Understanding Metadata

Acknowledgements

Understanding Metadata is a revision and expansion of *Metadata Made Simpler*: *A guide for libraries* published by NISO Press in 2001. NISO Press extends its thanks and appreciation to Rebecca Guenther and Jacqueline Radebaugh, staff members in the Library of Congress Network Development and MARC Standards Office, for sharing their expertise and contributing to this publication.

About NISO

NISO, a non-profit association accredited by the American National Standards Institute (ANSI), identifies, develops, maintains, and publishes technical standards to manage information in our changing and ever-more digital environment. NISO standards apply both traditional and new technologies to the full range of information-related needs, including retrieval, re-purposing, storage, metadata, and preservation. NISO Standards, information about NISO's activities and membership are featured on the NISO website <http://www.niso.org>.

This booklet is available for free on the NISO website (www.niso.org) and in hardcopy from NISO Press.

Published by:
NISO Press
National Information Standards Organization
4733 Bethesda Avenue, Suite 300
Bethesda, MD 20814 USA
Email: nisohq@niso.org
Tel: 301-654-2512
Fax: 301-654-1721
URL: www.niso.org

ISBN: 1-880124-62-9

Understanding Metadata

What Is Metadata?

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information.

The term metadata is used differently in different communities. Some use it to refer to machine understandable information, while others use it only for records that describe electronic resources. In the library environment, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital. Traditional library cataloging is a form of metadata; MARC 21 and the rule sets used with it, such as AACR2, are metadata standards. Other metadata schemes have been developed to describe various types of textual and non-textual objects including published books, electronic documents, archival finding aids, art objects, educational and training materials, and scientific datasets.

There are three main types of metadata:

- *Descriptive metadata* describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- *Structural metadata* indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- *Administrative metadata* provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that sometimes are listed as separate metadata types are:
  - *Rights management metadata,* which deals with intellectual property rights, and
  - *Preservation metadata,* which contains information needed to archive and preserve a resource.

Metadata can describe resources at any level of aggregation. It can describe a collection, a single resource, or a component part of a larger resource (for example, a photograph in an article). Just as catalogers make decisions about whether a catalog record should be created for a whole set of volumes or for each particular volume in the set, so the metadata creator makes similar decisions. Metadata can also be used for description at any level of the information model laid out in the IFLA (International Federation of Library Associations and Institutions) *Functional Requirements for Bibliographic Records*: work, expression, manifestation, or item. For example, a metadata record could describe a report, a particular edition of the report, or a specific copy of that edition of the report.

Metadata is key to ensuring that resources will survive and continue to be accessible into the future.

Metadata can be embedded in a digital object or it can be stored separately. Metadata is often embedded in HTML documents and in the headers of image files. Storing metadata with the object it describes ensures the metadata will not be lost, obviates problems of linking between data and metadata, and helps ensure that the metadata and object will be updated together. However, it is impossible to embed metadata in some types of objects (for example, artifacts). Also, storing metadata separately can simplify the management of the metadata itself and facilitate search and retrieval. Therefore, metadata is commonly stored in a database system and linked to the objects described.

What Does Metadata Do?

An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, provide digital identification, and support archiving and preservation.

Resource Discovery

Metadata serves the same functions in resource discovery as good cataloging does by:

- allowing resources to be found by relevant criteria;
- identifying resources;
- bringing similar resources together;
- distinguishing dissimilar resources; and
- giving location information.

Organizing Electronic Resources

As the number of Web-based resources grows exponentially, aggregate sites or portals are increasingly useful in organizing links to resources based on audience or topic. Such lists can be built as static webpages, with the names and locations of the resources "hardcoded" in the HTML. However, it is more efficient and increasingly more common to build these pages dynamically from metadata stored in databases. Various software tools can be used to automatically extract and reformat the information for Web applications.

Interoperability

Describing a resource with metadata allows it to be understood by both humans and machines in ways that promote interoperability. Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly.

Two approaches to interoperability are cross-system search and metadata harvesting. The Z39.50 protocol is commonly used for cross-system search. Z39.50 implementers do not share metadata but map their own search capabilities to a common set of search attributes. A contrasting approach taken by the Open Archives Initiative is for all data providers to translate their native metadata to a common core set of elements and expose this for harvesting. A search service provider then gathers the metadata into a consistent central index to allow cross-repository searching regardless of the metadata formats used by participating repositories.

Digital Identification

Most metadata schemes include elements such as standard numbers to uniquely identify the work or object to which the metadata refers. The location of a digital object may also be given using a file name, URL (Uniform Resource Locator), or some more persistent identifier such as a PURL (Persistent URL) or DOI (Digital Object Identifier). Persistent identifiers are preferred because object locations often change, making the standard URL (and therefore the metadata record) invalid. In addition to the actual elements that point to the object, the metadata can be combined to act as a set of identifying data, differentiating one object from another for validation purposes.

Archiving and Preservation

Most current metadata efforts center around the discovery of recently created resources. However, there is a growing concern that digital resources will not survive in usable form into the future. Digital information is fragile; it can be corrupted or altered, intentionally or unintentionally. It may become unusable as storage media and hardware and software technologies change. Format migration and perhaps emulation of current hardware and software behavior in future hardware and software platforms are strategies for overcoming these challenges.

Metadata is key to ensuring that resources will survive and continue to be accessible into the future. Archiving and preservation require special elements to track the lineage of a digital object (where it came from and how it has changed over time), to detail its physical characteristics, and to document its behavior in order to emulate it on future technologies.

Many organizations internationally have worked on defining metadata schemes for digital preservation, including the National Library of Australia, the British Cedars Project (CURL Exemplars in Digital Archives), and a joint Working Group of OCLC and the Research Libraries Group (RLG). The latter group developed a framework outlining types of presentation metadata. A follow-up group, PREMIS (PREservation Metadata: Implementation Strategies)—also sponsored by OCLC and RLG—is developing a set of core elements and strategies for the encoding, storage, and management of preservation metadata within a digital preservation system. Many of these initiatives are based on or compatible with the ISO *Reference Model for an Open Archival Information System* (OAIS).

Structuring Metadata

Metadata schemes (also called schema) are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content. Metadata schemes generally specify names of elements and their semantics. Optionally, they may specify content rules for how content must be formulated (for example, how to identify the main title), representation rules for content (for example, capitalization rules), and allowable content values (for example, terms must be used from a specified controlled vocabulary).

There may also be syntax rules for how the elements and their content should be encoded. A metadata scheme with no prescribed syntax rules is called syntax independent. Metadata can be encoded in any definable syntax. Many current metadata schemes use SGML (Standard Generalized Mark-up Language) or XML (Extensible Mark-up Language). XML, developed by the World Wide Web Consortium (W3C), is an extended form of HTML that allows for locally defined tag sets and the easy exchange of structured information. SGML is a superset of both HTML and XML and allows for the richest mark-up of a document. Useful XML tools are becoming widely available as XML plays an increasingly crucial role in the exchange of a variety of data on the Web.

Metadata Schemes and Element Sets

Many different metadata schemes are being developed in a variety of user environments and disciplines. Some of the most common ones are discussed in this section.

Dublin Core

The Dublin Core Metadata Element Set arose from discussions at a 1995 workshop sponsored by OCLC and the National Center for Supercomputing Applications (NCSA). As the workshop was held in Dublin, Ohio, the element set was named the Dublin Core. The continuing development of the Dublin Core and related specifications is managed by the Dublin Core Metadata Initiative (DCMI).

The original objective of the Dublin Core was to define a set of elements that could be used by authors to describe their own Web resources. Faced with a proliferation of electronic resources and the inability of the library profession to catalog all these resources, the goal was to define a few elements and some simple rules that could be applied by noncatalogers. The original 13 core elements were later increased to 15: *Title*, *Creator*, *Subject*, *Description*, *Publisher*, *Contributor*, *Date*, *Type*, *Format*, *Identifier*, *Source*, *Language*, *Relation*, *Coverage*, and *Rights*.

The Dublin Core was developed to be simple and concise, and to describe Web-based documents. However, Dublin Core has been used with other types of materials and in applications demanding some complexity. There has historically been some tension between supporters of a minimalist view, who emphasize the need to keep the elements to a minimum and the semantics and syntax simple, and supporters of a structuralist view who argue for finer semantic distinctions and more extensibility for particular communities.

These discussions have led to a distinction between qualified and unqualified (or simple) Dublin Core. Qualifiers can be used to refine (narrow the scope of) an element, or to identify the encoding scheme used in representing an element value. The element *Date*, for example, can be used with the refinement qualifier *created* to narrow the meaning of the element to the date the object was created. *Date* can also be used with an encoding scheme qualifier to identify the format in which the date is recorded, for example, following the ISO 8601 standard for representing date and time.

All Dublin Core elements are optional and all are repeatable. The elements may be presented in any order. While the Dublin Core description recommends the use of controlled values for fields where they are appropriate (for example, controlled vocabularies for the Subject field), this is not required. However, working groups have been established to discuss authoritative lists for certain elements such as Resource Type. While Dublin Core leaves content rules to the particular implementation, the DCMI encourages the adoption of application profiles (domain-specific rules) for particular domains such as education and government. An application profile for libraries is being developed by the Libraries Working Group.

Dublin Core Example

Title="Metadata Demystified"

Creator="Brand, Amy"

Creator="Daly, Frank"

Creator="Meyers, Barbara"

Subject="metadata"

Description="Presents an overview of metadata conventions in publishing."

Publisher="NISO Press"

Publisher="The Sheridan Press"

Date="2003-07"

Type="Text"

Format="application/pdf"

Identifier="http://www.niso.org/standards/resources/Metadata_Demystified.pdf"

Language="en"

Because of its simplicity, the Dublin Core element set is now used by many outside the library community— researchers, museum curators, and music collectors to name only a few. There are hundreds of projects worldwide that use the Dublin Core either for cataloging or to collect data from the Internet; more than 50 of these have links on the DCMI website. The subjects range from cultural heritage and art to math and physics. Meanwhile the Dublin Core Metadata Initiative has expanded beyond simply maintaining the Dublin Core Metadata Element Set into an organization that describes itself as "dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for discovery systems."

The Text Encoding Initiative (TEI)

The Text Encoding Initiative is an international project to develop guidelines for marking up electronic texts such as novels, plays, and poetry, primarily to support research in the humanities. In addition to specifying how to encode the text of a work, the *TEI Guidelines for Electronic Text Encoding and Interchange* also specify a header portion, embedded in the resource, that consists of metadata about the work. The TEI header, like the rest of the TEI, is defined as an SGML DTD (Document Type Definition)—a set of tags and rules defined in SGML syntax that describe the structure and elements of a document. This SGML mark-up becomes part of the electronic resource itself. Since the TEI DTD is rather large and complicated in order to apply to a vast range of texts and uses, a simpler subset of the DTD, known as *TEI Lite*, is commonly used in libraries.

It is assumed that TEI-encoded texts are electronic versions of printed texts. Therefore the TEI Header can be used to record bibliographic information about both the electronic version of the text and about the non-electronic source version. The basic bibliographic information is similar to that recorded in library cataloging and can be mapped to and from MARC. However, there are also elements defined to record details about how the text was transcribed and edited, how mark-up was performed, what revisions were made, and other non-bibliographic facts. Libraries tend to use TEI headers when they have collections of SGML-encoded full text. Some libraries use TEI headers to derive MARC records for their catalogs, while others use MARC records as the basis for creating TEI header descriptions for the source texts.

Metadata Encoding and Transmission Standard (METS)

The Metadata Encoding and Transmission Standard (METS) was developed to fill the need for a standard data structure for describing complex digital library objects. METS is an XML Schema for creating XML document instances that express the structure of digital library objects, the associated descriptive and administrative metadata, and the names and locations of the files that comprise the digital object.

The metadata necessary for successful management and use of digital objects is both more extensive than and different from the metadata used for managing collections of printed works and other physical materials. Structural metadata is needed to ensure that separately digitized files (for example, different pages of a digitized book) are structured appropriately. Technical metadata is needed for information about the digitization process so that scholars may determine how accurate a reflection of the original the digital version provides. Other technical metadata is required for internal purposes in order to periodically refresh and migrate the data, ensuring the durability of valuable resources.

METS was originally an outgrowth of the *Making of America II* project, a digitization project of major research libraries that attempted to address these metadata issues, in part by providing an encoding format for metadata for textual and image-based works. The Digital Library Federation (DLF) built on that earlier work to create METS, a standard schema for providing a method for expressing and packaging together descriptive, administrative, and structural metadata for objects within a digital library. Expressed using the XML schema language, METS provides a document format for encoding the metadata necessary for management of digital library objects within a repository and for exchange between repositories.

Metadata in Action

An oral historian makes tape-recordings of interviews with members of a particular ethnic group. Interviewees sign a paper release form giving intellectual property rights to the historian. Most interviewees grant permission to disseminate the interviews in print and electronically, but several restrict publication and dissemination until 25 years after death.

Information about each interview is kept in a database: *Interviewer*, *Interviewee*, *Date*, *Place*, etc. Each interview follows a questionnaire format. The questionnaire exists as a text file. The tapes, release forms, database, and text file are donated to a library that has a special collection focusing on the particular ethnic group.

The tapes are digitized. Since each interview runs over several tapes, technicians record structural metadata to keep component parts of each interview together. Technicians record administrative metadata such as file names, location of each interview in the files, equipment used, the methods of digitizing and assuring quality and completeness, file formats, etc. Different segments of this metadata allow the audio files to be automatically tracked, accessed, stored, refreshed, and migrated.

An archivist expands the database to include the persistent identifier of each interview, thereby linking the audio file to the descriptive metadata. The names of the data elements are revised to match Dublin Core terminology, including qualifiers used specifically for audio

(continued on page 5)

A METS document contains seven major sections:

- *METS Header* – Contains metadata describing the METS document itself, including such information as creator, editor, etc.
- *Descriptive Metadata* – Points to descriptive metadata external to the METS document (for example, a MARC record in an OPAC or an Encoded Archival Description finding aid maintained on a webserver), or to internally embedded descriptive metadata, or both.
- *Administrative Metadata* – Provides information regarding how the files are created and stored, intellectual property rights, the original source object from which the digital library object derives, and the provenance of the files comprising the digital library object.
- *File Section* – Lists all files containing content that comprise the electronic versions of the digital object.
- *Structural Map* – Outlines a hierarchical structure for the digital library object and links the elements of that structure to content files and metadata that pertain to each element.
- *Structural Links* – Allows METS creators to record the nodes in the hierarchy outlined in the Structural Map.
- *Behavior* – Associates executable behaviors with content in the METS object.

The METS header, file section, structural map, structural links, and behavior sections are defined within the METS schema. METS is less prescriptive about descriptive and administrative metadata, relying on extension schemas—externally developed metadata schemes—to provide specific elements. The METS Editorial Board has endorsed three descriptive metadata schemes: simple Dublin Core, MARCXML, and MODS (discussed below).

For technical metadata the METS website makes available schemas for text and digital still images. The latter standard is called MIX, *Metadata for Images in XML Schema*, and is based on a proposed NISO standard, Z39.87, *Data Dictionary: Technical Metadata for Digital Still Images*. Further work is in process on extension schemas for audio, video, and websites. Another current area of concentration for the METS development community is the creation of METS application profiles to give guidance regarding the creation of METS documents for particular object types.

Use of the METS schema is widespread. A list of implementation registries using METS, a tutorial, and other important information can be found on the METS website.

Metadata Object Description Schema (MODS)

The Metadata Object Description Schema (MODS) is a descriptive metadata schema that is a derivative of MARC 21 and intended to either carry selected data from existing MARC 21 records or enable the creation of original resource description records. It includes a subset of MARC fields and uses language-based tags rather than the numeric ones used in MARC 21 records. In some cases, it regroups elements from the MARC 21 bibliographic format. Like METS, MODS is expressed using the XML schema language.

Although the MODS standard can stand on its own, it may also complement other metadata formats. Because of its flexibility and use of XML, MODS may potentially be used as a Z39.50 Next Generation specified format, an extension schema to METS, a metadata set for harvesting, and for creating original resource metadata records in an XML syntax.

Rich description of electronic resources is a particular focus of MODS, which provides some advantages over other metadata

Metadata in Action

(continued from page 4)

materials. Information on rights and permissions is entered.

An archivist creates an EAD finding aid for the audio collection using the database as the core. Portions of the questionnaire text file are incorporated as a rich source of subject keywords. A MARC record is derived from the EAD finding aid and added to OCLC and RLIN.

A webpage is created where researchers can access the finding aid, search the database, and listen to the audio files. Interviews coded as restricted are invisible to the search program until the date when they become open to the public. Administrative, structural, and descriptive metadata is created for the webpage to hold all the pieces together, allow them to be managed, and allow them to be accessed.

The library participates in a metadata harvesting protocol to provide extracts of local metadata in a common format to a service provider so that information about the collection is automatically included in a number of relevant tools such as catalogs and portals.

The webpage is linked to the library's website dedicated to resources about the ethnic group, where it is available to researchers in context with archival and visual materials, digitized secondary sources, etc. Administrative, structural, and descriptive metadata at the website level has also been created.

A MODS Record Example

```
<mods>
    <titleInfo>
        <title>Metadata demystified</title>
    </titleInfo>
    <name type="personal">
        <namePart type="family">Brand</namePart>
        <namePart type="given">Amy</namePart>
        <role>
            <roleTerm authority="marcrelator" type="text">author</roleTerm>
          </role>
    </name>
    <typeOfResource>text</typeOfResource>
    <originInfo>
        <dateIssued>2003</dateIssued>
        <place>
            <placeTerm type="text">Bethesda, MD</placeTerm>
        </place>
        <publisher>NISO Press</publisher>
    </originInfo>
    <identifier type="isbn">1-880124-59-9</identifier>
</mods>
```

schemes. MODS elements are richer than the Dublin Core; its elements are more compatible with library data than the ONIX or Dublin Core standards; and it is simpler to apply than the full MARC 21 bibliographic format. With its use of XML Schema language, MODS offers enhancements over MARC 21, such as the use of an optional ID attribute to facilitate linking at the element level; the ability to specify language, script, and transliteration scheme at the element level; and the ability to embed a rich description of components in the related Item element.

The ability in MODS to give granular descriptions of constituent parts of an object works particularly well with the METS structural map for complex digital library objects.

The Encoded Archival Description (EAD)

The Encoded Archival Description (EAD) was developed as a way of marking up the data contained in finding aids so that they can be searched and displayed online.

In archives and special collections, the finding aid is an important tool for resource description. Finding aids differ from catalog records by being much longer, more narrative and explanatory, and highly structured in a hierarchical fashion. They generally start with a description of the collection as a whole, indicating what types of materials it contains and why they are important. If the collection consists of the personal papers of an individual there can be a lengthy biography of that person. The finding aid describes the series into which the collection is organized—such as correspondence, business records, personal papers, and campaign speeches—and ends with an itemization of the contents of the physical boxes and folders comprising the collection.

Like the TEI Header, the EAD is defined as an SGML DTD. It begins with a header section that describes the finding aid itself (for example, who wrote it) and then goes on to the description of the collection as a whole and successively more detailed information about the records or series within the collection. If individual items being described exist in digital form, the EAD can include pointers to the digital objects. The 2002 version of the EAD DTD provides support for both SGML and XML through the use of defined "switches" for turning off features used only in SGML and turning on features used only in XML. The EAD standard is maintained jointly by the Library of Congress and the Society of American Archivists.

The EAD is particularly popular in academic libraries, historical societies, and museums with large special collections. Many of these collections contain unique materials unavailable elsewhere and often the materials in the collections are not individually cataloged like traditional library materials. By creating searchable EAD finding aids, libraries and archives can increase awareness of their unique collections to the Internet community.

Learning Object Metadata

The IEEE Learning Technology Standards Committee (LTSC) developed the Learning Object Metadata (LOM) standard (IEEE 1484.12.1-2002) to enable the use and re-use of technology-supported learning resources such as computer-based training and distance learning. The LOM defines the minimal set of attributes to manage, locate, and evaluate learning objects. The attributes are grouped into eight categories:

- *General*, containing information about the object as a whole;
- *Lifecycle*, containing metadata about the objects evolution;
- *Technical*, with descriptions of the technical characteristics and requirements;
- *Educational*, containing the educational / pedagogical attributes;

- *Rights*, describing the intellectual property rights and use conditions;
- *Relation*, identifying related objects;
- *Annotation*, containing comments and the date and author of the comments; and
- *Classification*, which identifies other classification system identifiers for the object.

Within each category is a hierarchy of data elements to which the metadata values are assigned. Examples of learning-related metadata elements found in the Education category are *Typical Age Range* (of the intended user), *Difficulty*, *Typical Learning Time*, and *Interactivity Level*.

The IMS Global Learning Consortium has developed a suite of specifications to enable interoperability in a learning environment. Their *Meta-Data Information Model* specification is based on the IEEE LOM scheme with only minor modifications.

E-Commerce – <indecs> and ONIX

Metadata schemas are increasingly being developed to support electronic commerce applications. The <indecs> Framework (Interoperability of Data in ECommerce Systems) was an international collaborative effort supported by the European Commission's Info 2000 Programme. The collaborators were major rights owners, such as publishers and members of the recording industry, who wanted to develop a framework for metadata standards to support network commerce in intellectual property.

The foundation of the <indecs> work is a data model for intellectual property and its transfer. Rather than developing a new metadata scheme, <indecs> sought to develop a common framework to allow various schemes for transactions related to different genres such as music, journal articles, and books to be able to interchange information, particularly that related to intellectual property rights. In order to support this common framework, <indecs> has developed a minimal kernel of required metadata.

Several organizations have built on the <indecs> Framework to develop specific metadata schemas. Among them is the ONIX (Online Information Exchange) International standard. ONIX is an XML-based metadata scheme developed by publishers under the auspices of a number of book industry trade groups in the United States and Europe. The original ONIX specification was a direct response to the enormous growth in online book sales and the realization that books described with images, cover blurbs, reviews, and similar information significantly outsold books without this information. Therefore *ONIX for Books* has elements to record a wide range of evaluative and promotional information as well as basic bibliographic and trade data. *ONIX for Serials* is in development to define serials product metadata at the title, item, and subscription package levels.

While ONIX information was designed for use in the commerce cycle of a publication, it may also provide a source for enrichment of library-created catalog records; the Bibliographic Enrichment Advisory Team (BEAT) project at the Library of Congress is experimenting with this use. ONIX metadata may also be used by libraries in the future for the creation of a beginning bibliographic record. Mappings between *ONIX for Books* and both MARC 21 and UNIMARC have already been created.

Visual Objects – CDWA and VRA

Metadata used to describe visual objects such as a painting or sculpture has its own special requirements. The Art Information Task Force (AITF), developed a conceptual framework for describing and accessing information about objects and images called Categories for the Descriptions of Works of Art (CDWA). Some 30 categories were defined, most with multiple subcategories. Some examples of the specialized descriptive elements relevant to artworks included are: *Orientation*, *Dimensions*, *Condition*, *Inscriptions*, *Conservation Treatment*, and *Exhibition / Loan History*.

Typically, visual resources collections used in teaching art history and similar subjects do not contain original art works but rather slides or photographs of the original art. Metadata for these materials therefore has to accommodate the description of multiple levels of related resources, such as an original painting, a slide of the painting, and a digitized image of the slide. The VRA Core Categories build on and expand the CDWA work to define a single metadata element set that can be used to describe the work (the actual painting, photograph, sculpture, building, etc.) as well as the images (visual representations) of them.

Version 3.0 of the VRA Core Categories consists of 17 metadata elements which can be used as applicable to describe each of these versions and relate them to each other: *Record Type*, *Type*, *Title*, *Measurements*, *Material*, *Technique*, *Creator, Date*, *Location*, *ID Number*, *Style/Period*, *Culture*, *Subject*, *Relation*, *Description*, *Source*, and *Rights*. Like the Dublin Core, the VRA Core scheme does not specify any particular syntax or rules for representing content.

Both CDWA and VRA emphasize the use of controlled vocabularies for specified elements. A number of existing vocabularies are suggested and communities are encouraged to develop additional vocabularies as needed.

MPEG Multimedia Metadata

The ISO/IEC Moving Picture Experts Group (MPEG) has developed a suite of standards for coded representation of digital audio and video. Two of the standards address metadata: MPEG-7, *Multimedia Content Description Interface (ISO/IEC 15938),* and MPEG-21, *Multimedia Framework (ISO/IEC 21000)*.

MPEG-7 defines the metadata elements, structure, and relationships that are used to describe audiovisual objects including still pictures, graphics, 3D models, music, audio, speech, video, or multimedia collections. It is a multi-part standard that addresses:

- *Description Tools* including *Descriptors* that define the syntax and the semantics of each metadata element and *Description Schemes* that specify the structure and semantics of the relationships between the elements.
- A *Description Definition Language* to define the syntax of the Description Tools, allow the creation of new *Description Schemes*, and allow the extension and modification of existing *Description Schemes*.
- *System tools*, to support storage and transmission, synchronization of descriptions with content, and management and protection of intellectual property.

Descriptors for visual and audio are defined separately using a hierarchy of elements and sub-elements. For visual objects there are descriptors for *Basic Structure*, *Color*, *Texture*, *Shape*, *Motion*, *Localization*, and *Face Recognition*. Audio descriptors are divided into two categories: low-level descriptors that are common to audio objects across most applications, and high-level descriptors that are specific to particular applications of audio. The cross-application low-level descriptors cover *Structures* and *Features* (temporal and spectral). The domain-specific high-level descriptors include such elements as *Musical Instrument Timbre*, *Melody Description*, and *Spoken Content Description*.

The *Description Schemes* are based on XML, and can be expressed in textual form suitable for editing, searching, filtering, and human readability; or in a binary form for storage, transmission, and streaming delivery. Since the full description of a multimedia object can be quite complex, the standard provides for a *Summary Description Scheme* geared to browsing and navigation.

The standard envisions that search engines could use MPEG-7 metadata descriptions to identify audiovisual objects in entirely new ways, such as digitizing a musical phrase played on a keyboard and then retrieving a list of musical pieces that contain the sequence of notes; drawing some lines on an electronic drawing tablet and retrieving images with similar graphics; or using a voice excerpt to retrieve related speech files, photographs, video clips, and biographical information of the speaker. These retrieval mechanisms are outside the scope of MPEG-7, but the standards developers wanted to accommodate these futuristic capabilities and have included many interoperability requirements beyond the typical metadata elements.

MPEG-21 was developed to address the need for an overarching framework to ensure interoperability of digital multimedia objects. The multi-part standard is not yet fully completed but is intended to include the following:

- *Part 1: Vision, Technologies and Strategy* provides the overview of the complete vision and plan for the framework. It was issued as an ISO technical report (ISO/IEC TR 21000:1-2001) and is available as a free download from ISO's publicly available standards website. A second edition of the vision document is underway to address comments and suggestions received from other organizations following the initial publication.
- *Part 2: Digital Item Declaration*, issued in 2003, describes a model for defining Digital Items. It includes a description of the syntax and semantics of each of the Digital Item Declaration elements and a corresponding XML schema.
- *Part 3: Digital Item Identification*, also issued in 2003, describes how to uniquely identify Digital Items and how to link Digital Items with related information such as descriptive metadata.
- *Part 4: Intellectual Property Management and Protection* is still in development. It is intended to define the framework for ensuring interoperability of intellectual property management tools, including authentication, and accommodates the *Rights* information defined in the following two parts.
- *Part 5: Rights Expression Language*, issued in 2004, is a machine-readable language that can declare rights and permissions.
- *Part 6: Rights Data Dictionary* is still in development. It will define a standard set of terms to be used with the Rights Expression Language. It is also expected to include specifications for mapping and transforming rights metadata terminology. The *Rights Data Dictionary* and *Expression Language* are being viewed as models for the handling of intellectual property metadata for applications beyond audiovisual.

- *Part 7: Digital Item Adaptation*, also in development, is intended to standardize networking and interoperability description tools. Included in this part will be User Characteristic description tools that specify user preferences.

There are some seven additional parts identified and in various stages of development that deal with technical interoperability issues of less specific relevance to metadata. All of the published parts are available from ISO as ISO/IEC 21000-[part#].

Metadata for Datasets

Metadata schemes for datasets are enabling original data in the science and social science fields to be shared in a way that was never possible before the Internet. One of the most well developed element sets is the Federal Geographic Data Committee (FGDC) *Content Standard for Digital Geospatial Metadata* (CSDGM), officially known as FGDC-STD-001-1998.

Geospatial datasets include topographic and demographic data, GIS (geographic information systems), and computer-aided cartography base files. They are used in a wide variety of areas, including soil and land use studies, biodiversity counts, climatology and global change tracking, remote sensing, and satellite imagery. The FGDC Content Standard is required for use with resources created and funded by the U.S. Government and is also being used by many state governments.

An international standard, ISO 19115, *Geographic Information—metadata* was issued in 2003. A technical amendment that will allow datasets to be both ISO and FGDC compliant is underway along with an implementation model that can be used in conjunction with an XML schema.

A metadata scheme becoming well established in the social and behavioral sciences is the Data Documentation Initiative (DDI) standard for describing social science datasets. The DDI is defined as an XML DTD, and allows for top down hierarchical description of a social science study, the data files resulting from that study, and the variables used in the data files. There is also a header area that uses Dublin Core elements for a high-level description of the DDI document itself.

Extensions and Profiles

Despite the recent development of many of these metadata schemes, most have already been subject to the changes brought about by implementing them in real world situations. These modifications are of two types: extensions and profiles.

An extension is the addition of elements to an already developed scheme to support the description of an information resource of a particular type or subject or to meet the needs of a particular interest group. Extensions increase the number of elements.

Profiles are subsets of a scheme that are implemented by a particular interest group. Profiles can constrain the number of elements that will be used, refine element definitions to describe the specific types of resources more accurately, and specify values that an element can take.

In practice, many applications use both extensions and profiles of base metadata schemes. For example, the National Biological Information Infrastructure (NBII) has developed a Biological Data Profile of the FGDC Content Standard for use with biological information resources. The profile defines an extended set of data for describing biological data, such as the taxonomic name of the organism and its classification in the taxonomic hierarchy.

Metadata in Action

A county land planner is studying the impact of new zoning laws on a particular bird species. The study team is composed of an ecologist, hydrologist, civil engineer, and environmental protection specialist.

Remote sensing data for the last 20 years provides a trend analysis of the decrease in wetlands, the bird's habitat. These datasets have FGDC metadata. The biologists on the study team need to document the results of a field inventory. Using a biological profile to extend the FGDC element set, the biologists add the genus-species name and taxonomic hierarchy. The ecologists are concerned with collection methods and modeling tools. The data related to the changes in human population are documented using a metadata set developed by the Census Bureau.

This study results in a technical report which is assigned Dublin Core metadata by the author. When the technical report is cataloged into the organization's repository, the Dublin Core elements are used as the basis for automatic generation of a MARC cataloging record. This record is enhanced by the cataloger and included in the library's online public access catalog.

The U.S. Department of Education's Gateway to Educational Materials (GEM) project has based their own metadata scheme on the Dublin Core. The GEM profile limits the Dublin Core elements that can be used (for example, *Contributor* is not allowed) and makes some elements mandatory. GEM also defines additional elements such as *Audience*, *Grade*, *Quality*, and *Standards*, extending the base Dublin Core set for educational use.

Creating Metadata

Who creates metadata? The answer to this varies by discipline, the resource being described, the tools available, and the expected outcome, but it is almost always a cooperative effort.

Much basic structural and administrative metadata is supplied by the technical staff who initially digitize or otherwise create the digital object, or is generated through an automated process. For descriptive metadata, it is best in some situations if the originator of the resource provides the information. This is particularly true in the documentation of scientific datasets where the originator has significant understanding of the rationale for the dataset and the uses to which it could be put, and for which there is little if any textual information from which an indexer could work.

However, many projects have found that it is more efficient to have indexers or other information professionals create the descriptive metadata, because the authors or creators of the data do not have the time or the skills. In other cases, a combination of researcher and information professional is used. The researcher may create a skeleton, completing the elements that can be supplied most readily. Then results may be supplemented or reviewed by the information specialist for consistency and compliance with the schema syntax and local guidelines.

Creation Tools

Many metadata project initiatives have developed tools and made them available to others, sometimes for free. A growing number of commercial software tools are also becoming available. Creation tools fall into several categories:

- *Templates* allow a user to enter the metadata values into pre-set fields that match the element set being used. The template will then generate a formatted set of the element attributes and their corresponding values.
- *Mark-up tools* will structure the metadata attributes and values into the specified schema language. Most of these tools generate XML or SGML Document Type Definitions (DTD). Some templates include such a mark-up as part of their final translation of the metadata.
- *Extraction tools* will automatically create metadata from an analysis of the digital resource. These tools are generally limited to textual resources. The quality of the metadata extracted can vary significantly based on the tool's algorithms as well as the content and structure of the source text. These tools should be considered as an aid to creating metadata. The resulting metadata should always be manually reviewed and edited.
- *Conversion tools* will translate one metadata format to another. The similarity of elements in the source and target formats will affect how much additional editing and manual input of metadata may be required.

Metadata tools are generally developed to support specific metadata schemas or element sets. The websites for the particular schema will frequently have links to relevant toolsets.

Metadata Quality Control

The creation of metadata automatically or by information originators who are not familiar with cataloging, indexing, or vocabulary control can create quality problems. Mandatory elements may be missing or used incorrectly. Schema syntax may have errors that prevent the metadata from being processed correctly. Metadata content terminology may be inconsistent, making it difficult to locate relevant information.

The *Framework of Guidance for Building Good Digital Collections,* available on the NISO website, articulates six principles applying to good metadata:

- Good metadata should be appropriate to the materials in the collection, users of the collection, and intended, current and likely use of the digital object.
- Good metadata supports interoperability.
- Good metadata uses standard controlled vocabularies to reflect the what, where, when and who of the content.
- Good metadata includes a clear statement on the conditions and terms of use for the digital object.
- Good metadata records are objects themselves and therefore should have the qualities of archivability, persistence, unique identification, etc. Good metadata should be authoritative and verifiable.
- Good metadata supports the long-term management of objects in collections.

There are a number of ongoing efforts for dealing with the metadata quality challenge:

- Metadata creation tools are being improved with such features as templates, pick lists that limit the selection in a particular field, and improved validation rules.
- Software interoperability programs that can automate the "crosswalk" between different schemas are continuously being developed and refined.
- Content originators are being formally trained in understanding metadata and controlled vocabulary concepts and in the

use of metadata-related software tools.

- Existing controlled vocabularies that may have initially been designed for a specific use or a narrow audience are getting broader use and awareness. For example, the *Content Types* and *Subtypes* originally defined for MIME email exchange are commonly used as the controlled list for the Dublin Core *Format* element.
- Communities of users are developing and refining audience-specific metadata schemas, application profiles, controlled vocabularies, and user guidelines. The *MODS User Guidelines* are a good example of the latter.

A Dublin Core description represented in RDF

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF SYSTEM "http://purl.org/
   dc/schemas/dcmes-xml-20000714.dtd">
<rdf:RDF xmlns:rdf="http://www.w3.org/
   1999/02/22-rdf-syntax-
   ns#"xmlns:dc="http://purl.org/dc/elements/
   1.1/">
   <rdf:Description about="http://
      www.niso.org/standards/resources/
      Metadata_Demystified.pdf">
         <dc:title>Metadata Demystified</
            dc:title>
         <dc:creator>Brand, Amy</dc:creator>
         <dc:creator>Daly, Frank</dc:creator>
         <dc:creator>Meyers, Barbara</
            dc:creator>
         <dc:subject>metadata</dc:subject>
         <dc:description>Presents an overview
            of metadata conventions in publish-
            ing.</dc:description>
         <dc:publisher>NISO Press</
            dc:publisher>
         <dc:publisher>The Sheridan Press</
            dc:publisher>
         <dc:date>2003-07</dc:date>
         <dc:format>application/pdf</
            dc.format>
   </rdf:Description>
</rdf:RDF>
```

Interoperability and Exchange of Metadata

Some people ask: Do we need so many metadata standards? With all the metadata standards, initiatives, extensions, and profiles, how can interoperability be ensured?

It is important to remember that different schemes serve distinct needs and audiences. Complementary schemes can be used to describe the same resource for multiple purposes and to serve a number of user groups. For example, a technical report could have a MARC metadata set in a library's online catalog, an FGDC description as part of the National Spatial Data Infrastructure Clearinghouse Mechanism, and an embedded set of Dublin Core elements.

The Resource Description Framework (RDF), developed by the World Wide Web Consortium (W3C), is a data model for the description of resources on the Web that provides a mechanism for integrating multiple metadata schemes. In RDF a namespace is defined by a URL pointing to a Web resource that describes the metadata scheme that is used in the description. Multiple namespaces can be defined, allowing elements from different schemes to be combined in a single resource description. Multiple descriptions, created at different times for different purposes, can also be linked to each other. RDF is generally expressed in XML.

Metadata Crosswalks

The interoperability and exchange of metadata is further facilitated by metadata crosswalks. A crosswalk is a mapping of the elements, semantics, and syntax from one metadata scheme to those of another.

A crosswalk allows metadata created by one community to be used by another group that employs a different metadata standard. The degree to which these crosswalks are successful at the individual record level depends on the similarity of the two schemes, the granularity of the elements in the target scheme compared to that of the source, and the compatibility of the content rules used to fill the elements of each scheme.

Crosswalks are important for virtual collections where resources are drawn from a variety of sources and are expected to act as a whole, perhaps with a single search engine applied. While these crosswalks are key, they are also labor intensive to develop and maintain. The mapping of schemes with fewer elements (less granularity) to those with more elements (more granularity) is problematic.

Table 1 on page 12 shows a crosswalk between Dublin Core, MARC 21, and VRA Core for selected elements. In this case, there is no attempt to map at the content level.

Metadata Registries

Registries are an important tool for managing metadata. Metadata registries can provide information on the definition, origin, source, and location of data. Registration can apply at many levels, including schemes, usage profiles, metadata elements, and code lists for element values. The metadata registry provides an integrating resource for

Table 1. Example of Metadata Crosswalk Mapping			
	Dublin Core	**EAD**	**MARC 21**
Title Element	Title	<titleproper>	245 00$a (Title Statement/Title proper)
Author Element	Creator	<author>	700 1#$a (Added Entry--Personal Name) (with $e=author) 720$a (Added Entry–Uncontrolled Name/Name) (with $e=author)
Date Created Element	Date.Created	<unitdate>	260 ##$c (Date of publication, distribution, etc.)

legacy data, acts as a lookup tool for designers of new databases, and documents each data element.

Registries can also document multiple schemes or element sets, particularly within a specific field of interest. A good example is the U.S. Environmental Protection Agency's Environmental Data Registry that provides information about thousands of data elements used in current and legacy EPA databases.

Standards relevant to metadata registries include ISO/IEC 11179, *Specification and Standardization of Data Elements,* and ANSI X3.285, *Metamodel for the Management of Shareable Data*.

Future Directions

Most early metadata standards have focused on the descriptive elements needed for discovery, identification, and retrieval. As metadata initiatives developed, administrative metadata, especially in the rights and preservation areas was further emphasized. Technical metadata is one area that still does not get much attention in metadata schemas. The effective exchange and use of the digital objects described by the metadata often requires knowledge of specific technical aspects of the objects beyond its filename and type. Newer standards are beginning to address this need. The NISO/AIIM standard, *Z39.87, Data Dictionary—Technical Metadata for Digital Still Images*, focuses solely on the technical data needed to facilitate interoperability between systems of digital image files. The metadata elements defined in the standard cover basic image parameters such as compression and color profile, information about the equipment and settings use to create the image, and performance assessment data such as sampling frequency and color maps.

Metadata work is ongoing across a number of standards development organizations. In the International Organization for Standardization (ISO), a subcommittee of Technical Committee (TC) 46 (*Information and documentation*), is addressing metadata development for bibliographic applications. ISO TC 211 (*Geographic information / Geomatics*) is developing metadata standards for applications in geographic information systems. The *Data management and interchange* subcommittee of ISO-IEC JTC1 (*Information technology*) is developing standards for the specification and management of metadata and has recently issued a technical report on *Procedures for achieving metadata registry content consistency* (ISO/IEC 20943).

Many organizations that developed metadata specifications outside the formal standards community are seeking to have their specifications turned into international standards. The Dublin Core is an example of this approach. It was originally developed in 1995 at a workshop sponsored by OCLC and the National Center for Supercomputing Applications. In 2001, it became an official ANSI/NISO standard (Z39.85) and in 2003 Dublin Core was issued as an international standard (ISO 15836).

The World Wide Web Consortium's (W3C) metadata activity has been incorporated into the Semantic Web, their initiative to "provide a common framework that allows data to be shared and reused across application, enterprise, and community boundaries." The RDF framework is one of the key enabling standards. The Semantic Web efforts are directed to standards that increase the interoperability of metadata, rather than specific metadata schemas.

The World Wide Web has created a revolution in the accessibility of information. The development and application of metadata represents a major improvement in the way information can be discovered and used. New technologies, standards, and best practices are continually advancing the applications for metadata. The resources in the following section will give you a head start in tracking developments and contain links to more information on the projects discussed throughout this document.

More Information on Metadata

General Resources

Digital Libraries: Metadata Resources (IFLA)
http://www.ifla.org/II/metadata.htm

A Framework of Guidance for Building Good Digital Collections
http://www.niso.org/framework/forumframework.html

Introduction to Metadata: Pathways to Digital Information
by Martha Baca
http://www.getty.edu/research/conducting_research/standards/intrometadata/index.html

Metadata: Cataloging by Any Other Name
by Jessica Milstead and Susan Feldman
ONLINE, January 1999
http://www.onlinemag.net/OL1999/milstead1.html

Metadata and Its Application
by Brad Eden
Library Technology Reports (September-October 2002)

Metadata Demystified: A Guide for Publishers
by Amy Brand, Frank Daly, Barbara Meyers
NISO Press & The Sheridan Press, 2003,
ISBN 1-880125-49-9
http://www.niso.org/standards/resources/Metadata_Demystified.pdf

Metadata Fundamentals for All Librarians
by Priscilla Caplan
ALA, 2003, ISBN: 0-8389-0847-0

Metadata Information Clearinghouse Interactive (MICI)
http://www.metadatainformation.org

Metadata Portals and Multi-standard Projects
by Candy Schwartz
http://web.simmons.edu/~schwartz/meta.html

Metadata Primer – A "How To" Guide on Metadata Implementation [for digital spatial data]
by David Hart and Hugh Phillips
http://www.lic.wisc.edu/metadata/metaprim.htm

Metadata Principles and Practicalities
Duval, Erik, Wayne Hodgins, Stuart Sutton, and Stuart L. Weibel
D-Lib Magazine 8(4) (April 2002)
http://www.dlib.org/dlib/april02/weibel/04weibel.html

Metadata Resources (UKOLN)
http://www.ukoln.ac.uk/metadata/resources

Metadata Standards
http://www.chin.gc.ca/English/Standards/metadata_intro.html

Metadata Standards, Crosswalks, and Standards Organizations
http://staff.library.mun.ca/staff/toolbox/standards.htm

Metadata.net – Projects, Tools & Services, and Schema Registry (Australia)
http://metadata.net/

Preservation Metadata for Digital Objects: A Review of the State of the Art
A White Paper by the OCLC/RLG Working Group on Preservation Metadata, January 31, 2001
www.oclc.org/research/projects/pmwg/presmeta_wp.pdf

Schemes, Initiatives, and Related Sites

Application profiles: mixing and matching metadata schemas
Rachel Heery and Manjula Patel, Ariadne, Issue 25, September 2000.
http://www.ariadne.ac.uk/issue25/app-profiles/intro.html

The Cedars Project (CURL exemplars in digital archives)
http://www.leeds.ac.uk/cedars/metadata.html

CDWA (Categories for the Description of Works of Art)
http://www.getty.edu/research/conducting_research/standards/cdwa/

DDI (Data Documentation Initiative)
http://www.icpsr.umich.edu/DDI/

DOI (Digital Object Identifier)
http://www.doi.org/

Dublin Core Metadata Initiative (DCMI)
http://dublincore.org

EAD (Encoded Archival Description)
http://www.loc.gov/ead/

Environmental Data Registry (EPA)
http://www.epa.gov/edr/

FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
http://www.fgdc.gov/metadata/

Gateway to Educational Materials (GEM)
http://www.geminfo.org/

IFLA Functional Requirements for Bibliographic Records
http://www.ifla.org/VII/s13/frbr/frbr.htm

IMS Global Learning Consortium
http://www.imsglobal.org

<indecs> interoperability of data in ecommerce systems
http://www.indecs.org/

LOM (Learning Object Metadata)
http://ltsc.ieee.org/wg12/

MARC 21 (Machine-Readable Cataloging)
http://www.loc.gov/marc

MetaWeb Project
http://www.dstc.edu.au/Research/Projects/metaweb/

METS (Metadata Encoding and Transmission Standard)
http://www.loc.gov/standards/mets/

MIX (Metadata for Images in XML Schema)
http://www.loc.gov/standards/mix/

MODS (Metadata Object Description Schema)
http://www.loc.gov/standards/mods/

MPEG (Moving Picture Experts Group)
http://www.chiariglione.org/mpeg/

NBII (National Biological Information Infrastructure)
http://www.nbii.gov/

Nordic Metadata Projects
http://www.lib.helsinki.fi/meta/

NSDI (National Spatial Data Infrastructure)
http://www.fgdc.gov/nsdi/

OAI (Open Archives Initiative)
http://www.openarchives.org/

OAIS (Open Archival Information System)
http://www.ccsds.org/documents/650x0b1.pdf

ONIX (Online Information Exchange)
http://www.editeur.org/onix.html

Open GIS Consortium
http://www.opengis.org/

PADI (Preserving Access to Digital Information)
http://www.nla.gov.au/padi/topics/32.html

PREMIS (PREservation Metadata: Implementation Strategies)
http://www.oclc.org/research/projects/pmwg

PURL (Persistent Uniform Resource Locator)
http://purl.org

RDF (Resource Description Framework)
http://www.w3.org/RDF/

SCHEMAS: Forum for Metadata Schema Implementors (UKOLN)
http://www.ukoln.ac.uk/metadata/schemas/

TEI (Text Encoding Initiative)
http://www.tei-c.org/

VRA (Visual Resources Association) Core Categories
http://www.vraweb.org/vracore3.htm

XML (Extensible Markup Language)
http://www.w3.org/XML/

Z39.50
http://www.loc.gov/z3950/agency/

ZING (Z39.50 Next Generation)
http://www.loc.gov/z3950/agency/zing/zing-home.html

Crosswalks and Lists of Crosswalks

All about Crosswalks
http://www.oclc.org/research/projects/mswitch/1_crosswalks.htm

Dublin Core / MARC / GILS Crosswalk
http://www.loc.gov/marc/dccross.html

FGDC to MARC
http://www.alexandria.ucsb.edu/public-documents/metadata/fgdc2marc.html

Issues in Crosswalking Content Metadata Standards
by Margaret St. Pierre and William P. LaPlant, Jr.
http://www.niso.org/press/whitepapers/crsswalk.html

MARC 21 to Dublin Core
http://www.loc.gov/marc/marc2dc.html

Metadata: Mapping between Metadata Formats (UKOLN)
http://www.ukoln.ac.uk/metadata/interoperability/

Metadata Mappings (Crosswalks)
http://libraries.mit.edu/guides/subjects/metadata/mappings.html

Metadata Standards Crosswalk (Getty)
http://www.getty.edu/research/conducting_research/standards/intrometadata/3_crosswalks/crosswalk1.html

Metadata Standards Crosswalks (Canadian Heritage Information Network)
http://www.chin.gc.ca/English/Standards/metadata_crosswalks.html

Metadata Registries & Clearinghouses

DCMI Registry Working Group
http://dublincore.org/groups/registry/

DESIRE Metadata Registry
http://desire.ukoln.ac.uk/registry/

Environmental Data Registry
http://www.epa.gov/edr/

FGDC Clearinghouse Registry
http://registry.gsdi.org/

MICI (Metadata Information Clearinghouse Interactive)
http://www.metadatainformation.org/

NBII Metadata Clearinghouse
http://metadata.nbii.gov/

The SCHEMAS Registry
http://www.schemas-forum.org/registry/

Tools for Metadata Creation

DDI Tools
http://www.icpsr.umich.edu/DDI/users/tools.html#a01

Dublin Core tools
http://dublincore.org/tools/

FGDC Metadata Tools
http://www.nbii.gov/datainfo/metadata/tools/

Metadata Software Tools
http://ukoln.bath.ac.uk/metadata/software-tools/

OAI-Specific Tools
http://www.openarchives.org/tools/tools.html

RDF Editors and Tools
http://www.ilrt.bris.ac.uk/discovery/rdf/resources/#sec-tools

TEI Software
http://www.tei-c.org/Software/index.html

Glossary

AACR2 (*Anglo-American Cataloging Rules)* – A standard set of rules for cataloging library materials. The "2" refers to the second edition.

administrative metadata – metadata related to the use, management, and encoding processes of digital objects over a period of time. Includes the subsets of technical metadata, rights management metadata, and preservation metadata.

ANSI (American National Standards Institute) – administers and coordinates the U.S. voluntary standardization and conformity assessment system.

CDWA (Categories for the Descriptions of Works of Art) – a metadata element set for describing artworks.

crosswalk – a mapping of the elements, semantics, and syntax from one metadata scheme to another.

CSDGM (Content Standard for Digital Geospatial Metadata) – a metadata standard developed by the FGDC. Officially known as FGDC-STD-001.

dataset – a collection of computer-readable data records.

DC (Dublin Core) – a general metadata element set for describing all types of resources.

DDI (Data Documentation Initiative) - a specification for describing social science datasets.

descriptive metadata – metadata that describes a work for purposes of discovery and identification, such as creator, title, and subject.

DLF (Digital Library Federation) – a membership organization dedicated to making digital information widely accessible.

DOI (Digital Object Identifier) – a unique identifier assigned to electronic objects of intellectual property which can be resolved to the object's location on the Internet.

DTD (Document Type Definition) – a formal description in SGML or XML syntax of the structure (elements, attributes, and entities) to be used for describing the specified document type.

EAD (Encoded Archival Description) – a metadata scheme for collection finding aids.

element set – information segments of the metadata record, often called semantics or content.

encoding rules – the syntax or prescribed order for the elements contained in the metadata description.

extension – an element that is not officially part of a metadata scheme, which is defined for use with that scheme for a particular application.

FGDC (Federal Geographic Data Committee) – a U.S. Federal government interagency committee responsible for developing the National Spatial Data Infrastructure.

GEM (Gateway to Educational Materials) – a U.S. Department of Education initiative that has defined an extension to the Dublin Core element set to accommodate educational resources.

GIS (Geographic Information System) – a computer system for capturing, managing, and displaying data related to positions on the Earth's surface.

HTML (Hypertext Mark-up Language) – a set of tags and rules derived from SGML used to create hypertext documents for the World Wide Web. Officially, a W3C Recommendation.

<indecs> (Interoperability of Data in ECommerce Systems) – a framework for metadata to support commerce in intellectual property.

interoperability – the ability of multiple systems, using different hardware and software platforms, data structures, and interfaces, to exchange and share data.

Glossary

ISO (International Organization for Standardization) – the primary international standards development organization.

IEC (International Electrotechnical Commission) – an international standards development organization for all electrical, electronic and related technologies. Co-sponsors with ISO the Joint Technical Committee 1 on Information Technology.

LOM (Learning Object Metadata) – a metadata scheme for technology-supported learning resources.

MARC 21 (MAchine Readable Cataloging) -- a formatting, record structure, and encoding standard for electronic bibliographic cataloging records developed by the Library of Congress. The “21” refers to the version of MARC issued in 1998 that integrated the U.S. and Canadian versions of MARC.

MARCXML – a metadata scheme for working with MARC data in a XML environment

metadata – structured information that describes, explains, locates, and otherwise makes it easier to retrieve and use an information resource.

metadata harvesting – a technique for extracting metadata from individual repositories and collecting it in a central catalog

METS (Metadata Encoding and Transmission Standard) – a metadata scheme for complex digital library objects.

MODS (Metadata Object Description Schema) – a metadata scheme for rich description of electronic resources.

MPEG (Moving Pictures Experts Group) – Standards Committee 29, Working Group 11 of ISO/IEC JTC1, which develops standards for digital audio and video. Also refers to a suite of standards developed by the group.

namespace – in RDF, a way to tie a specific use of a metadata element to the scheme where the intended definition is to be found.

NISO (National Information Standards Organization) – a standards development organization, accredited by the American National Standards Institute, that develops library and information-related standards.

ONIX (Online Information Exchange) – a metadata scheme for book bibliographic, trade, and promotional data.

preservation metadata – a form of administrative metadata dealing with the provenance of a resource and its archival management.

profile – a subset of a scheme defined and used by a particular interest group to customize the scheme for its purposes.

PURL (**Persistent URL)** – a naming and resolution system developed by OCLC utilizing an intermediate redirection service to locate a resource’s URL.

qualifier – an optional sub-element to a Dublin Core element that is used to further refine the element or support a specific encoding scheme.

RDF (**Resource Description Framework)** – a language for representing metadata about Web resources so it can be exchanged between applications without loss of meaning. Officially, a suite of W3C specifications.

registry – a formal system for the documentation of the element sets, descriptions, semantics, and syntax of one or more metadata schemes.

rights management metadata – a form of administrative metadata dealing with the intellectual property rights of a resource.

scheme (schema)– a metadata element set and rules for using it.

semantics – the names and meanings of metadata elements.

SGML (Standard Generalized Markup Language) – a language used to mark-up electronic documents with tags that define the relationship between the content and the structure. Officially, international standard ISO 8879, *Information processing—Text and office systems—Standard Generalized Markup Language (SGML)*.

structural metadata – metadata that indicates how compound objects are structured, provided to support use of the objects.

syntax – rules for how metadata elements and their content are encoded.

technical metadata – a form of administrative metadata dealing with the creation or storage encoding processes or formats of the resource.

TEI (Text Encoding Initiative) – a metadata scheme for electronic text

URL (Uniform Resource Locator) – A unique address for identifying and locating a resource on the Internet.

VRA (Visual Resources Association) Core – a metadata scheme for describing a visual work and its representations

W3C (World Wide Web Consortium) – an international consortium that develops consensus protocols and specifications to ensure the interoperability of the World Wide Web.

XML (Extensible Mark-up Language) – an application profile of SGML designed for use in Web applications. Officially, a W3C Recommendation.

Z39.50 – a NISO and ISO standard protocol for cross-system search and retrieval. Officially, international standard, ISO 23950, *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*, and ANSI/NISO standard Z39.50.

Support the leaders in our community who support NISO as Voting Members:

3M

American Association of Law Libraries

American Chemical Society

American Library Association

American Society for Information Science and Technology

American Society of Indexers

American Theological Library Association

ARMA International

Armed Forces Medical Library

Art Libraries Society of North America

AIIM International

Association of Information and Dissemination Centers

Association of Jewish Libraries

Association of Research Libraries

Auto-Graphics, Inc.

Barnes & Noble, Inc.

Book Industry Communication

California Digital Library

Cambridge Information Group

Checkpoint Systems, Inc.

College Center for Library Automation

Colorado State Library

CrossRef

Davandy, L.L.C.

Docutek Information Systems

Dynix Corporation

EBSCO Information Services

Elsevier Science Inc.

Endeavor Information Systems, Inc.

Entopia, Inc.

ExLibris USA

Fretwell-Downing Informatics

Gale Group

Geac Library Solutions

GIS Information Systems, Inc.

H.W. Wilson Company

Helsinki University Library

Index Data

Infotrieve

Innovative Interfaces, Inc.

Institute for Scientific Information

The International DOI Foundation

Ithaka/JSTOR/ARTstor

John Wiley & Sons, Inc.

KINS, Inc.

Library Binding Institute

Library of Congress

The Library Corporation

Los Alamos National Laboratory

Lucent Technologies

Medical Library Association

MINITEX

Modern Language Association

Motion Picture Association of America

MuseGlobal, Inc.

Music Library Association

National Agricultural Library

National Archives and Records Administration

National Federation of Abstracting and Information Services

National Library of Medicine

National Security Agency

Nylink

OCLC, Inc.

Openly Informatics, Inc.

ProQuest Information and Learning

Random House, Inc.

Recording Industry Association of America

The Research Libraries Group

SAGE Publications

Serials Solutions, Inc.

SIRSI Corporation

Society for Technical Communication

Society of American Archivists

Special Libraries Association

Synapse Corporation

TAGSYS, Inc.

Talis Information Ltd.

Triangle Research Libraries Network

U.S. Department of Commerce, NIST, Office of Information Services

U.S. Department of Defense, DTIC (Defense Technical Information Center)

U.S. Department of Energy, Office of Scientific & Technical Information

U.S. Government Printing Office

U.S. National Commission on Libraries and Information Science

VTLS, Inc.

WebFeat

NISO Press

ISBN 1-880124-62-9