METADATA AND DIGITAL CURATION

The purpose of this lesson is to introduce you to the use of metadata for digital curation and to make you aware of other several schema that are used for certain kinds of materials and/or tasks in this field. Metadata is used for a variety of purposes, such as the retrieval of digital objects, the management of digital resources, and the preservation of digital objects. Thus metadata is essential to Digital Curation.

Before getting to the formal schema, we should say something about social tagging. In last week's lesson, you were introduced to the use of crowdsourcing for the collection of metadata. In your readings this week, there are two articles with different views on the use of folksonomies (metadata collected through crowdsourcing). One of the main differences between the use of taxonomies with controlled vocabularies and folksonomies is that the former places the descriptive term in a hierarchical structure that defines its relationship to the terms above, below and beside it in the structure. These relationships are sometimes referred to as "semantic" relationships. This context makes it easier to avoid the kind of miscommunication that can occur with any form of written communication.

The trend in cultural institutions has been to augment traditional cataloging with social tagging, when the latter is employed. Using social tagging alone as the means for providing intellectual access to digital objects in less common, however, there have been some efforts map to folksonomy terms to controlled vocabulary terms so as to place these terms within taxonomies. There have also been efforts to structure social tagging so that it maps to various metadata schema.

The primary metadata schema used for digital curation are METS, MODS, PREMIS, and Dublin Core, although there are others used for specific types of materials or in specific contexts. These will be covered in much more detail in the courses on Metadata and Digital Preservation, so we do not expect that you will become competent in the use of these from this one lesson, however, you will hopefully come to understand, if you don't already, the primary functions for each of these.

If you are new to metadata (which just means data about data) this may seem a bit intimidating, but, in fact, metadata is used all around us all the time and standards for different disciplines and industries abound. Please take a look at the poster, "Seeing Standards: A Visualization of the Metadata Universe, " http://www.dlib.indiana.edu/~jenlrile/metadatamap/seeingstandards.pdf to see just how pervasive these are.

And, if you doubt the importance of these standards, please viewe this video ("Metadata and Nuclear Disaster", a tongue-in-cheek look at the importance of metadata]:
http://www.youtube.com/watch?v=pbBa6Oam7-w

The primary purpose of these standards is the usability of data across different contexts.   With the growth of the Internet, it became apparent that there was a the need for simplified metdata structure to support description of various types of materials that were being represented in regional and national bibliographic databases and on the web.   Dublin Core was developed as a response to this need. By comparison with the metadata included in MARC21, the library standard, it was small set of descriptive metadata.   The fields have been organized into three categories as follows:
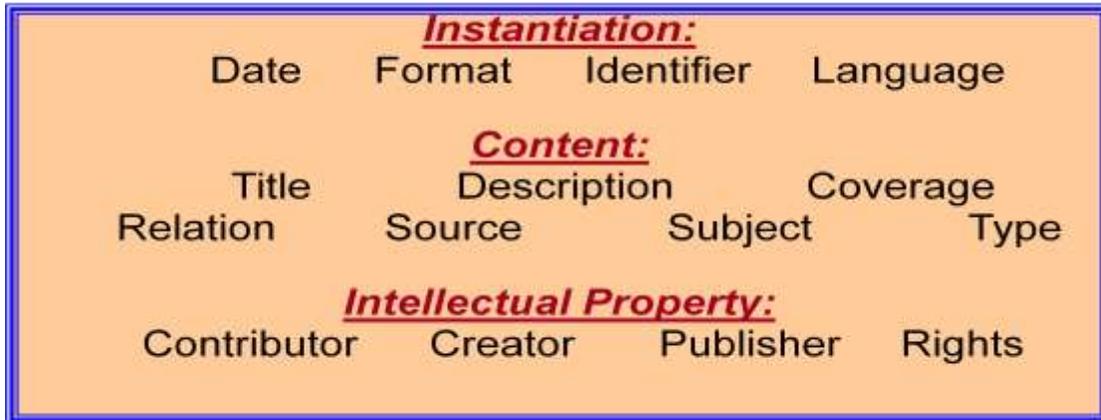

Developed by the Library of Congress, the purpose of MODS (Metadata Object Description Schema) is essentially the same as Dublin Core: to organize descriptive metadata in a way that will allow of interoperability, however it is has added a number of elements that exist in MARC that do not exist in Dublin Core.   This allows libraries, and other institutions that use bibliographic databases, to export more of the data from this utilities for use in describing digital objects.

METS (Metadata Encoding and Transmission Standards) is an XML "wrapper" or "envelope" in which metadata about a digital object can be stored and remain in association with that object. It is often used in tandem with MODS, but also accommodates additional metadata, such as administrative metadata and preservation metadata.   This is an outline of the METS, showing the different sections in which the metadata can be input.
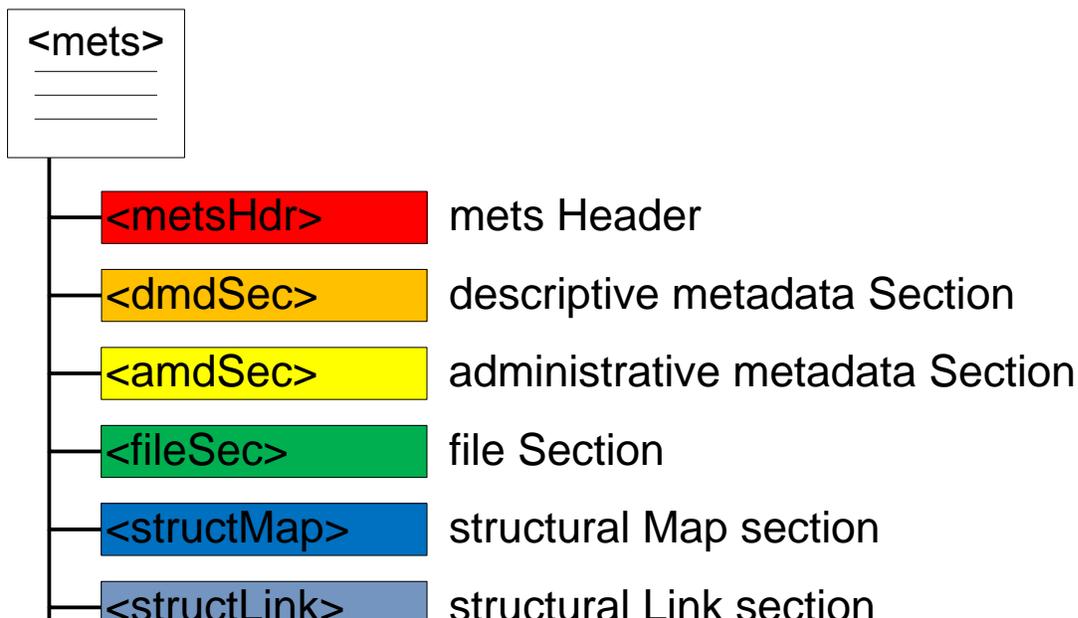
PREMIS (Preservation Metadata Implementation Strategy) is a schema and data dictionary for recording metadata required for the ongoing preservation of digital objects. It supports the linking of multiple files in different formats that are part of the same intellectual object, such as different files use for a web-page or site. It provides elements for keeping a record of any action taken in relationship to an object (such a migration to another format) and allows for recording intellectual property rights. Like MODS, this is often used in conjunction with METS.
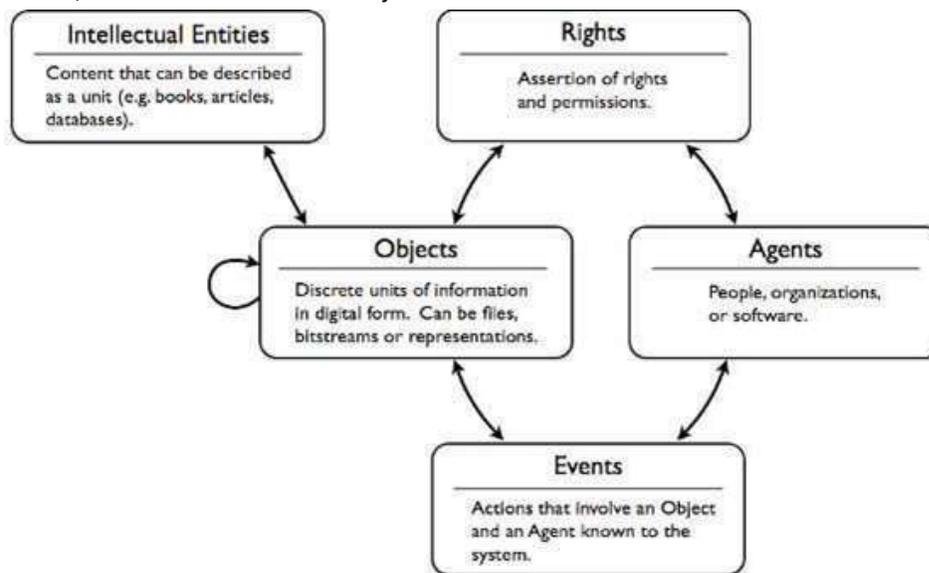
conjunction with

You will no doubt notice that there is quite a bit of overlap between these and other metadata standards. It is quite possible to map metadata from one schema to another and a number of "crosswalks" have been developed to aid this process. Please take a look at the metadata crosswalk developed by the Getty Museum, which maps metadata across a number of different standards:

http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html

Since you have quite a bit of material to cover this week, the only assignment will be to post on at least 3 of the readings.