org.uk/onix.html.

The Network Development and MARC Standards Office at the Library of Congress maintains an ONIX to MARC 21 crosswalk at http://lcweb.loc.gov/marc/onix2marc.html. View an ONIX to UNIMARC mapping written by Alan Danskin of The British Library at www.editeur.org/onixmarc.html.

A detailed report on the importance of ONIX to the library community by the ALCTS Committee on Cataloging: Description and Access (CC:DA) Task Force on ONIX International is available at www.ala.org/alcts/organization/ccs/ccda/tf-onix3.pdf.

Text-Encoding Initiative (TEI)

www.tei-c.org

In December 2000, after a year's negotiation, a new nonprofit corporation called the TEI Consortium (TEI-C) was set up to maintain and develop the TEI standard. The consortium has executive offices in Bergen, Norway, and hosts at the University of Bergen, Brown University, Oxford University, and the University of Virginia. The consortium is managed by a board of directors, and its technical work is overseen by an elected council.

Community of use

Humanities scholars, librarians, archivists, manuscript repositories, and literary enthusiasts.

Purpose and goals

The Text Encoding Initiative (TEI) is an international research effort intended to produce a community-based standard for encoding and interchanging texts. The basic principles on which the TEI-C is established are that the TEI guidelines, other documentation, and document type definition (DTD) should be free to users; that participation in TEI-C activities should be open (even to nonmembers) at all levels; that the TEI-C should be both international and interdisciplinary; and no role with respect to the TEI-C should be without term.

The TEI Consortium establishes and maintains a home for the Text Encoding Initiative (TEI) in the form of a permanent organizational structure to ensure the continued funding of TEI-C activities, including editorial maintenance and development of the TEI guidelines and DTD, training and outreach activities, and services to members. It also creates and maintains a governance structure for the TEI-C with broad representation of TEI user-communities.

Description

The Text-Encoding Initiative (TEI) is an SGML document type definition (DTD) that is used for marking up humanities and literary texts. TEI was one of the early metadata standards (developed in the late 1980s), and the TEI consortium was established to provide an organizational structure that supports and assists in the development of the standard. The impetus behind TEI was to develop a metadata standard that would support the encoding of literary and linguistic texts, and facilitate sharing of these texts across diverse computing platforms.

TEI has a large and exhaustive tag library, but the section that is of most interest to information organizations is the TEI Header. The current version of the TEI, known as TEI P4, is a huge document with 39 chapters and various appendixes. Chapter 5 contains a description of the TEI Header, which was developed in conjunction with librarians and is meant to represent the descriptive cataloging information found in a MARC record. Every TEI-conformant text must carry a TEI header, tagged <teiHeader>, which consists of four major parts:

- A *file description*, tagged <fileDesc>, containing a full bibliographical description of the computer file itself, from which a user of the text could derive a proper bibliographic citation, or which a librarian or archivist could use in creating a catalog entry recording its presence within a library or archive. The term computer file here refers to the whole entity or document described by the header, even when this document is divided and is stored in several distinct operating system files. The file description also includes information about the source or sources from which the electronic document was derived.
- An *encoding description*, tagged <encodingDesc>, which describes the relationship between the electronic text and its source or sources. The encoding description details whether (or how) the text was normalized (issues of formating, justification, syntax, and grammar) during transcription, how the encoder resolved ambiguities in the source, what levels of encoding or analysis were applied, and similar matters.
- A *text profile*, tagged <profileDesc>, containing classification and contextual information about the text, such as subject matter, the situation in which it was produced, the people described by or participating in producing it, and so forth. Such a text profile is particularly useful in highly structured composite texts such as corporate or language collections, where enforcing a controlled descriptive vocabulary or performing retrievals from a body of text in terms of text type or origin is desirable. The text profile may, however, be of use in any form of automatic text processing.
- A *revision history*, tagged <revisionDesc>, which allows the encoder to provide a history of changes made during the development of the electronic text. The revision history is important for version control and for resolving questions about the history of a file.

The <fileDesc> element contains three mandatory and four optional elements. They are self-explanatory to those knowledgeable about the MARC format:

<titleStmt> (245 |A MARC tag)

<editionStmt> (250 MARC tag)

<extent> (245 |C MARC tag)

<publicationStmt> (260 MARC tag)

<seriesStmt> (4XX MARC tag)

<notesStmt> (5XX MARC tag)

<sourceDesc> (5XX MARC tag).

Within each of these elements, further subdivision and opportunity for bibliographic description is possible.

TEI Lite is a DTD that includes only a small subset of the whole TEI tagset, selected to include the most commonly needed tags. TEI Lite is packaged neatly so users don't have to work out which TEI modules to include or configure the TEI to include them. It is the most commonly used subset of TEI.

Potential for information organizations

Since TEI is well-established and well-known, it has already established a broad base of practice and support in the digital text community. The establishment of this metadata standard assisted in the development of electronic text centers in academic universities, building on the example set by the University of Virginia, one of the early developers and initiators of TEI. Information organizations that have significant text-based collections, particularly in the humanities, will find this standard well-suited to their needs. Crosswalks have already been established between TEI and many of the other major metadata standards, including MARC, EAD, and Dublin Core.

Key projects

The official TEI Web site (www.tei-c.org/Applications/index.html) lists 92 projects, which can be sorted alphabetically by project, by corpus language, or by subject. Most of the major early projects were developed at the University of Virginia and the University of Michigan.

A TEI tutorial page at the official Web site assists in implementation and practical application, www.tei-c.org/Tutorials/index.html

Uniform Resource Identifiers (URIs)

www.w3.org/Addressing

The World Wide Web Consortium (W3C) is the sponsoring agency for all issues related to resource identification of Web resources.

Community of use

Web resource creators, Web resource indexers, search engines, all information organizations.

Purpose and goals

The purpose of uniform resource identifiers (URIs) is to provide a label, consisting of a string of letters or numbers, as a unique, concise way of referring to a particular resource on the Web.

Description

URIs encompass a broad category of schemes for resource identification: uniform resource locators (URLs), uniform resource names (URNs), uniform resource characteristics (URCs), and to some extent, digital object identifiers (DOIs). URLs are an informal term associated with popular URI schemes such as http, ftp, mailto, gopher, and so on. URNs are URIs that have an institutional commitment to persistence and availability, and are intended to serve as location-independent and persistent resource identifiers. URCs have been abandoned as a concept since the URI standard was proposed. DOIs began as a way to maintain digital rights management with resource content on the Web, but the movement has wavered between a focus on DOIs as unique resource identifiers for Web resources or as a business structure to maintain digital rights. Whether URIs can accomplish all or any of these objectives is still undecided.

The need for unique identifiers is a critical concern in both the information organization and Web environments. Information organizations have always had problems with establishing identifiers in the print environment, and the Web only exacerbates the situation. URLs are not reliable long-term mechanisms for providing access to digital objects—the issues of disappearance, versioning, and location movement of these types of resources require a more permanent solution to the problem.

Versioning is when numerous revisions or versions of a document or item are available electronically. Problems develop when URLs are reused as access points when documents and items that they link to are constantly being updated or revised without notifying the user.

The syntax, structure, and implementation of a URI mechanism are underway in the W3C. Whether any major decisions or solutions will be made in the near future by the W3C seems questionable.

Potential for information organizations

The potential for unique and persistent identification of digital objects is immense for information organizations. Librarians have been trying to solve this problem in the print environment for a long time. The Web has engendered a whole new set of complexities and concerns with digital objects. Most of these concerns are economic and have aroused new interest in this topic.

Key projects

"Identifiers and their role in networked information applications" by Clifford Lynch, www.arl.org/newsltr/194/identifier.html.

ALCTS Committee on Cataloging: Description and Access (CC:DA) Task Force on Uniform Resource Identifiers and AACR2, www.ala.org/alcts/organization/ccs/ccda/tf-uri.pdf.

The Digital Object Identifier (DOI) System, www.doi.org.

"What is … a URI?" by Ian Peacock, www.ariadne.ac.uk/issue18/what-is.

"DOI: implementing a standard digital identifier as the key to effective digital rights management" by Norman Paskin, www.doi.org/doi_presentations/Syndey.PDF.

Universal Resource Characteristics (URCs), www.hypernews.org/HyperNews/get/www/URCs.html.

Transportation models

Transportation model metadata standards assist in the interoperability, transfer, and transport of information among standards. Two standards, Resource Description Framework (RDF) and eXtensive Markup Language (XML), provide a wrapper around metadata to enable translation and adaptability of the information into different environments.

Resource Description Framework (RDF)

www.w3.org/RDF

RDF is maintained by the World Wide Web Consortium (W3C).

Community of use

All metadata standards users and creators, and Web content creators.

Purpose and goals

The objective of the RDF is to support the interoperability of metadata. RDF allows descriptions of Web resources to be made available in machine-understandable form, and enables the semantics of digital objects to be translatable and useful in different venues and environments.

Description

RDF is designed to be an infrastructure that enables the exchange, encoding, and reuse of structured metadata. RDF is an XML application that assists in

providing unambiguous methods of expressing semantics via structural constraints. In addition, this framework provides a means for human-readable and machine-processable vocabularies to reuse and extend metadata semantics among disparate information communities. In layman's terms, RDF is a container to assist in the transport and translation of metadata into different communities and packages.

RDF provides a model for describing resources. A resource is defined as any object that is uniquely identifiable by a Uniform Resource Identifier (URI). RDF is technically not a metadata standard but attempts to provide interoperability between applications that exchange metadata. Some application areas that benefit from RDF are: site-maps, content rating systems, collaborative services, electronic commerce, and privacy preferences services. Members of these communities have assisted the W3C in this endeavor by reaching consensus on their syntactical needs and deployment efforts.

RDF consists of two key components: a schema and a data model. The schema (similar to other metadata schemas) enables information organizations to define and relate semantics. The data model connects uniquely identified resources or objects through a system of named relationships. Explaining RDF has proved difficult for both its originators and those interested in understanding it, which is why numerous visual examples of how RDF works exists in the presentations and papers on the official Web site.

RDF was developed in close proximity with Dublin Core, and it is a practical realization of the Warwick Framework (a container architecture designed to aggregate metadata sets) and the W3C's Platform for Internet Content Selection (PICS) begun in 1995.

Potential for information organizations

The initial concepts behind RDF are valid and necessary in the current digital environment. Although RDF was one of the first interoperability models to appear, it has yet to prove itself as a model for the future. The industry has made much discussion as to why this RDF is still unproven. In the meantime, information organizations should become knowledgeable about RDF and understand its basic principles and applicability for the future. Work on the Open Archive Initiative (OAI) may incorporate aspects of RDF in its structure, or RDF may be lost in the resultant popularity of the OAI model.

Key projects

The Dublin Core Metadata Initiative (DCMI) has been intimately involved with the development of RDF. One of the major projects that has resulted from this collaboration is the Cooperative Online Resource Catalog (CORC), a research project from OCLC that explores the cooperative sharing and creation of metadata by libraries, http://corc.oclc.org.

The Open Directory Project is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors and incorporates RDF technology, http://dmoz.org.

Another project using RDF is the Publishing Requirements for Industry Standard Metadata (PRISM), www.prismstandard.org.

The MusicBrainz Metadata Initiative (www.musicbrainz.org/MM) is designed to create a portable and flexible means of storing and exchanging metadata related to digital audio and video tracks. The MusicBrainz Metadata Initiative is a content description model for audio and video tracks on the Internet.

Many RDF tools and examples are available at the Forth Institute of Computer Science RDF Suite Web site at http://139.91.183.30:9090/RDF.

Many more projects and resources are available on Dave Beckett's Resource Description Framework (RDF) Resource Guide at www.ilrt.bris.ac.uk/discovery/rdf/resources/#sec-apps.

eXtensible Markup Language (XML)

www.w3.org/XML

XML is a project of the World Wide Web Consortium (W3C). The development of XML is being directed by its XML working group.

Community of use

All users, content providers, and coders of the Web.

Purpose and goals

The eXtensible Markup Language (XML) is an abbreviated version of SGML and a more content-oriented language than HTML. XML helps Web content providers define document types and helps programmers write programs to handle them. It omits all the options and most of the more complex and less-used parts of SGML. In return it is easier to understand and write applications for. XML also is more suited to delivery and interoperability over the Web. It is still a form of SGML, however, so XML files may still be processed in the same way as any other SGML file.

Description

XML has been proclaimed the next-generation upgrade to HTML, the markup language on which most of the Web has been constructed. XML is simpler than SGML and contains more content-related information than HTML, so it might solve the content problems that HTML has not been able to handle.

XML allows users to bring together multiple files to form compound documents. It identifies where illustrations are to be incorporated into text files and the format used to encode each illustration. It provides processing control information to supporting programs, such as document validators and browsers, and it adds editorial comments to a file. XML is not a predefined set of tags—of the type defined for HTML—that can be used to mark up documents, nor is it a standardized template for producing particular types of documents. Because XML is similar to other metadata standards, users must create a document type definition (DTD) to define tag sets that identify the relationships between the various elements that form their documents.

Many other applications are associated with the use of XML. These include the eXtensible Stylesheet Language (XSL) and the three linking and addressing languages: XPath, XPointer, and XLink.

Potential for information organizations

Information professionals are interested XML's potential to organize information. MARC users who have been frustrated by its complexity and limitations see both a threat and an opportunity in the success of XML. XML document type definitions (DTD) linked to MARC have already been set up at the Library of Congress and by other information organizations.

Key projects

Some projects listed below replace MARC using XML technology.

The XML Cover Pages, everything you need to know about XML and its applications, www.oasis-open.org/cover.

"Moving from MARC to XML" (three-part essay), http://ihome.ust.hk/~lblkt/xml/marc2xml.html.

Meta Content Framework (MCF) (a metadata standard using XML), www.textuality.com/mcf/NOTE-MCF-XML.html and www.textuality.com/mcf/MCF-tutorial.html.

Papers about XML (especially recommended on XML and databases), www.rpbourret.com/xml.

James (Java MARC events), www.bpeters.com.

"MARC to XML: An Enhanced Name Authority Record" (to obtain project report, contact author directly), http://senna.sjsu.edu/lmain/Nomen.

XMLMARC, http://xmlmarc.stanford.edu.

eXtensible rights Markup Language (XrML), www.xrml.org.

Organization for the Advancement of Structural Information Organization (OASIS) Designs and develops industry standard specifications for interoperability based on XML, www.oasis-open.org.

Education metadata standards

Educational metadata standards were developed to assist educators, educational vendors, and learners in manipulating and using the electronic environment as a learning tool. The two standards listed below appear to be well-positioned to be the major standards in this area.

Instructional Management System (IMS)

www.imsproject.org

The Instructional Management System (IMS) is sponsored and supported by the IMS Global Learning Consortium, Inc., a nonprofit organization supported by members of a worldwide consortium that includes more than 40 contributing members and more than 200 developer network subscribers.

Community of use

Educational and distributed learning content creators, from K-12 through university.

Purpose and goals

The three main objectives of the IMS are:

- To support the inherently collaborative and dynamic nature of learning
- To develop standards for locating and operating interactive materials
- To build a structure for developing and sharing content among educational courseware developers, educational vendors, and users

IMS Global Learning Consortium, Inc. (IMS), which is the organizational representative of the IMS standard, develops and promotes open specifications for facilitating online distributed learning activities such as locating and using educational content, tracking learner progress, reporting learner performance,

and exchanging student records between administrative systems. IMS has two key goals: to define the technical specifications for interoperability of applications and services in distributed learning, and to support the incorporation of the IMS specifications into products and services worldwide.

IMS promotes the widespread adoption of specifications that allows distributed learning environments and content from multiple authors to work together. IMS is a global consortium with members from educational, commercial, and government organizations. Membership fees provide funding, with organizations choosing to join as either contributing members or network subscribers.

Description

The Instructional Management System (IMS) standard was developed by Educom/EDUCAUSE as a result of a common need among educational institutions for nonproprietary, Internet-based strategies for customizing and managing the instructional process, and for integrating content from multiple publishers in distributed/virtual learning environments. The overall objective of the IMS standard is to enable an open architecture for learning.

The entire standard has been called a course management system, a learning server, and even an integrated learning system. IMS is concerned with standards for learning servers, learning content, and even the commercial integration of these capabilities. Although the idea for the IMS originated in higher education, IMS now encompasses corporate and government entities, K-12 interests, and continuing education opportunities. The IMS standard defines technical specifications that developers and creators of educational products and services should incorporate so that their courseware products are interoperable with one another. As such, the IMS standard is not a software product, but a metadata standard that assists developers and creators of educational products in the production of interactive, Internet-based courseware that is platform-independent. Although everything is proprietary, the goal of the IMS is to standardize on an open solution.

Five specifications have been developed or are under development by the IMS to achieve this objective: the IMS Learning Resources Meta-data Specification, the IMS Enterprise Specification, the IMS Content and Packaging Specification, the IMS Question and Test Specification, and the IMS Learner Profiles Specification. The IMS Learning Resource Meta-data Specification, released in August 1999, creates a uniform metadata standard for learning resources so these resources can be more easily found using metadata-aware search tools that reflect the unique needs of users in learning situations. The IMS Enterprise Specification, released in November 1999, is aimed at administrative applications and services that share data about learners, courses, and performance across platforms, operating systems, and user interfaces.

The IMS Content and Packaging Specification was released in two parts—Part I in February 2000 and Part II in June 2000. This specification makes creating reusable content objects easier, that is, educational courseware components that are useful in a variety of learning systems.

The IMS Question and Test Specification, released in February 2000, addresses the ability to share test items and other assessment tools across different systems. The IMS Learner Information Package, released in March 2001, examines ways to organize learner information so learning systems can respond better to the specific needs of each user. Most of these specifications are still in some Version 1.0 manifestation.

More than 200 participants in the IMS Developers Network, including corporations, universities, publishers, educational software companies, and digital

educational libraries, are under agreement to create and implement IMS-compliant materials and environments. Included are IBM, Microsoft, Sun Microsystems, Princeton University CIT, University of California, Texas A&M University, NetDimensions, the NEEDS project, eCollege, and @learning.

NIST is National Institute of Standards and Technology; **ARIADNE** is the Greek legend name used by European Electronic Libraries Programme Project; **IEEE** is the Institute of Electrical and Electronics Engineers.

Groups within NIST, ARIADNE, and IEEE have been working together since 1997 with the IMS to develop open, market-based standards for online learning, including specifications for learning content metadata. In 1998, IMS and ARIADNE submitted a joint proposal and specification to IEEE. Their recommendations form the basis for the IEEE Learning Object Meta-data (LOM) base document. LOM is a classification for a predraft IEEE Specification.

The IEEE LOM base document defines a set of metadata elements for use in describing learning resources. It includes element names, data types, field lengths, and definitions. The specification also defines a conceptual structure for the metadata, as well as how metadata documents must be organized and how applications must behave to be considered IEEE-conforming.

The IMS community recommended that a select core of elements must be identified from this document to simplify initial implementation efforts. The IMS developed a representation of the metadata in XML (eXtensible Markup Language) and surveyed its member institutions around the world to identify the core elements. The *IMS Meta-data Best Practice and Implementation Guide* (MBPIG) is the result of this survey. MBPIG provides access to four documents that assist in the implementation of the IMS standard. These documents are:

- The IEEE Learning Object Meta-data Base Document
- The IMS Learning Resource Meta-data XML Binding Specification
- The IMS Core and Standard Extension Library
- The IMS Taxonomy and Vocabulary Lists

The IMS MBPIG identifies a minimum set of IEEE metadata elements called the IMS Core. The remaining IEEE metadata elements form the IMS Standard Extension Library (SEL). Choosing this smaller set of elements will foster a base level of metadata interoperability and will enable easier implementation of basic metadata capabilities into software vendors' existing products.

The IMS MBPIG also explains how an application may use the core and extended metadata element., It also includes a sample XML representation and document type definition (DTD) of a conforming metadata record to assist developers with their metadata implementations. View the 19 IMS Core metadata elements and the 67 SEL elements at the official Web site.

Potential for information organizations

Educators will actively incorporate IMS-compliant technology in the development of interactive courseware modules in any number of educational fields and practices in the near future. Those in information organizations may become implementers of the IMS, actively use the IMS standard as a creator of instructional courseware, or even search IMS-compatible software as a user. The IMS metadata standard is on its way to providing platform-independent educational resources.

Key projects

See a list of contributing members and developer networks under the Membership tab at the official Web site that provides links to the various implementations and projects using the IMS metadata standard.

Learning Object Metadata Standard (LOMS)

http://ltsc.ieee.org/wg12

This standard is being constructed and maintained by the Institute of Electrical and Electronics Engineers (IEEE) Learning Technology Standards Committee (LTSC) Learning Object Metadata working group.

Community of use

Educational and distributed learning content creators, from K-12 through university.

Purpose and goals

The purpose and goals of the Learning Object Metadata Standard (LOMS) (known as IEEE 1484.12) are as follows (as stated on its official Web site):

- To enable learners or instructors to search, evaluate, acquire, and use learning objects
- To enable the sharing and exchange of learning objects across any technology supported learning systems
- To enable the development of learning objects in units that can be combined and decomposed in meaningful ways
- To enable computer agents to automatically and dynamically compose personalized lessons for an individual learner
- To complement the direct work on standards that is focused on enabling multiple learning objects to work together within a open distributed learning environment; to enable, where desired, the documentation and recognition of the completion of existing or new learning and performance objectives associated with Learning Objects
- To enable a strong and growing economy for learning objects that supports and sustains all forms of distribution: nonprofit, not-for-profit and for profit
- To enable education, training, and learning organizations, government, public, and private, to express educational content and performance standards in a standardized format that is independent of the content itself
- To provide researchers with standards that support the collection and sharing of comparable data concerning the applicability and effectiveness of learning objects
- To define a standard that is simple yet extensible to multiple domains and jurisdictions so that it is easily adopted and applied
- To support necessary security and authentication for the distribution and use of learning objects

Description

LOMS is a basic data element set used to describe learning objects or educational courseware. These data elements are divided into categories. The base scheme consists of nine categories: general, lifecycle, meta-metadata, technical, educational, rights, relation, annotation, and classification. Each of these categories has its own set of data elements. For each data element in these categories, the standard defines its name, explanation, size, order, value space, and data type, and gives an example.

In defining the standard, a learning object must be defined. Learning objects are special units based on the author's intent for encapsulating the data or

program object. The intent of the author, designer, or learning engineer is to use the object for development of a learning environment or path of learning activities. The intent of the author on the use of the learning object establishes its initial classification for metadata. As the object is used, it may prove ineffective or useful. (Many objects, such as a calculator tool object to be used in math assignments, may be used in nonlearning situations as well, for accounting or financial work.)

Multiple points of view are crucial, as a teacher, student, or designer may have different perspectives in regard to the same content. For example, a teacher may be concerned with the presentation mechanisms of the object, a student with the actual functions of the object, and the designer with the properties or characteristics that support the needs of both the learner and the teacher.

Learning objects can have three elements: a learning objective, a learning exercise, and an assessment. Others have defined learning objects as anything used by a learner or created by a course developer for learning with learning. Objects can be as large as a book with hundreds of pages.

LOMS is the predominant metadata standard for digital and nondigital educational resources. The Instructional Management System (IMS) metadata standard is built on the IEEE LOMS standard, and efforts are underway to coordinate the activities and work of the two groups.

Potential for information organizations

As with IMS, LOMS will become an important metadata standard assisting in the interoperability and interactivity of educational learning objects and their distribution in the online environment. Information organizations will need to help set up mechanisms for retrieval and creation using this metadata standard. A Dublin Core/LOMS/IMS crosswalk is in place to aid practitioners in this endeavor at the official Web site.

Key projects

Alliance of Remote Instructional Authoring and Distribution Networks for Europe, (ARIADNE), http://ariadne.unil.ch/Metadata.

Alexandria Digital Earth Prototype (ADEPT), http://alexandria.sdc.ucsb.edu/~acoleman/alommv3.html.

American Mathematics Metadata Task Force (AMMTF), http://mathmetadata.org/ammtf/projects.html.

Various LOMS translations to other languages, www.cenorm.be/isss/Workshop/lt/LOM-localization/default.htm.

An article on various educational metadata standards, www.learnativity.com/standards.html.

Media-specific metadata standards

Media-specific metadata standards concern various medias available in electronic form. Medias such as art, music, audio, moving pictures, streaming video all have metadata standards that have been developed to assist in their description, transport, and rights management over the Internet.

Digital audio standards

www.aes.org

Most digital audio standards are issued by the Audio Engineering Society (AES).

AES Standards Committee,
http://aessec.aessc.aes.org

Community of use

Digital audio community.

Purpose and goals

The Audio Engineering Society (AES) is devoted exclusively to audio technology, and its membership consists of those involved in related activities. AES encourages and implements standards related to audio technology through its Standards and Technical Council committees.

Description

Many digital audio standards have been issued by AES and are available at its official Web site, www.aes.org.

Potential for information organizations

Anyone involved with digital audio, especially in relation to media centers, should be interested in standards set up by the AES.

Key projects

The Colorado Digitization Project produced the best guide to digital audio standards available at http://coloradodigital.coalliance.org/digaudio1.pdf.

The Council on Library and Information Resources (CLIR) report "Building a National Strategy for Digital Preservation: Issues in Digital Media Archiving" is a excellent set of six reports on preservation issues in digital media:

- *Peter Lyman, Archiving the World Wide Web, www.digitalpreservation.gov/ndiipp/repor/repor_back_web.html*
- *Dale Flecker, Preserving Digital Periodicals, www.digitalpreservation.gov/ndiipp/repor/repor_back_perio.html*
- *Samuel Brylawski, Preservation of Digitally Recorded Sound, www.digitalpreservation.gov/ndiipp/repor/repor_back_sound.html*
- *Frank Romano, E-books and the Challenge of Preservation, www.digitalpreservation.gov/ndiipp/repor/repor_back_ebooks.html*
- *Mary Ide, Dave MacCarn, Thom Shepard, and Leah Weisse, Understanding the Preservation Challenge of Digital Television, www.digitalpreservation.gov/ndiipp/repor/repor_back_tv.html*
- *Howard D. Wactlar and Michael G. Christel, Digital Video Archives: Managing through Metadata, www.digitalpreservation.gov/ndiipp/repor/repor_back_archi.html*

European Broadcasting Union (EBU) Project Metadata (P/META)

www.ebu.ch/pmc_meta.html

Sponsored by the European Broadcasting Union.

Community of use

Media and broadcasting industry.

Purpose and goals

The goals of the European Broadcasting Union Project Metadata (EBU P/META) are:

- To establish understanding among EBU members of the media-related data interchange requirements of media commissioner and publishers (broadcasters), suppliers (producers) and consumers, using the BBC Standard Media Exchange Framework (SMEF) as the core information architecture
- To validate and extend the SMEF model as appropriate against members' requirements in terms of data and process, noting local synonyms (translations)
- To create a European SMEF—an E-SMEF—that would extend the thinking to the development of a commercial process framework for exchange of media among EBU members using E-SMEF to apply emerging Society of Motion Picture and Television Engineers (SMPTE) metadata standards to the production and broadcast or distribution process and to study the feasibility of creating and adopting common exchange formats for metadata
- To establish understanding of the use of unique identifiers in metadata as a crucial linkage tool between unwrapped data (metadata) and wrapped or embedded metadata in media files or streams and to develop protocols for their management among members—as an aid to commercial and system interoperability between members and in cooperation with standards bodies in related industries such as music and print publishing
- To collate all relevant unique identifier schemes and map them against one another, which could be in done in collaboration with the EU in*d*ecs project and the International DOI Foundation and extend to cover their data models

International DOI Foundation, www.doi.org.

Description

The EBU P/META project has identified a set of core information required to support the exchange of broadcasting program material among parties. The standard is built on the BBC's SMEF. EBU P/META is linked extensively to this standard, as well as to the MPEG-21 effort (see that section later in this chapter). The EBU P/META attribute set is a list of more than 200 tags used in its MetadataScheme.

According to the official Web site, this standard is in Version 0.9 (February 2001). A suite of documents and the standard are available via FTP after a registration process. These documents include a MetadataScheme presentation, an EBU P/Meta Scheme Notation, a set of HTML files (accessed by hyperlink from within the MetadataScheme Presentation document), and a README.TX file.

To review current documents regarding this standard, the EBU Secretariat must be contacted to receive a password; her name is Chris Blanc, and her email is blanc@ebu.ch.

Potential for information organizations

Metadata standards in the media and broadcasting industry are in flux and need to be monitored by information organizations.

Key projects

For a discussion of metadata in broadcast-related media, see http://www-1.ibm.com/industries/media/pdf/MECollateralNAB2002DTVWhitepaper.pdf.

The Society of Motion Picture and Television Engineers (SMPTE) (www.smpte.org) leads the effort to develop standards within the media industry. View its official Web site at www.smpte.org and a short document on its work at www.indecs.org/washington/wrappers.pdf.

Metadata for art

The major metadata standards constructed for art objects have as their major purpose and goals the construction of a conceptual framework for the description and access of art objects and images. No official Web site exists for these standards. Most of these standards have been developed in the last 10 years. The key standard is the Categories for the Description of Works of Art (CDWA), which is considered the parent standard for many art metadata efforts, including the VRA Core Categories and Object ID.

Potential for information organizations

Anyone involved in the description and access of art objects and images should be knowledgeable about the various metadata standards in this area. Some have been adopted more readily than others; as yet, none has been adopted as the single standard.

Key projects

Categories for the Description of Works of Art (CDWA), www.getty.edu/research/institute/standards/cdwa/index.html.

Consortium for the Computer Interchange of Museum Information (CIMI), www.cimi.org/now.html.

DIG35 metadata standard (International Imaging Industry Association)(I3A), www.i3a.org/i_dig35.html.

International Guidelines for Museum Object Information: The CIDOC Information Categories, www.cidoc.icom.org/guide/guidecat.htm and www.cidoc.icom.org/guide0.htm.

Museum Computer Network (MCN) metadata standards page, www.mcn.edu/stand_subscribe.htm.

RLG REACH element set for shared description of museum objects, www.rlg.org/reach.elements.html.

Technical metadata for digital still images (NISO Standards committee), www.niso.org/committees/committee_au.html.

ALCTS Networked Resources and Metadata (NRMC) Committee, Metadata for Art Research Materials report (1999), www.ala.org/alcts/organization/div/nrmc/artmeta.html.

Visual Arts, Museums and Cultural Heritage Metadata Workshop 1997 Report, http://vads.ahds.ac.uk/reports/Metadata_workshop/metadata_index.html.

Metadata for music notation

Metadata for music notation has no official Web site or sponsoring agency. See www.lib.ox.ac.uk/immpwg for the most up-to-date description of music metadata standards. The sites are maintained by the Music Library Association (MLA).

No official metadata standard exists that is focused exclusively on music, but the related portions of the site adapt some metadata standards to present music notation on the Web, and some markup languages specifically address digital music notation in score format.

No centralized authority or Web site for information regarding metadata standards exists for music notation or the presentation of music on the Web. The International Music Metadata Projects Working Group (IMMPWG) within the Music Library Association (MLA) attempts to document and keep current with various metadata standards that are associated with music. This group is charged with the following tasks:

- Survey the current landscape to identify music metadata projects.
- Monitor developments in existing music metadata projects and identify new projects.
- Prepare a written and verbal report for the MLA annual meeting in New York City (2001) that inventories these projects, explains the meaning and relevance of them to music libraries, and apprises the membership of particularly noteworthy activities.
- Make recommendations to the BCC for any appropriate actions that should be taken, both to keep music librarians informed of important developments and to ensure music librarianship is represented in those developing areas as needed.

The working group was asked to take a broad view of the terms *music* and *metadata*, to include textual information about music, musical notation, musical sound files, and music video or still images and survey and learn broadly about various metadata standards, not restricting itself to one or two specific standards.

The first report of the IMMPWG resulted in a list of 115 digital projects that included or incorporated some aspect of music. Phase II of the working group narrowed this list to 41 projects grouped under six topics: MARC, EAD, Dublin Core, Music metadata markup languages, American Memory projects, and a Defined Parameters project. This report was presented at the MLA conference in February 2002.

Four music metadata markup standards exist: Standard Music Description Language (SMDL), MusicXML, Music Markup Language Project, and MusiCat. Links to these Web sites are available from the IMMPWG project report under the Key Projects section below.

Maintaining currency with metadata standards in this area is key for those information organizations working in the areas of sound, music, and multimedia. Most digital projects that involve music work with Dublin Core, EAD, and MARC. The American Memory projects have been built on a combination of two or more of these standards along with SGML and TEI; the METS metadata standard will be incorporated in future American Memory projects. Those who are involved with digitizing musical scores or musical notation will want to keep track of the various music metadata markup languages.

Key projects

International Music Metadata Projects Working Group (IMMPWG), www.lib.ox.ac.uk/immpwg.

HARMONICA – Survey of existing music projects (last updated 1999), www.svb.nl/project/harmonica/harm_survey.htm.

Standard Music Description Language (SMDL), www.techno.com/smdl.htm.

MusicXML, www.recordare.com/xml.html.

Music Markup Language (MML), www.mmlxml.org.

Multimedia Content Description Interface (MPEG-7)

http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm

This site is maintained by the Moving Pictures Expert Group (MPEG), an ISO/International Engineering Consortium (IEC) working group.

Community of use

Digital audio and video community.

Purpose and goals

The MPEG-7 standard provides a rich set of standardized tools to describe multimedia content. MPEG-7 aims to standardize a language to define description schemes and descriptors, called the Description Definition Language (DDL), to create a core set of descriptors to describe the various features of multimedia content; to provide coded representations of descriptions to enable efficient storage and fast access; and to assist in predefining structures of descriptors and their relationships, called description schemes (DSs).

Description

The "Multimedia Content Description Interface," popularly known as MPEG-7, describes audiovisual information regardless of its storage, coding, display, transmission, medium, or technology. MPEG-7 meets the needs of many media types such as graphics, audio, video, photographs, 3D models, speech, and type combinations. It handles both retrieval from digital archives and filtering of streamed audiovisual broadcasts on the Internet. The standard operates in both real-time and nonreal-time environments.

MPEG-7 addresses the proliferation of audiovisual and media formats in digital form. It is an ISO/IEC standard developed by MPEG that addresses applications in which video and audio content can be stored or streamed. The focus of the standard is on description elements of multimedia data.

The MPEG-7 standard consists of the following parts:

- MPEG-7 systems—the binary format for encoding MPEG-7 descriptions and the terminal architecture
- MPEG-7 Description Definition Language (DDL)—the language for defining the syntax of the MPEG-7 description tools and for defining new description schemes
- MPEG-7 visual—the description tools dealing with (only) visual descriptions
- MPEG-7 audio—the description tools dealing with (only) audio descriptions
- MPEG-7 multimedia description schemes—the description tools dealing with generic features and multimedia descriptions
- MPEG-7 reference software—a software implementation of relevant parts of the MPEG-7 standard with normative status
- MPEG-7 conformance—guidelines and procedures for testing conformance of MPEG-7 implementations (under development)

- MPEG-7 extraction and use of descriptions—informative material (in the form of a Technical Report) about the extraction and use of some of the description tools (under development).

The five major functionalities of MPEG-7 are listed as the first five bullets above. The multimedia description schemes comprises the bibliographic metadata section. The MPEG-7 Visual section has five visual-related basic structures: the grid layout, the time series, multiple view, the spatial 2d coordinates, and temporal interpolation. The MPEG-7 Audio section has 17 temporal and spectral descriptors that may be used in a variety of applications. The MPEG-7 DDL is an XML schema that is divided into three groups of components. A more detailed description of the MPEG-7 standard is available from the official Web site.

Potential for information organizations

MPEG standards have become the model for the establishment of standards in the multimedia community. MPEG-7 appears to be well situated to assist in the standardization of descriptive metadata for the digital audio and video environment, and so information organizations should be prepared to support it.

Key projects

"MPEG-7: Behind the Scenes" by Jane Hunter, www.dlib.org/dlib/september99/hunter/09hunter.html.

The MPEG-7 DDL home page, http://archive.dstc.edu.au/mpeg7-ddl.

"Moving Pictures Expert Group: MPEG-7 Standard" by Robin Cover (links to many projects using MPEG-7), www.oasis-open.org/cover/mpeg7.html.

The MPEG-7 schema page, http://pmedia.i2.ibm.com:8000/mpeg7/schema.

Multimedia Content Description Framework (MPEG-21)

http://mpeg.telecomitalialab.com/standards/mpeg-21/mpeg-21.htm

This site is maintained by the Moving Pictures Expert Group (MPEG), an ISO/IEC working group.

Community of use

Digital audio and video community.

Purpose and goals

MPEG-21 integrates critical electronic technologies, enabling transparent and augmented use of multimedia resources across a wide range of networks and devices to support functions such as: content creation, content production, content distribution, content consumption and usage, content packaging, intellectual property management and protection, content identification and description, financial management, user privacy, terminals and network resource abstraction, content representation, and event reporting.

Description

The "Multimedia Content Description Framework," or MPEG-21, will identify and define the key elements needed to support the multimedia delivery chain from production to presentation to delivery to archiving. MPEG-21 will elaborate the elements of the MPEG-7 standard by defining the syntax and semantics of its characteristics, such as interfaces to the elements. MPEG-21 also addresses the necessary framework functionality, such as the protocols associated with the interfaces, mechanisms to provide a repository, composition, and conformance.

Seven key elements make up the MPEG-21 standard:

- **A digital item declaration:** A uniform and flexible abstraction and interoperable schema for declaring or naming digital items
- **A digital item identification and description:** A framework for identifying and describing any entity regardless of its nature, type or granularity
- **A content handling and usage component::** Provides interfaces and protocols that enable creation, manipulation, search, access, storage, delivery, and (re)use of content
- **An intellectual property management and protection component:** Provides the means to enable content to be persistently and reliably managed and protected across a wide range of networks and devices
- **Terminals and networks:** Provides interoperable and transparent access to content across networks and terminals
- **A content representation component:** Indicates how the media resources are represented
- **An event reporting module:** Provides the metrics and interfaces that enable users to precisely understand the performance of all reportable events within the framework

The MPEG-7 standard can be described as bibliographic metadata. The MPEG-21 standard is more of a packaging standard, similar to a shipping manifest. This idea relates closely to what other metadata standards such as METS and OAIS are doing, and it is especially useful for preservation purposes. One important component of the MPEG-21 standard is the Digital Item Declaration Language (DIDL), which is similar to the METS metadata standard.

Potential for information organizations

Those involved in the packaging, distribution, and preservation of audio and video objects will be especially affected by the MPEG-21 standard. Given the quick production schedules of most MPEG standards, this standard will most likely be developed and implemented quickly as well.

Key projects

No current projects, but view the paper "From MPEG-1 to MPEG-21: Creating an Interoperable Multimedia Infrastructure" by Rob Koenen at http://mpeg.telecomitalialab.com/documents/from_mpeg-1_to_mpeg-21.htm.

Standard Media Exchange Format (SMEF)

www.bbc.co.uk/guidelines/smef

The British Broadcasting Corp. (BBC) directs the Standard Media Exchange Format (SMEF).

Community of use

Media and broadcasting industry.

Purpose and goals

The Standard Media Exchange Format Data Model (SMEF-DM) integrates key information systems with an appropriate systems architecture and harmonizes other related systems as appropriate to improve commonality of data. SMEF-DM also will be used to manage data definitions across media applications and embedded metadata definitions in media formats.

Description

SMEF-DM is available in Version 1.7. To obtain the standard, you must read and sign an extensive license agreement and send the information in an e-mail to the BBC. Access to and use of the standard is free, but you must agree to the extensive contract as written before being allowed to even view the standard.

Potential for information organizations

Not known.

Key projects

See EBU P/META information.

Visual Resources Association (VRA) Core Categories

www.vraweb.org/vracore3.htm

The Visual Resources Association (VRA) is maintained and supported by the Visual Resources Association (VRA) Data Standards Committee.

Community of use

Digital graphics creators, visual culture curators, and digital image collections and organizers.

Purpose and goals

Version 2.0 of the VRA Core Categories is available at http://severn.dmu.ac.uk/elise/el2_dels/vra_guide.htm.

The VRA Core Categories standard (Version 3.0) is not intended as a complete application but as a point of departure to facilitate the sharing of information among visual resources collections about works and images. This standard is a single element set that can be applied as many times as necessary to create records to describe works and images of visual culture.

Description

The 17-element set VRA Core Categories Version 3.0 is a data dictionary. The element set allows for the use of qualifiers similar to the Dublin Core metadata standard. It was constructed around the concepts of what a *work* and an *image* are. Given that the world of digital images and objects is much larger than mere visual resources collections, other metadata standards are being developed without reference to the VRA Core Categories. No priority of importance is given to any of the 17 elements, and the order of the elements can be applied arbitrarily.

Potential for information organizations

This metadata element set was developed to closely follow Dublin Core. For those involved in maintaining and describing visual resources collections, the core categories may be a viable standard to use. Implementation at the local level is up to each organization.

Key projects

ALCTS Committee on Cataloging: Description and Access (CC:DA) Task Force on VRA Core Categories, www.ala.org/alcts/organization/ccs/ccda/tf-vra1.pdf.

"Towards a Shared Cataloguing Tool for VR Collections," www.diglib.org/collections/vrtool/toolframe.htm.

Preservation metadata standards

Preservation metadata standards manage and preserve digital objects for the long term. Although long-term preservation is a concern of most metadata standards, these particular standards attempt to either function as additional components attached to other metadata standards or as containers that wrap around the metadata to assist in its documentation and management.

Metadata Encoding and Transmission Standard (METS)

www.loc.gov/standards/mets

The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress and is being developed as an initiative of the Digital Library Federation.

Community of use

Digital library implementers, librarians, and metadata creators.

Purpose and goals

The Metadata Encoding and Transmission Standard (METS) is a direct result and continuation of the Library of Congress's Making of America II project (MOA2). MOA2 is a Digital Library Federation project to create a proposed digital library object standard by encoding defined descriptive, administrative and structural metadata, along with primary content, inside a digital library object. The cornerstone of the MOA2 effort is the production of an XML document type definition (DTD) that defines the digital object's elements and encoding.

The project also has developed a relational database that allows a library to capture the metadata created by an MOA2 library—a program that reads the database of captured metadata. It automatically creates XML-encoded digital objects, a repository manager that provides distributed network access to the objects, and a viewer that displays MOA2 objects from the repository. METS provides an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users).

Description

METS is an encoding format for descriptive, administrative, and structural metadata for textual and image-based works within a digital library. It also serves as a transmission standard. A METS document consists of four major components: descriptive metadata, administrative metadata, file inventory, and structural map. The file inventory and structural map are required by the schema. The descriptive metadata is optional, and it is used for identification and discovery by those metadata creators who have not already chosen one of the other metadata standards to work with.

METS can be internal, point to an external source, or do both. It can point to internally embedded descriptive metadata (attached to the digital object), external metadata such as a MARC record or EAD finding aid, or both. Administrative metadata is optional and has four optional subcomponents: rights metadata, preservation metadata, source metadata, and technical metadata.

The METS file inventory is a listing of all the files associated with a digital object, both internal and external. Files can be grouped into areas such as master files, thumbnails, and so on. The structural map is a tree structure that describes the digital object. The map can be simple or complex, but it needs to outline a hierarchical structure linking metadata and content files to the digital object.

Many working groups are involved in METS production, even though this standard emerged recently. These groups are working in the areas of audio, video, image, and text technical metadata extensions of the METS standard.

Potential for information organizations

The development of METS is being supported and enabled by the Library of Congress. Many information professionals hope that its applicability and usefulness to libraries, especially in the digital environment, will be immense. MOA2 projects were innovative and daring in their use of multiple metadata formats and numerous digital objects in many formats, and METS is attempting to bridge the project's interoperability gap as well as the problems and concerns that digital libraries bring to the information landscape. The Research Libraries Group (RLG) will soon take over as the new sponsoring organization and coordinator of METS.

Key projects

Library of Congress Audio-Visual Prototyping Project—motivated by the decreasing viability of analog copying, this project is developing approaches for the digital reformatting and preservation of historical sound recordings and moving image collections, http://lcweb.loc.gov/rr/mopic/avprot/metsmenu.html.

University of California, Berkeley—most of UCB's digital library projects already use the MOA2 document type definition (DTD). METS is a natural evolution for current and future projects as well as digital objects contained in the California Digital Library (CDL), digital library projects, http://sunsite.berkeley.edu.

The METAe project, a European initiative, is creating an automated metadata extraction system for books and journals and plans to use METS as the basis for its construction, http://meta-e.uibk.ac.at.

METS Java Toolkit is a site that describes a prototype alpha Java toolkit produced at Harvard University to help construct, serialize, validate, and deserialize METS files, http://hul.harvard.edu/~stephen/mets.html.

Open Archival Information System (OAIS)

http://ssdoo.gsfc.nasa.gov/nost/isoas

Donald Sawyer, head of the NASA Science Office of Standards and Technology, was the early creator and influence of this standard. Substantial collaboration with RLG and OCLC is in process.

Community of use

Librarians, archivists, and anyone involved in preservation and long-term archiving of digital objects and information.

Purpose and goals

The Open Archival Information System (OAIS) is a conceptual framework for an archival system that preserves and maintains long-term access to digital information. The reference model that the standard has created increases awareness and understanding of concepts relevant for archiving digital objects, especially

among nonarchival institutions; elucidates terminology and concepts for describing and comparing data models and archival architectures; expands consensus on the elements and processes endemic to digital information preservation and access; and creates a framework to guide the identification and development of standards.

Description

OAIS is regarded as a best-practice model for the long-term preservation and archiving of digital objects and associated metadata. An OAIS archive is expected to meet these certain minimum requirements:

- Negotiate and accept appropriate information from information producers
- Obtain sufficient control of the information to ensure long-term preservation
- Determine the scope of its specified class of users (called the Designated Community)
- Ensure the information can be understood by the Designated Community without the assistance of the information producers
- Follow documented policies and procedures to ensure the information is preserved against reasonable contingencies, and to enable the information to be disseminated as authenticated copies of the original or as traceable to the original
- Make the information available to the Designated Community

The OAIS environment is driven by the interaction of four entities: producers, consumers, management, and the archive itself. Producers supply the information that the archive preserves. Consumers use the preserved information. Management is responsible for establishing the broad policy objectives of the archive (for example, determining what types of information are to be archived and identifying funding sources). The management entity does not include the day-to-day administration of the archive; administration is performed by a designated group of people.

The OAIS archive model contains three types of information packages: the Submission Information Package (SIP), which is sent from the information producer to the archive; the Archive Information Package (AIP), which is the information package actually stored by the archive; and the Dissemination Information Package (DIP), which is the information package transferred from the archive in response to a request by a consumer.

Five functional software units use these information packages:

- The Ingest unit receives information from producers and prepares the information for storage and management within the archive.
- The Archival Storage function handles the storage, maintenance, and retrieval of the AIPs held by the archive.
- The Data Management function coordinates the Descriptive Information pertaining to the archive's AIPs, in addition to system information used in support of the archive's operation.
- The Administration function manages the day-to-day operation of the archive.
- The Access function helps consumers identify and obtain descriptions of relevant information in the archive, and it delivers information from the archive to consumers.

In the OAIS reference model, technical metadata is referred to as representation information (RI), and descriptive metadata is referred to as preservation description information (PDI). When an Information Package is delivered to an OAIS digital repository, it arrives as a SIP, which contains the digital object and associated metadata. The SIP is enhanced as necessary by the requesting agency or person on submission, and encapsulated as an AIP. When a user requests access to an object, a DIP is provided, which contains a copy of the digital object and its metadata, as well as appropriate mechanisms to retrieve and use the digital object.

Potential for information organizations

RLG and OCLC published a document titled "Trusted digital repositories: attributes and responsibilities" in March 2002. Both organizations agree that OAIS is the best working model for large-scale, heterogeneous collections held by cultural organizations. OCLC held a videoconference in April 2002 that focused exclusively on the OAIS standard, so information organizations will most likely be working with this standard in the future.

Key projects

"Trusted Digital Repositories: Attributes and Responsibilities," an RLG-OCLC report, www.rlg.org/longterm/repositories.pdf.

CURL exemplars of digital archives (CEDARS) project, www.leeds.ac.uk/cedars/index.html.

National Library of Australia, "Preservation Metadata in Digital Collections: Exposure Draft, www.nla.gov.au/preserve/mpeta.html.

Three excellent Web sites are devoted to information on OAIS and other preservation metadata standards:

- *Electronic Resource Preservation and Access NETwork (ERPANET), www.erpanet.org*
- *Council on Library and Information Resources (CLIR), www.clir.org/pubs/pubs.html*
- *Preserving Access to Digital Information, www.nla.gov.au/padi/topics/18.html*

Rights metadata standards

Rights metadata standards deal exclusively with the challenges and problems associated with copyright in the digital arena. Most of these standards would work in tandem and alongside other metadata standards, or in some cases (as standalone modules) be incorporated with other metadata standards.

Common Information System (CIS)

www.cisac.org/cisac/webcontent.nsf/homepage

Common Information System (CIS) is administered by the International Confederation of Societies of Authors and Composers (CISAC).

Community of use

Authors, composers, and those interested in digital rights.

Purpose and goals

The Common Information System (CIS) defines and implements standards to allow uniform data exchange and access between affiliated rights societies such as the American Society of Composers, Authors, and Publishers (ASCAP) and the International Confederation of Societies of Authors and Composers (CISAC), which increases efficiency by reducing duplication of digitizing objects. CIS is a strategic development by rights societies who recognize they must invest today if they are to offer effective protection to their members in the future environment of electronic commerce.

Description

Although much has been accomplished since 1994 when it first began, CIS has been stalled. Working groups evaluate and define standards and pass recommendations to a steering committee. The CIS working groups meet four times a year and are split between general processes and application development. These working groups are: data modeling, data standards and rules, international communications, interested parties information, musical works information, audio-visual work information for musical rights, international documentation on audio-visual works for audio-visual rights, international standard audio-visual number, literary works information, visual works information, sound carrier and recording information, and agreements and schedule information.

Some of the more important milestones for CIS are:

- The International Standards Work Code (ISWC), known as ISO 15707
- The Works Information Database (WID), located at ASCAP in New York
- The CISAC Standards Database (CSD)
- The Interested Party Information (IPI) located at SUISA in Zurich (a Swiss corporation for lyricists, composers, and music publishers).

Potential for information organizations

The ISO 15707 standard (April 2002) provides a system for assigning a unique international identification code to musical works and is of tremendous importance to information organizations to catalog and classify musical works.

Key projects

Press release on ISO 15707, www.iso.ch/iso/en/commcentre/pressreleases/2002/Ref821.html.

Interoperability of Data in E-Commerce Systems <indecs>

www.indecs.org

<in*d*ecs> is a not-for-profit company composed of partners such as Kopiosto, CAL, Editeur, IFPI, MUZE Inc. and IDF—all various national copyright organizations. These companies and copyright agencies all have a vested interest in the success of digital rights initiatives.

Community of use

Rights management, e-commerce companies, and intellectual property domain.

Purpose and goals

<in*d*ecs> develops a data model for intellectual property, including a formal structure for describing intellectual property, the people, and the businesses

that are involved in its trading, and the agreements that they make. The project also develops high-level specifications for the services required to implement a global intellectual property trading system based on this standard.

Description

The Interoperability of Data in E-Commerce Systems, <in*d*ecs>, metadata standard was designed as a fast-track, infrastructure project aimed at finding practical solutions to interoperability affecting all types of rights-holders in a networked, e-commerce environment. Its main concern is the management of intellectual property rights. The <in*d*ecs> schema sets forth four axioms about electronic commerce: metadata is critical, data is complex, metadata is modular, and transactions need automation. Along with these axioms, <in*d*ecs> supports interoperability on at least six levels: across media (such as books and audiovisual), across functions (such as cataloging and workflow), across levels of metadata (from simple to complex), across linguistic and semantic barriers, across territorial barriers, and across technology platforms.

The <in*d*ecs> framework recognizes metadata relating to any types of creation, the integration of descriptive metadata with commercial transactions and rights, and that metadata should be created once and used many times for different purposes. Four principles dictate the development of well-formed metadata that <in*d*ecs> espouses: unique identification, functional granularity, designated authority, and appropriate access.

<in*d*ecs> accounts for three distinct but overlapping perspectives: a general view, a commerce view, and an intellectual property view. A fourth, specific, view is known as the generic attribute structure. The general view recognizes items as either animate (beings) or inanimate (things), and relations as being static (situations) or dynamic (events). The commerce view relates to descriptive metadata and is concerned with how objects are made. The intellectual property view shows people making and using intellectual property in which rights are owned.

The <in*d*ecs> framework includes a metadata model, principles for mappings to other schemas, a Directory of Parties Proposal, and a detailed metadata dictionary of terms used in the standard. Some of the schemas that have been mapped to <in*d*ecs> include: Dublin Core, Functional Requirements for Bibliographic Records (FRBR), DCMS (recording industry standard), the SMPTE data dictionary (audiovisual community), ONIX (publishers), and Common Information System (CIS) (author's rights).

Potential for information organizations

This metadata standard focuses on intellectual property rights and e-commerce. The influence of the <in*d*ecs> project will be felt in all areas related to intellectual property rights in the digital environment, making knowledge of the standard critical, especially for those people involved in digital and digitization projects that incorporate copyrighted materials.

Key projects

EDItEUR ONIX International, www.editeur.org/onix.html

EDItEUR, the international booktrade e-commerce standards organization, is implementing ONIX International along with <in*d*ecs> as a mechanism for the communication of rich bibliographic data between the different partners in the supply chain.

The International DOI Foundation is the leading identification system for intellectual property in the network environment. It is in the process of imple-

menting <in*d*ecs>-based metadata standards to ensure the future interoperability of its metadata. www.doi.org/welcome.html

Muze, Inc. is a major supplier of descriptive data to global e-commerce vendors. It has remodeled its database design on <in*d*ecs> principles to permit the management of descriptions of all types of media—books, audio and audiovisual—in a single unified database structure. www.muzebiz.com

Open Digital Rights Language (ODRL)

http://odrl.net

Little content regarding this standard, its development, and history, is available. The sponsoring agency appears to be the ODRL Initiative.

Community of use

Digital rights community.

Purpose and goals

Open Digital Rights Language (ODRL) fosters support for open and free standards for the specification of media commerce rights languages.

Description

The Open Digital Rights Language (ODRL) attempts to develop a freely available media commerce rights language that has no licensing requirements. ODRL seems to be a response to the XrML metadata standard, which requires licensing and user fees. A Version 1.1 specification of ODRL was available as of June 2002. The specification is available in both PDF and HTML, and XML schemas for a data dictionary and expression language also are available. ODRL intends to support and interoperate with the MPEG-21 multimedia framework.

Potential for information organizations

The area of digital rights metadata standards is in flux and inundated with many competing standards. Information organizations should be ready to work with and incorporate whatever emerges as the industry standard. Although ODRL is well-positioned to be a leader among digital rights metadata standards, only time will tell whether it will be implemented and achieve widespread acceptance.

Key projects

None available, but many supporters are listed on the official Web site.

Secure Digital Music Initiative (SDMI)

www.SDMI.org

Secure Digital Music Initiative (SDMI) is a forum of more than 200 companies that are interested in this standard.

Community of use

Companies and people who play, store, and distribute digital music.

Purpose and goals

SDMI is a forum that brings together the worldwide recording, consumer electronics, and information technology industries to develop open technology specifications for protected digital music distribution. The specifications re-

leased by SDMI will ultimately provide consumers with convenient access to music in both online and emerging digital distribution systems, enable copyright protection for artists' work, and promote the development of music-related businesses and technologies.

Description

Part in a standard indicates something is complete and pusblished. **Phase** indicates indicates a process, not a published work.

SDMI work is defined into two phases. The first phase is a standard for portable devices, which has already been accomplished. The second, longer-term phase is working toward completion of an overall architecture for delivery of digital music in all forms. SDMI is not producing a single metadata format, technology, or design, but the framework allows a variety of competing technologies and download formats to be used within its system. In essence, SDMI is more of a transportation model for digital music rights.

The reference model of the SDMI Portable Device Specification Part I (Version 1.0, released in February 1999) consists of three layers: an application layer where all SDMI-compliant applications reside, a licensed compliant module (LCM) layer that transfers content to portable devices (PD) that use one or more digital music formats, and a portable device layer that identifies and receives only protected content from the LCM-PD interface.

See more information about this specification at www.SDMI.org/download/port_device_spec_part1.pdf.

SDMI is not in competition with the MP3 file format; but where MP3 players can only play MP3 content, SDMI-compliant devices will be able to play content from both SDMI-compliant and noncompliant devices. SDMI has decided not to initiate Phase 2, since SDMI lacks consensus from participants regarding any combination of the proposed technologies.

Potential for information organizations

If broadly adopted, SDMI-compliant PDs will rapidly proliferate and be used to download and play digital music. Information organizations should be prepared to integrate this technology into their environments.

Key projects

No implementations of this standard are available, but some informative PowerPoint presentations are available at www.SDMI.org/doc_archive.htm, specifically on the dates Feb. 26, 1999, and Sept. 23, 1999.

eXtensible Media Commerce Language (XMCL)

www.xmcl.org

eXtensible Media Commerce Language (XMCL) is supported by a broad range of well-known industry-leading technology and media companies led by RealNetworks, Inc.

Community of use

Those interested and involved in digital media commerce.

Purpose and goals

XMCL is an open, XML-based language designed to establish industry-wide standards for Internet media commerce. XMCL was created for the interchange of commercial and industry standards to be applied to media between business systems (Web businesses) and trusted delivery and playback systems. XMCL describes the minimum, self-complete set of business rules under which digital media. The document is licensed for consumer use. The goal is to have an XMCL document generated whenever a business system authorizes a customer transaction for digital media, which is then submitted to the trusted system.

Description

Digital Rights Management (DRM) software packages are not capable of providing the broad range of services that media businesses require to create, package, publish, distribute, license, and sell digital media over the Web. XMCL assists media owners and distributors by providing the interchange language necessary for all these functions to be accomplished in one software package.

XMCL is a rights specification language that is built using the XML markup language. XMCL combines Dublin Core and TEI syntax and language.

Potential for information organizations

The issue of patrons downloading (either free or for a fee) multimedia digital items over the Internet has become important for libraries, so the adoption of this standard by media companies should be closely monitored in regards to liability and accessibility concerns. XMCL is only one of many initiatives in the area of digital rights management.

Key projects

No products support XMCL, but 25 companies have announced their interest since the announcement in 2001.

eXtensible rights Markup Language (XrML)

www.xrml.org

ContentGuard, Inc., is the main originator and sponsor of this standard.

Community of use

Those people implementing rights management in the digital environment.

Purpose and goals

The goal of eXtensible rights Markup Language (XrML) is to expand the usefulness of digital content, resources, and services to rights holders and users by providing a flexible, extensible, and interoperable industry standard language that is platform-, media-, and format-independent.

Description

The eXtensible rights Markup Language (XrML) is a general-purpose, XML-based specification for expressing rights and conditions associated with digital content, resources, and services. XrML was submitted to the MPEG working group (ISO/IEC) in November 2001 as a proposal for standardization. Users of XrML have to register with ContentGuard, Inc., to use XrML, since much of the software that uses XrML is patented by this company. XrML is extensible and interoperable since it is built using the XML markup language. XrML is available in Version 2.0 (November 2001).

Potential for information organizations

As with other rights management metadata initiatives, information organizations should monitor and familiarize themselves with this standard.

Key projects

The digital rights marketplace is examining many proposed standards in this area. No company has adopted XrML or any of the other major digital rights standards. Until broad experimentation and practical implementations of one or more of the many digital rights metadata standards is initiated, determining which ones will emerge as industry best practices in the future is difficult.