

LIS 530 - ASSIGNMENT 1

METADATA SCHEMA EVALUATION

Gabor Por - April 22, 2007

<http://pgabor.com>

1. Unit of Analysis

The primary unit of analysis of MODS is the manifestation or edition. It cannot be a collection, literary unit (aka extended bibliographic family), work (aka immediate bibliographic family), because all of those entail more than one (physical) object. In contrast a MODS record always refers to one specific item. For the same reason its unit of analysis is not at the expression level.

For example MODS can specify a certain edition of a book. In this case (<http://www.loc.gov/standards/mods/v3/mods99042030.html>) it is the 1999 Cornell University Press edition of Alterman's Sound and Fury. The record does not include any other editions. Therefore the literary unit is not one of the more encompassing categories. On the other hand it does not refer to specific printing or a unique item. Hence the unit of analysis is not at these levels neither.

However the unit of analysis can be analytic. This metadata schema can be used to a part of a single unit, like a chapter in a book. (See <http://www.loc.gov/standards/mods/v3/modsbook-chapter.html>) The existence of the (Part:Extent(unit="page") "Start" and "End" elements enables this level.

2. Functions/Types

Access: In this example <http://www.loc.gov/standards/mods/v3/mods83691515.html> MODS is used for providing information of a certain map for retrieval, such as the item's title, publisher and year of publications. These are all access points users may search by.

Descriptive: The same example offers additional information of the item, such as its "Physical Description" (aka size), "Statement of Responsibility" or "Note". These correspond to the descriptive function, because they present extra information, which are not necessarily access points.

Contextual: The last elements of the example above are labeled "Record Information" and include "Record Content Source," "Record Creation Date," "Record Change Date," and "Record Identifier." All of these describe the record's history, providing context. These are not necessary for the record itself, therefore they are considered extrinsic to the resource.

Relational: In the following example of an article in a serial (<http://www.loc.gov/standards/mods/v3/modsjournal.html>) the "PartDetail:Volume" and "PartDetail:Issue" elements tell the user which volume and issue of the serial the article can be found, respectively. This signifies the relations to other issues and volumes of the same serial. Also, one of the top level elements is "relatedItem", which provides additional relational functions.

Structural: The music example of MODS (<http://www.loc.gov/standards/mods/v3/mods85753651.html>) identifies three separate arias of notated music. In other words the "Constituent" element is used for showing the structure of the record.

Administrative function: The book example (<http://www.loc.gov/standards/mods/v3/mods99042030.html>) includes the "Access Condition:restrictionOnAccess" element. This describes access restrictions, which is one of the administrative functions.

3. Principles

I see no reason why the three principles of descriptions should not apply in the utilization of MODS. They are not explicitly mentioned in the explanation of MODS, but looking at the examples I am reasonably sure they are followed.

Title Page Sanctity: When the MODS is used for describing anything that has the equivalent of a title page I assume that the principle of the title page sanctity is observed and the element values are represented exactly the same way as they appear on the title page without any change, correction, or embellishment.

Transcription: Similarly to the above, the content of the records should be copied directly from the document itself and not from a surrogate or substitute.

Chief Source of Information: The records preserved in MODS may follow specific standards in regard what they consider the chief source of information, from where content has to be transcribed. For example, according to AACR2 the title of a book needs to be transcribed form the title page and not the spine, or cover) of the book.

Duval et al identified four metadata principles: modularity, extensibility, refinement and multilingualism. I will write about first two under point 4b.

Refinement: Duval formulated this principle this way: "The design of metadata standards should allow schema designers to choose a level of detail appropriate to a given application." Considering that MODS is a subset of MARC 21 it is only refinable to the extent that MARC21 is. I read through the MARC 21 Principles (<http://www.loc.gov/marc/96principl.html>) and found no reference to varying degree of refinements. MARC 21 is extensive, but it does not allocate for future refinements. Therefore neither does MODS.

Multilingualism: I looked through the list of 23 registered MODS implementation (<http://www.loc.gov/standards/mods/registry.php>) With three exceptions (Chile, Greece, German) they all came from organizations located in countries where English is the first spoken language. I attempted to find whether the elements at the exceptions are in English or not, but failed to locate the actual files. I suspect that they may be in English, in which case MODS does not satisfy the multilingual principle. MARC 21 uses numeric tags while MODS is using language-based. The former is more universal therefore in this regard, because (Arabic, as opposed to Roman) numerals are symbols which are recognized in most written languages. MARC21 can be considered language independent, while MODS cannot be.

4a. Form of element values

Before going into the details of form elements values it is important to note that "All elements and attributes are optional. Attributes are not in a mandated sequence and not repeatable."

(<http://www.loc.gov/standards/mods/mods-outline.html>)

There are elements that can only have values from an enumerated list, i.e. from a controlled vocabulary. An example of this is "typeOfResource", which can be any one of these: text, cartographic, notated music, sound recording-musical, sound recording-nonmusical, sound recording, still image, moving image, three dimensional object, software, multimedia, or mixed material.

Authority control is also practiced, e.g. the "genre" element's value has to be in the relevant MARC authority file: <http://www.loc.gov/marc/sourcecode/genre/genresource.html>

For other elements MODS suggest a type of value. For example here is the list for the type of the "identifier" is suggestions: "hdl, doi, isbn, isrc, ismn, issn, issue number, istc, lccn, local, matrix number, music publisher, music plate, sici, uri, upc, videorecording identifier, stock number."

A final example for element value related rules is the order of the "hierarchicalGeographic" subelement of the "subject" top level element. The value has to follow this order/hierarchy: Continent, country, province, region, state, territory, county, city, island, and area.

4b. Instructions

MODS is an XML schema, therefore it has to work like all other XML schemas.

Interoperability: The Metadata Encoding and Transmission Standard (METS) (<http://www.loc.gov/standards/mets/>) is designed with interoperability in mind. It encapsulates the data presented through the MODS XML schema and through this wrapper allows a greater degree of interoperability with other schemas and applications.

Modularity: As mentioned above all elements and attributes of MODS are optional. Furthermore the elements do not have to follow any specific, predefined order. The combination of these two features results in a high degree of modularity. The records can be used to create building blocks of information. One of the advantages of this approach is that simplified MARC records can be shared in XML. That in turn can be reformatted according to the end user's needs. The record can also be imported and/or intertwined with other XML schemas, such as METS.

Extensibility: MODS is not extensible by the users. But its development is ongoing. For example the latest draft of version 3.3. is ten days old, dated April 12, 2007. The process of the schema development includes user input. This is the limit the users can influence its extension.

4c. Professional audience

The intended users of this schema are individuals and organizations, who have to maintain large collection of records. Their system may or may not already be MARC21 compliant. By working with MODS it will be. If they are already MARC21 compatible than implementing MODS can be easily down using the MARC to MODS mapping (<http://www.loc.gov/standards/mods/mods-mapping.html>) conversion tool. A warning is due though: while that mapping lists all equivalences, but "is not intended to be a crosswalk that allows for bi-directional conversions without some loss of data."

Browsing through the list of implementations I noticed that almost all of them are universities or university libraries. I conclude that they were the primary audience targeted. On the other hand I was surprised to see that only one Congress (107th) was archived using this schema. I would have thought that once they deployed the system it would have been a relatively simple task, using the same technology, to archive the materials of later Congresses (108th and 109th. The current Congress is the 110th.)

To conclude: the target audience of MODS is universities, libraries and government organizations.

5. Features

The schema is maintained by the Network Development and MARC Standards Office of the Library of Congress. The Library of Congress was established in 1800. It established an office for network development in 1976 and expanded it to cover MARC in 1984 (See <http://www.loc.gov/marc/ndmsso.html>) Based on this information I do not doubt the longevity of the maintaining organizations. Furthermore they develop the standard with the help of user input. This makes the project not only more democratic, but also creates a feeling of ownership by the individuals and organization that provide input. This characteristic is essential for wider adoption.