<glossary of metadata standards>

AACR2
Anglo-American Cataloging Rules, 2nd edition, Revised
http://www.aacr2.org/

AACR2 is the primary content standard used in the library field in the US, Canada, the UK, and Australia. Its use is almost exclusive to libraries, although there have been calls for the archives and museum communities to adopt it for the description of "bibliographic" types of materials. While primarily focused on descriptive metadata, instructions exist that cover technical, rights, and structural metadata as well. AACR2 is scheduled to be replaced by RDA.

AAT
Art & Architecture Thesaurus
http://www.getty.edu/research/conducting_research/vocabularies/aat/

The AAT is one of a suite of controlled vocabularies maintained by the Vocabulary Program at the Getty Research Institute in Los Angeles. It focuses on generic terms for the description of works of art, architecture, and material culture. The AAT is organized hierarchically within seven facets: associated concepts, physical properties, styles and periods, agents, activities, materials, and objects. The vocabulary may be searched one term at a time freely on the web, and is available for license in bulk.

ADL
AES31-3-2008: AES standard for network and file transfer of audio - Audio-file transfer and exchange - Part 3: Simple project interchange (Audio Decision List)
http://www.aes.org/publications/standards/

The AES Audio Decision List (ADL) is a text-based file format and metadata standard for encoding the results of audio editing actions. The format records cuts, fades, the results of processing and filtering actions, and other edits to audio files made by a sound engineer. AES31-3 ADL support is included to some extent in audio editing software such as WaveLab and Pyramix.

AES Core Audio
AES-X098B: Descriptive metadata for audio objects - Core audio schema
http://www.aes.org/standards/meetings/project-status.cfm

The AES Core Audio schema (in draft as X098B) is part of the Audio Engineering Society's suite of standards for descriptive metadata for audio objects, although the AES uses the term "descriptive" differently than the library community does. The scope of the AES Core Audio standard is wide, including analog originals, digitally reformatted copies, and native digital recordings. The specification allows the capture of basic audio properties such as sample rate for digital files and groove width for physical discs. It also breaks audio objects down into "faces" (physical sides or directions for playback contiguously), "regions" (specific formats such as playing speed within a face), and "streams" (specific audio channels within a region). The AES Core Audio Schema and documentation are currently in draft status with no firm release date yet scheduled.

AES Process History
AES-X098C: Administrative metadata for audio objects - Process history schema
http://www.aes.org/standards/meetings/project-status.cfm

The AES Process History standard is a data dictionary and XML Schema for recording information about processes that have been performed on an audio object over time. This includes but is not limited to transfer of audio between physical formats or from a physical format to a digital one. The standard provides elements to track extensive detail about device settings, signal chains, and even equipment serial numbers. AES-X098C is currently in draft status.

AGLS
Australian Government Locator Service
http://www.naa.gov.au/records-management/create-capture-describe/describe/AGLS/index.aspx

AGLS is an Australian government metadata standard intended for the description of government resources on the Web. It uses DCMI Terms properties, to which it adds a few additional properties such as function and mandate. AGLS can be expressed either in HTML or RDF/XML. AGLS usage guidelines frequently suggest appropriate controlled vocabularies for specific properties.

Atom
Atom Syndication Format
http://tools.ietf.org/html/rfc4287

Atom is a syndication format for Web content in XML, allowing frequently updated information such as news feeds to be pushed to subscribed users. The most frequent use of Atom is to embedd an Atom-encoded news feed into an otherwise human-readable web page such as a news service or a blog.

The main alternative to Atom for syndicated content is RSS. Atom can also refer to a full Web publishing protocol in addition to the syndication format.

BISAC
Book Industry Standards and Communications
http://www.bisg.org/publications/product.php?p=14

BISAC is a subject vocabulary for books created by the publishing industry, specifically the Book Industry Study Group (BISG). It is arranged hierarchically and includes codes as well as textual labels for entries. BISAC is commonly used in bookstores, and has been seen in action in Google Book Search.

CanCore
http://cancore.athabasecau.ca/en/index.html

CanCore is a set of guidelines for the implementation of the IEEE LOM metadata standard. It arose from Canadian efforts on metadata for educational materials, and as such, its focus is on learning resources.

CCO
Cataloging Cultural Objects
http://www.vrafoundation.org/ccowb/

CCO is a content standard for the description of works of art, architecture, and material culture. It was developed in partnership between the Visual Resources Association and the Getty Foundation, and as such attempts to meet the needs of both the visual resources (frequently tied to libraries) and museum communities.

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

CDWA is a long-standing metadata standard from the museum community designed as a framework for the description of works of art and material culture. It is an extensive set of descriptive elements, including 532 categories and subcategories. Usage guidelines distinguish between data elements intended for display and those intended for indexing. CDWA defines only category labels and definitions - it does not define a specific syntax for encoding them, although the CDWA guidelines suggest a relational structure providing for easy re-use of authority records. CDWA is commonly implemented in museum management software.

CDWA Lite
Categories for the Description of Works of Art Lite
http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

CDWA Lite is an XML representation of a subset of the full CDWA category set, explicitly designed for the sharing of descriptions of works of art and material culture via OAI-PMH. The OAICatMuseum OAI-PMH data provider software is designed to share CDWA Lite records in addition to Simple Dublin Core. There are ongoing efforts to harmonize CDWA Lite and MuseumDat into a new format called LIDO.

CIDOC/CRM
CIDOC Conceptual Reference Model
http://cidoc.ics.forth.gr/

CIDOC/CRM defines concepts and relationships essential for the description of cultural heritage materials. Beyond the traditional descriptive information about physical objects, CIDOC/CRM also focuses on space and time information, including modeling of events that affect the physical objects held by cultural heritage institutions. CIDOC/CRM is strongly allied with the museum community. In addition to a textual document intended for human implementers, CIDOC/CRM is defined in a formal OWL ontology and in RDF. The CIDOC/CRM has been standardized as ISO 21127:2006.

CQL
Contextual Query Language
http://www.loc.gov/standards/sru/specs/cql.html

CQL is a query language for information systems maintained at the Library of Congress. It operates using the concept of "context sets," allowing implementers to create new indexes, operators, etc., but still maintain common query parsing rules. CQL can be implemented at various conformance levels, and implementations are required to return diagnostics when a query feature is not supported. CQL is the query language most commonly used with the SRU search protocol. It attempts to be at once both simple and robust. The current version is 1.2, which represents a name change from Common Query Language in CQL 1.1.

DACS
Describing Archives: A Content Standard
http://www.archivists.org/governance/standards/dacs.asp

DACS is a product and publication of the Society of American Archivists, and thus reflects the descriptive priorities of the archival community. It replaces the older Archives, Personal Papers, and Manuscripts (APPM) content standard. It primarily focuses on the description of personal papers and institutional records. DACS is generally used in a multi-level description environment although it is possible to apply it for item-level description as well.

DC
Dublin Core Metadata Element Set
http://www.dublincore.org/documents/dces/

Dublin Core is a widely misunderstood metadata standard. The Dublin Core Metadata Element Set (DCMES) is also known as Simple Dublin Core. Simple Dublin Core is a basic 15-element set designed to represent core features across all resource formats. It is standardized as ISO 15836-2003, ANSI/NISO Z39.85-2007, and IETF RFC 5013. The Dublin Core Usage Guidelines sometimes suggest (but do not require) specific content guidelines or controlled vocabularies. Simple Dublin Core is widely known as the baseline metadata format required for all resources shared via OAI-PMH. Encoding of the DCMES in HTML <meta> tags was popular in the early days of search engines, but today most search engines prefer to weigh page text and linking patterns more heavily then page creator-supplied structured metadata.

DCAM
Dublin Core Metadata Initiative Abstract Model
http://dublincore.org/documents/abstract-model/

The DCMI Abstract Model is a framework for the components of resource description and how they relate to one another. The structure of the DCAM is very similar to and inspired by the RDF model. The full model has three main sub-parts: the DCMI Resource Model, the DCMI Description Set Model, and the DCMI Vocabulary Model. These three work together to allow robust semantic relationships to be recorded between resources. The DCAM is a far cry from the 15 element set of simple Dublin Core that is familiar to many in the cultural heritage community, and represents a different and more robust approach to resource description. The DCAM is significantly more complex than the original simple Dublin Core, but offers a corresponding significant improvement in functionality and reusability. Encodings of Dublin Core metadata in HTML, XML, and RDF all implement different subsets of the full DCAM.

DDC
Dewey Decimal Classification
http://www.oclc.org/dewey/

The Dewey Decimal Classification is primarily used in public libraries, and is currently in its 22nd edition. Dewey divides knowledge into 10 primary classes, with further subdivisions possible in multiples of 10. A process of "number building" is used to read the Dewey schedules and construct a potentially long number combining different intellectual aspects of a resource.

DIF
Discovery Interchange Format
http://gcmd.nasa.gov/User/difguide/difman.html

DIF is an early metadata initiative from the Earth sciences community, intended for the description of scientific data sets. It includes elements focusing on instruments that capture data, temporal and spatial characteristics of the data, and projects with which the dataset is associated. It is defined as a W3C XML Schema. DIF is fully compatible with the ISO 19115/TC211 geospatial metadata standard by providing places for elements from that standard.

DIG35
Digital Imaging Group 35
http://www.i3a.org/technologies/metadata/

DIG35 is a metadata format for still images that grew out of industry work, specifically from the International Imaging Industry Association (I3A). DIG35 is divided into five blocks: basic image parameter metadata, image creation metadata, content description metadata, history metadata, and intellectual property rights metadata. DIG35 is defined primarily as a human-readable data dictionary, but a W3C XML Schema is also available.

DTD
Document Type Definition
http://xmlfiles.com/dtd/

DTDs are mechanisms for defining XML languages, and serve as an alternative to W3C XML Schema and RelaxNG for this purpose. The DTD language dates back to SGML, but currently is also used for XML applications. DTD syntax is significantly simpler than W3C XML Schema, but lacks some more advanced functionality, such as strong data typing of element or attribute content.

DwC
Darwin Core
http://www.tdwg.org/activities/darwincore/

Darwin Core is a "concept list" defining categories of information useful for the description of biological data - specifically, where organisms and species exist in time and space. The specification exists as a textual representation of the defined concepts and as an XML Schema. Darwin Core also provides extensions for curatorial, geospatial, paleontological, and interaction information. Darwin Core is managed under the auspices of Biodiversity Information Standards, a nonprofit devoted to promoting the sharing of biodiversity data.

EAC-CPF
Encoded Archival Context - Corporate Bodies, Persons, and Families
http://eac.staatsbibliothek-berlin.de/

EAC-CPF is an XML representation of data about corporate bodies, persons, and families conformant to the model presented in the ISAAR (CPF) specification. In contrast to traditional library authority records for these entities that exist primarily to establish and justify controlled headings, EAC-CPF reflects its roots in archival description by focusing more on the context in which these entities operate. While EAC-CPF has a long development history, a significantly revised version was released in 2010, and as of yet the companion EAD standard has not had the opportunity to evolve to allow the two to be used in concert more effectively. EAC-CPF is maintained by the Society of American Archivists in partnership with the Berlin State Library.

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

EAD is a markup language for archival finding aids. It provides XML elements for structural and presentational data typically found in finding aid documents. While EAD is a markup language in the sense that it "flags" data structures as they appear in a pre-existing text, it is also the primary source of (semi-) structured descriptive metadata in archives.

EML
Ecological Markup Language
http://knb.ecoinformatics.org/software/eml/eml-2.0.1/

EML grew out of early metadata efforts from the Ecological Society of America. It is an extremely detailed specification that is intended to support the description of any type of ecological information, including raw data, published research papers, rights information, and research protocols. EML is defined as a series of W3C XML Schemas, and can wrap data packages together with metadata. At the highest level, EML models four primary entities: datasets, literature, software, and protocols, although not all are always applicable or are required for use.

FGDC/CSDGM
Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata
http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/index_html

The standard commonly referred to as FGDC (although FGDC is the maintenance agency, and CSDGM is the actual element set) is a large and early metadata standard for geospatial information created by agencies of the US federal government. The FGDC web site describes the scope of this standard as to allow users to "determine the availability of a set of geospatial data, to determine the fitness [of] the set of geospatial data for an intended use, to determine the means of accessing the set of geospatial data, and to successfully transfer the set of geospatial data."

The current production version of FGDC is 2.0, from 1998. Since this time, an international standard for geospatial information (ISO 19115) has emerged. Plans have been announced to create a US national geospatial metadata standard as a profile of ISO 19115, and to create version 3.0 of CSDGM as an implementation of that. This work has not yet been finalized.

FOAF
Friend of a Friend
http://www.foaf-project.org/

FOAF is an RDF syntax for describing people, intended to be used on the Semantic Web. It includes features for encoding names, email addresses, personal interests, home pages, and various online identities. Although the language is focused on people, encoding traditional library authority files in FOAF is challenging due to its assumption that each individual has only one FOAF identity and its focus on online presence for current living persons.

FRAD
Functional Requirements for Authority Data
http://www.ifla.org/publications/ifla-series-on-bibliographic-control-34

FRAD is a companion document to the earlier FRBR conceptual model developed by IFLA. FRAD expands on FRBR by adding additional attributes to each of the Group 1, 2, and 3 entities; adding a new Group 2 entity (Family); and adding new entities intended to support the authority control process (Name, Identifier, Controlled Access Point, Rules, and Agency). Perhaps the strongest promise of the FRAD model is support for multi-lingual catalogs that can display to a user different forms of names for various entities depending on a user's location or language preferences.

In addition to expanded entities and attributes, FRAD defines a different set of user tasks for authority data than FRBR did for bibliographic data. Here, the user tasks are Find, Identify, Contextualize, and Justify. The final FRAD report was published by IFLA in 2009.

FRBR
Functional Requirements for Bibliographic Records
http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records

FRBR is a 1998 conceptual model of the bibliographic universe, created in order to better understand the user tasks catalogs can and should support, and to suggest how bibliographic data might be viewed in support of these tasks. The most commonly known features of the FRBR report are its four user tasks (Find, Identify, Select, and Obtain) and the Group 1 Entities which categorize the products of intellectual and artistic endeavors (Work, Expression, Manifestation, and Item). The FRBR report has other features as well, including Group 2 Entities representing the creators of Group 1 Entities (Person and Corporate Body), Group 3 Entities which are the subjects of Works (Group 1 Entities, Group 2 Entities, plus Concept, Object, Event, and Place), and minimal standards for national bibliographic records.

The FRBR conceptual model has received a great deal of discussion in the cultural heritage community, but only in the very late 2000s have concrete implementations of the conceptual model into working systems begun to appear.

FRSAD
Functional Requirements for Subject Authority Data
http://www.ifla.org/node/1297

The FRSAD initiative is intended to provide a more complete conceptual model for FRBR Group 3 entities in their role serving as the subjects of FRBR Works. A draft of FRSAD for public comment was issued in early 2009. This draft abandoned the FRBR Group 3 entity structure (Concept, Object, Event, Place) in favor of conceptual entities (Thema) that are known by name tokens (Nomen).

GEM
Gateway to Educational Materials
http://www.thegateway.org/about/documentation2/schemas/index_html/

GEM is an RDF metadata vocabulary designed for the description of educational resources. The GEM model includes all the properties available in DCMI Terms, to which are added education-specific properties such as educational standards and pedagogical methods. The current version of GEM is 2.0.

GEM has also created a number of controlled vocabularies, including lists for audience level, assessment methods and instruments, and resource type. GEM Consortium members have access to the GemCat metadata creation tool, which produces GEM-compliant metadata.

GILS
Government Information Locator Service
http://www.gils.net

GILS was an early metadata standard for the encoding of descriptive information for government records. It contained fields for the recording of creators, titles, identifiers, topical and geographic subjects, time periods, and access information. Usage of GILS has dropped off significantly in recent years.

GML
OpenGIS Geography Markup Language
http://www.opengeospatial.org/standards/gml

GML is an element set intended for the description of geographic information, as well as providing for the creation of application schemas for more specific uses of GML. The GML schema is extremely detailed in its ability to describe sppatial and temporal features, topologies, and observation methods. GML is written in W3C XML Schema language, and is standardized as ISO 19136:2007.

ID3
http://www.id3.org/

ID3 tags are data stored inside an MP3 audio file to assist with the identification of the content on the file. ID3v2 includes a number of predefined "frames" (essentially, fields) for use, including Album title, Composer, Date of recording, Original artist(s)/performer(s), and File owner/licensee. Images and other content files can also be embedded inside the ID3 chunk. ID3v2 also allows for user-defined frames. Most audio players, such as iTunes and Windows Media Player, can display ID3 tags to users and allow them to be edited.

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

The LOM standard is a "conceptual data schema" for the description of learning objects (by a broad definition of the term). LOM was developed and formalized through the IEEE and their Learning Technology Standards Committee. The stated purpose of LOM is to "facilitate search, evaluation, acquisition, and use of learning objects, for instance by learners or instructors or automated software processes." LOM data elements are grouped into nine categories: general, lifecycle, meta-metadata, technical, educational, rights, relation, annotation, and classification.

In addition to the conceptual data schema outlined in LOM documentation, a binding of the LOM model to XML has been creteated, and expressed in XML Schema language. Following the development of the DCMI Abstract Model, efforts have commenced to harmonize IEEE/LOM with this model.

indecs
<indecs> Metadata Framework
http://www.doi.org/topics/indecs/indecs_framework_2000.pdf

<indecs> describes itself as a "model of commerce," operating under a simple basic premise: "People make stuff. People use stuff. People do deals about stuff." The basic entities of <indecs> are as follows: Entities (something interesting) break down into Percepts (things percieved) which are further broken down into Beings and Things, Relations which are further broken down into Events and Situations, and Concepts.

<indecs> shares many common features with the FRBR model, but is different in that it focuses heavily on events that act on entities over time, an area FRBR avoids completely. While <indecs> defines a robust conceptual model, it is unclear if many systems in either the cultural heritage or business communities have built systems that implement all or part of the model.

ISAAR(CPF)
International Standard Archival Authority Record for Corporate Bodies, Persons and Families
http://www.icacds.org.uk/eng/isaar2ndedn-c_3_1.pdf

ISAAR(CPF) is a descriptive metadata model for contextual information in archives, covering the descriptions of corporate bodies, persons, and families; construction of access points for these entities; and documenting relationships among them, and between them and resources. The standard is intended to promote the sharing of archival authority records between institutions. Like the IDBD, ISAAR(CPF) is divided into several areas of description: identity, description, relationships, and control.

The first edition of ISAAR(CPF) was published in 1996, and the second edition was published in 2004. ISAAR(CPF) is intended to be used with ISAD(G) for resource description. The EAC structure standard for archival authority data is intended to support the encoding of ISAAR(CPF)-compliant records.

ISAD(G)
International Standard Archival Description (General)
http://www.ica.org/en/node/30000

ISAD(G) is a statement of general principles for archival description, throughout the archival management process, and applicable to any type of material controlled archivally regardless of format or media type. ISAD(G) defines 26 elements of archival description, and defers to national or local rules for the structure of the values of those elements. The definitions of the archival description elements presented in ISAD(G) conform to the archival principle of respect des fonds and are structured to allow multi-level description. Like ISBD, ISAD(G) is organized into "areas" of description. These are: Identity Statement, Context, Content and Structure, Condition of Access and Use, Allied Materials, Note, and Description Control Areas.

ISBD
International Standard Bibliographic Description
http://www.ifla.org/publications/international-standard-bibliographic-description

ISBD is a standard from IFLA designed to make bibliographic description more consistent across a wide range of applications. It serves two distinct functions: to define the selection and order of data elements to be recorded, and to prescribe punctuation to be used inside a bibliographic description. ISBD is divided into 8 "areas" of description: title and statement of responsibility; edition; material or type of resource specific; publication, production, distribution, etc; physical description; series; note; and resource identifier and terms of availability. The structure of AACR2 is strongly informed by the ISBD areas of information.

ISO 19115
Geographic Information - Metadata
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020

ISO 19115 is an international geospatial metadata standard which was built on the framework of the earlier US FGDC/CSDGM. Its initial version was released in 2003, and a revision was completed in 2009. Plans have been announced to create a US national geospatial metadata standard as a profile of ISO 19115, and to create version 3.0 of CSDGM as an implementation of that. This work has not yet been finalized.

KML
Keyhole Markup Language
http://code.google.com/apis/kml/documentation/

KML is a markup language for geographic data used in the Google Maps and Google Earth services. It can be used to describe placemarks (single points), ground overlays, paths, and polygons. The language allows for 3-D spatial data, including altitude in addition to latitude and longitude. KML's relative simplicity and the availability of the Google Maps API have contributed to quick and fairly widesparead adoption of this language.

LCC
Library of Congress Classification
http://www.loc.gov/catdir/cpso/lcc.html

The Library of Congress Classification is used primarily in academic libraries. It is divided into 21 basic classes, each of which start with one or more uppercase letters. Full class numbers use a mixture of letters and numbers, with subtopics offset by a period. Libraries typically append Cutter numbers at the end of LC class numbers to create a full call number for physical shelving of materials.

LCSH
Library of Congress Subject Headings
http://authorities.loc.gov

LCSH is a long-standing controlled vocabulary maintained by the Library of Congress, covering topical subjects, genres, and geographic places among other related areas of study. It is a precoordinated vocabulary, built upon the principle of literary warrant. Libraries can contribute new terms for consideration via the SACO initiative. Despite its function as a controlled vocabulary, LCSH is not a fully enumerated list, allowing the presence of "standard subdivisions" on explicitly authorized terms according to human-readable rules.

With the development of a new service, http://id.loc.gov, that makes Library of Congress-hosted vocabularies available to machine applications, LCSH and other vocabularies are now more readily available to applications outside the library community and especially outside the cultural heritage community.

Linked Data
Linked Data
http://www.w3.org/DesignIssues/LinkedData.html

Linked data is a broad term that refers to a framework and a set of best practices for exposing data on the Semantic Web and making connections between resources. Linked data implementations are guided by four principles outlined by Tim Berners-Lee in 2006: 1) Use URIs as names for things, 2) Use HTTP URIs so that people can look up those names, 3) When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL), and 4) Include links to other URIs. so that they can discover more things. One of the highest profile uses of linked data in the cultural heritage community is http://id.loc.gov, although that service does not systematically implement the fourth principle of linked data - linking to other things. The records at http://id.loc.gov point to other records in the same service, but not to data elsewhere on the web. Additional information can be found at linkeddata.org.

MADS
Metadata Authority Description Schema
http://www.loc.gov/standards/mads/

MADS is a companion to MODS, intended to encode authority data that is referenced by MODS bibliographic records. The structure and design of MADS are heavily influenced by the MARC Authority format. As such, it provides for the encoding of headings and cross references traditionally established by the library community, including personal names, corporate names, name/title entries, title entries, subject, genres, and geographic places. While MADS allows for more of a complete description of an entity than MARC Authority does, it still retains a focus on documenting and justifying choice of headings. MADS elements use the same name as MODS elements whenever feasible. MADS is maintained by the Library of Congress, and its content is managed by the MODS/MADS Editorial Committee.

MARC
Machine Readable Cataloging
http://www.loc.gov/marc/

MARC was first developed in the late 1960s at the Library of Congress, and represented the first major attempt to encode bibliographic data in machine-readable form. MARC uses a mixture of fixed and variable fields to record information. The variable fields are themselves a mixture of coded and textual data. The MARC format is defined in ISO2709, which prescribes numeric field names that contain alphanumeric subfields. The MARC format in use in the US is known as MARC21. UNIMARC is a variant common in Europe. While there are five formats in the MARC21 suite, the Bibliographic and Authority formats are the most commonly used.

MARC Relator Codes
MARC Relator Codes
http://www.loc.gov/marc/relators/relaterm.html

The MARC Relator Codes list is provided by the Library of Congress for use in specifying the role of an individual or group in connection with a resource. The list is expressed both in three-letter codes and in full English-language terms. Codes and terms from this list are commonly used in MARC and in MODS.

In cooperation with the DCMI community, the Library of Congress has developed a version of the MARC relator codes suitable for use in Dublin Core Application Profiles. These may be found at http://id.loc.gov/vocabulary/relators.

MARCXML
MARC in XML
http://www.loc.gov/standards/marcxml/

MARCXML, first released in 2002, is a representation of the ISO2709 MARC format in an XML syntax. MARCXML is designed to be fully interchangeable with MARC21 - records can be moved back and for the between the two formats without any loss of data. The MARCXML Schema, however, allows any 3-number field name and any alphanumeric subfield name, not restricting values to those defined in MARC21. MARCXML is primarily used as an intermediate step between MARC21 and other XML formats, as MARCXML can be converted to other XML formats with XSLT, which is not possible directly from MARC21.

MathML
Mathematical Markup Language
http://www.w3.org/Math/

MathML is a W3C Recommendation for the low-level encoding of mathematical information, with the intention of representing this information on the Web. It is defined by an XML DTD. MathML elements exist both in support of presentation of mathematical data and for the content of the mathematical data itself.

MEI
Music Encoding Initiative
http://www2.lib.virginia.edu/innovation/mei/

MEI is a markup language for Western common music notation. It is strongly inspirted by the structure and design of TEI, and was developed in response to an identified need for a music notation format that facilitates research into the structure of musical corpora. In addition to the full notation encoding, MEI includes a header for bibliographic information about the notation file. MEI is developed and maintained as an XML DTD by Perry Roland at the University of Virginia.

MESH
Medical Subject Headings
http://www.nlm.nih.gov/mesh/

MeSH is produced by the US National Library of Medicine for the description of biomedical journal literature, books, and other formats collected by the Library. It is also used for subject indexing in the PubMED database. The MeSH vocabulary contains a full syndetic structure of broader, narrower, and "use for" terms. The full vocabulary is available online for individual searches and downloads in XML and ASCII formats.

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

METS is an XML metadata standard intended to package all the information needed to represent a complex object, including both primary files and metadata that describes them. It defines its own structure for representing files and the relationships between them, and allows embedding or referencing descriptive, technical, rights, source, and digital provenance metadata defined by other schemas. METS has various levels of support in digital asset management systems, including DigiTool, Greenstone, and the Archivists' Toolkit. This standard grew out of early work on representing complex digital objects by the Making of America II project. METS is maintained at the Library of Congress and through a volunteer Editorial Board.

METS Rights
METS Rights Declaration Schema
http://www.loc.gov/standards/mets/news080503.html

METS Rights was developed by the METS Editorial Board as a simple and easy to implement rights schema, as an alternative or temporary solution before implementing a more comprehensive rights metadata format. It focuses on a simple structure for access and ownership rights for locally-controlled digital resources.

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

MIX is an XML representation of the Data Dictionary - Technical Metadata for Digital Still Images (ANSI/NISO Z39.87-2006). As a technical metadata format for still images, MIX can be used to describe images born digitally such as those taken with digital cameras, and images that have been reformatted from analog originals such as scans of photographs or pages of text. The data dictionary on which MIX is based includes four basic areas of metadata: basic digital object information, basic image information, image caputure metadata, and image assessment metadata. The MIX XML Schema is maintained by the Library of Congress.

MO
Music Ontology
http://musicontology.com/

The Music Ontology is a framework for the description of musical materials intended to push these descriptions to the Semantic Web. It is divided into three levels allowing incremental increases in complexity. Level 1 is for basic descriptive information such as tracks, artists, and releases. Level 2 adds the music creation workflow such as arrangement, performance, and recording. Level 3 adds support for complex events such as timelines and relationships between performances. The Music Ontology uses FRBR principles to separate a musical Work from its Manifestations. It is expressed in RDF/OWL.

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

MODS was developed by the Library of Congress Network Development and MARC Standards Office as a MARC-compatible metadata format expressed in XML and using language-based element names. MODS takes a similar approach to resource description as MARC, with some rearranging, removing, and adding of data elements. MODS is frequently used as a descriptive metadata structure standard inside METS metadata wrappers for storage or exchange of digital objects.

MPEG-21 DIDL
MPEG-21 Digital Item Description Language
http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=41112

DIDL is a component of the ISO/IEC 21000-2:2005 standard MPEG-21, and as such carries the same standarization weight MPEG-2, MPEG-4, and MPEG-7 carry. DIDL is a packaging format for digital objects, defining a data model for representing both content files and their metadata, and an XML wrapper format that conforms to the DIDL data model. The DIDL data model describes Containers, which can have Items, which group Components, which group individual datastreams called Resources. Descriptors apply to Containers, Items, or Resources. While MPEG21-DIDL is much less well known in the digital library community than METS, there are some high-profile implementations, such as that at the Los Alamos National Laboratory Digital Library.

MPEG-7
MPEG Multimedia Content Description Interface
http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

MPEG-7, unline MPEG-1 and MPEG-2, is a standard for the description of the content of multimedia files, rather than a format for the multimedia files themselves. It is intended to provide structures for data both for human and machine users. The standard provides "description schemes" for a wide variety of uses. In addition to the high-level descriptions of content that will allow search and browse, there are description schemes for the creation process, rights information, technical information, user history, and low-level features such as color, lighting level, and sound timbre.

MuseumDat
http://www.museumdat.org/index.php?lu=en

MuseumDat is a metadata structure standard for museums. It is based upon CDWA Lite, but while CDWA Lite has a heavy focus on works of art and material culture, MuseumDat also is appropriate for other types of museums such as technology and natural history. MuseumDat is defined in a W3C XML Schema. The current version is 1.0. There are ongoing efforts to harmonize CDWA Lite and MuseumDat into a new format called LIDO.

MusicXML
http://www.recordare.com/xml.html

MusicXML is an XML encoding format for musical notation. It focuses on modern Western music notation, covering a full range of note types, accidentals, clefs, dynamics, and textual notations such as metronome markings and tempo indications. As such, MusicXML documents are extremely verbose and intended for machine processing rather than human consumption. MusicXML files can be structured by part or by measure. The format also includes a header for bibliographic information about the score. MusicXML is supported by many music notation software packages. The format was developed and is maintained by the company Recordare.

MXF
Material Exchange Format
http://mxf.info/

MXF is a wrapper for a large set of formats for digital audio and video maintained by the standards body Society of Motion Picture and Television Engineers (SMPTE). The primary goal of the MXF wrapper and contained data formats is to exchange digital objects and their attendant metadata between audio and video devices such as cameras and video editing packages. In contrast to many standards emerging from the cultural heritage community, MXF focuses more heavily on low-level features of audio and video such as edit decision lists in video production, and less on high-level descriptive or preservation metadata.

NewsML
News Markup Language
http://www.newsml.org/

The G2 version of NewsML is intentionally broad, covering textual news, articles, photos, graphics, audio, and video—the components that make up or express news items. Its data elements cover both factual information such as places and people but also higher-level conceptual information providing interpretation of events. NewsML is expressed both as a textual data model and an XML Schema.

OAI-ORE
Open Archives Initiative Object Re-use and Exchange
http://www.openarchives.org/ore/

OAI-ORE defines formats for the description and exchange of complex digital resources, which the framework calls Aggregations. Aggregations are then described by Resource Maps. Aggregations are groups of related content, whether different formats of the same content such as a PDF vs. a Word document, or content related by derivation such as a source data set and a paper written describing work done based on that data. OAI-ORE is explicitly designed to work with existing web technologies and therefore expose structured metadata to web-based applications. Serializations of the OAI-ORE model are available in Atom, RDF/XML, and RDFa.

OAI-PMH
Open Archives Initiative Protocol for Metadata Harvesting
http://www.openarchives.org/pmh/

The Open Archives Initiative Protocol for Metadata Harvesting is a technology used to share metadata in a mostly automated way. "Data providers" set up servers where descriptions of resources are available using requests governed by the OAI-PMH protocol, and "service providers" collect metadata from multiple data providers and create value-added services on top of the aggregated data, such as cross-repository discovery. The protocol requires at a minimum a Simple Dublin Core record for every resource exposed, but also allows supplemental metadata formats as long as they are represented by a W3C XML Schema. The OAI-PMH protocol grew out of communities wishing to share pre-prints of scientific papers, but was quickly adopted by the larger cultural heritage community. While OAI-PMH is primarily about sharing metadata, some implementers have experimented with using it to share content as well, by providing links to thumbnail images or sharing full METS packages encapsulating or linking to full digital objects.

OAIS
Open Archival Information System
http://public.ccsds.org/publications/archive/650x0b1.pdf

OAIS is known as a "reference model," defining concepts and responsibilities essential for ensuring preservation of digital information. The most well-known feature of OAIS is its categorization of information packages by their function. The Submission Information Package (SIP) is the content and metadata received from an information producer by a preservation respository. An Archival Information Package (AIP) is the set of content and metadata managed by a preservation repository, and organized in a way that allows the repository to perform preservation services. The Dissemination Information Package (DIP) is distributed to a consumer by the repository in response to a request, and may contain content spanning multiple AIPs. Preservation repository software frequently is described as "OAIS-compliant" to indicate a certain amount of functionality and standardization of features.

ODRL
Open Digital Rights Language
http://odrl.net/

ODRL is a language for encoding rights management metadata, for abstract content and for specific manifestations (formats) of that content. ODRL is designed to record in a machine-readable way the information needed for Digital Rights Management (DRM) systems. The ODRL model defines Assets, Rights, and Parties, plus the relationships between them.

ONIX
Online Information Exchange
http://www.bisg.org/what-we-do-21-15-onix-for-books.php

ONIX is a metadata standard for published material (essentialy books) that has emerged from the publishing industry. ONIX metadata is intended to accompany books throughout the supply chain, from production to retail distribution. ONIX is implemented as an XML Schema. Over 200 data elements are defined, with 31 identified as best practice to use. ONIX 3.0 was released in April 2009.

Some level of communication between the RDA and ONIX communities has occurred as part of the RDA development process. This interaction has the (as yet unrealized) potential for a greater level of partnership and data sharing between the publishing and library communities.

Ontology for Media Resource
Ontology for Media Resource
http://www.w3.org/TR/2009/WD-mediaont-10-20090618/

The Ontology for Media Resource is a W3C Working Draft designed to provide a vocabulary for media resources, especially those on the Web. A "media resource" is defined as either a tangible, retrievable resource or the abstract work represented by a tangible thing. The Ontology defines a relatively small number of core properties in RDF, including properties for basic description, technical information, and user ratings. The specification also provides mappings to a wide variety of related standards.

OpenURL
ANSI/NISO Z39.88 - The OpenURL Framework for Context-Sensitive Services
http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=d532040905f160be4697d046613f7f19a773a89e

OpenURL is a technology that facilitates the discovery of full text content by users affiliated with an institution that provides access to licensed resources. An information service such as an abstracting and indexing database might support OpenURL by providing a link in each search result in an OpenURL format that includes among other things name/value pairs with appropriate bibliographic information to identify the located resource. Once constructed, OpenURLs are then sent to link resolvers run by individual institutions with which users are affiliated, which check the bibliographic information about the located resource against a local database of licensed and open access resource. The user is then presented with a list of options for how to access different versions of the resource in print and in licensed databases.

SFX was the first mainstream OpenURL resolver used in libraries after it was purchased by Ex Libris. OCLC is the official OpenURL maintenance agency.

PB Core
Public Broadcasting Core Metadata Dictionary
http://www.pbcore.org/

PB Core is an extensive metadata structure supporting the description and exchange of media assets in the public broadcasting community, including both individual clips and full, edited, aired productions. Its elements are divided into sections focusing on intellectual content, intellectual property, instantiations, and extensions. PB Core is maintained under the auspices of the US Corporation for Public Broadcasting, and was influenced heavily by Dublin Core.

PREMIS
Preservation Metadata Implementation Strategies
http://www.loc.gov/standards/premis/

PREMIS is a data dictionary and XML Schema for the encoding of information necessary to support the digital preservation process. Its data elements are divided into 5 categories, reflecting information on the PREMIS container, objects, events, agenda, and rights. A key feature of the PREMIS model is the definition of Objects as made up of Representations, Files, and Bitstreams. Also of note is the fact that PREMIS considers Objects immutable; if an action is taken on an Object that changes it, the result is a new but related Object.

PREMIS intentionally excludes format-specific technical metadata from its scope, assuming implementers will use other relevant standards for tracking this informatin. The Library of Congress is the official PREMIS maintenance agency.

PRISM
Publisher Requirements for Industry Standard Metadata
http://prismstandard.org/

PRISM emerged from the IDEAlliance, a membership organization for the publishing industry and related companies focusing on topics such as information technology and digital content creation. The PRISM XML specification supports publishing and content aggregation workflows. As such, it provides a heavy focus on both descriptive and rights metadata. It re-uses some Dublin Core descriptive elements. The PRISM specification is formalized in XML DTDs and W3C XML Schemas, and in RDF. PRISM 2.1 was released in 2009.

QDC
Qualified Dublin Core
http://www.dublincore.org/documents/dcmi-terms/

Qualified Dublin Core, also known as DC Terms, is an extension of Simple Dublin Core through the use of additional elements, element refinements, and encoding schemes. Qualified Dublin Core is seen in widely differing implementations, often using locally-defined refinements and encoding schemes. Some digital asset management systems such as CONTENTdm and DS