



Metadata with Levels of Description: New Challenges to Catalogers and Metadata Librarians

Myung-Ja Han

Assistant Professor/Metadata Librarian
University of Illinois at Urbana-Champaign
Urbana, Illinois, USA
E-mail: mhan3[at]illinois.edu

Meeting:

80 — *Inspired moments in cataloguing* — Cataloguing

Abstract:

As libraries collect and provide more resources in digital formats, new units of granularity have emerged in access and description. Digital surrogates or born-digital resources make it possible for users to access resources at more granular levels, e.g., a book or journal title can be accessed at chapter level, article level, and page level in addition to the book or journal title as a whole. To provide granular levels of access, metadata should be created at granular levels as well. This paper discusses why the emerging need for granularity of access and description makes the cataloging and metadata process a highly collaborative work, and suggests a way to design and create a metadata schema for describing granular levels of resources.

Background

Libraries have a long history of organizing and managing their resources in a structured way, i.e., cataloging, which is then used to provide access to resources. Library cataloging has evolved over time: from the index card, which featured a main entry, added entry(ies), a call number, and

classification information for each item on a 3x5 inch card, to using multiple sets of elements, e.g., metadata standards, including MACHine-Readable Cataloging (MARC), with a content standard when creating metadata for the resources that libraries hold.

Since it was first introduced in the mid-1960s, MARC has been used as the main metadata standard for bibliographic description in the library domain using Anglo-American Cataloguing Rule (AACR) as a content standard. Because MARC has a rich set of semantics, it has been used to describe printed books, serials, maps, and microforms without any problems. However, as libraries have started collecting more resources in diverse formats, notably in audio/video and electronic/digital formats, the need for facilitating other metadata standards to best describe resources in non-print formats has increased due to the fact that some of the resources would be described better in metadata standards other than MARC for their technical information. In addition, these resources would be better accessed via systems other than integrated library systems (ILSs), as the majority of ILSs used by libraries are designed for only MARC format records. Consequently, metadata standards such as Dublin Core, Metadata Object Description Schema (MODS), Encoded Archival Description (EAD), and Visual Resource Association (VRA) Core, have been widely used for describing special formats of resources in digital collections, and libraries have been developing local access systems other than ILSs for their unique digital collections. According to a survey conducted by the Association of College and Research Libraries (ACRL), most college and research libraries use other metadata standards along with MARC to describe resources in digital formats, which requires elements for describing format specific information or relationships between original and digitally re-formatted resources, and then house these resources in systems other than their ILSs (Ma 2007). Also, as many libraries have started digitizing the resources they house in special collections units or archives, notably images or manuscript collections, libraries are creating metadata not only for items included in the collections, but also for collections or collection level finding aids in Dublin Core Collections Application Profile, EAD, or MODS, to enhance the discoverability of the resources. However, selecting a metadata standard to best describe the resources in a collection is always a challenge to catalogers and metadata librarians.

Decisions as to which metadata standard and content standard to use can be determined by several factors, including the characteristics of the resources, the systems where the metadata and digital contents are stored and accessed, the domains where the resources will be used, and the functions of the metadata. In addition, the available information and the catalogers' subject knowledge can be further factors to consider because these have a direct impact on metadata content. "*Seeing Standards: A Visualization of the Metadata Universe*" created by Jenn Riley (2010) showcased the metadata and content standards that are currently available for use. The list includes more than a hundred metadata standards used in not only the library domain, but also other cultural heritage institutions, including museums and archives. It also includes metadata standards that are used for functions other than descriptive purposes. Although there are already many specific metadata standards to choose from, there is still a need for yet another metadata standard that will work for more specific and at more granular levels of descriptions to provide granular levels of access services.

Granular Levels of Description for Granular Levels of Access Service

"Virtually any content we digitize and make available to our clientele requires metadata for discovery and access (Tennant 2002, p. 32)." Metadata makes it possible for all of our resources to be searchable and retrievable by users in an optimal way. Active discussions on the importance of creating high-quality and consistent metadata are also related to the library's efforts to provide the best access services. The levels of access services that libraries can provide also depend on the information the metadata contains.

Until recently, the default level of access that libraries provided was book or journal title - or issue/volume of the journal in some cases. In other words, "in a physical library of printed books, the volume-on-the-shelf is the dominant unit of granularity," (Cole, Han, and Vannoy 2012). By the same token, the default level of description for resources, metadata, was book or journal title level. However, as the number of specialized databases for articles and electronic resources purchased by libraries has increased, users now have access to more granular levels of resources, such as chapter or article level, and they expect libraries to provide even more granular levels of access to their digitized resources on top of collection, book, or journal title level access. In order

to provide granular levels of access services, libraries must develop new metadata standards that are designed to accommodate the granular levels of description that users want access to.

Granularity of elements

Research on the granularity of metadata has focused on two areas: the granularity of elements and the granularity of content values added into the metadata. Roy Tennant (2004) examined the granularity of elements focusing on name elements in MARC and Text Encoding Initiative (TEI). He pointed out that MARC does not have semantics for distinguishing the first name and last name of the author in the data field 100. He added that TEI provides granular elements in the types of documents, and that name information can be added to a different part of the TEI metadata with the elements <name> or <author>. However, there are no sub-elements or attributes for distinguishing the first name from the last name, which can help machines parse them in a more granular way. As compared to MARC and TEI, a top-level element <name> in MODS has the sub-element <namePart> that allows the separation of the first and last names using an attribute <type>. Dublin Core, one of the most used metadata standards other than MARC, also does not have sub-elements or attributes for the <creator> element (Ma 2007). Table 1 shows how these four metadata standards have different granularities for the element used to describe names associated with the resource.

Metadata Standard	Describing <name>
MARC ¹	100 1 _ \$a Last name, First name. \$d 1111-1222, \$e role.
TEI	<name type="person">First name and Last name</name> or, <docAuthor>Last name, First name.</docAuthor>
MODS ²	<name type="personal"> <namePart type="given">Last name</namePart> <namePart type="family">First name</namePart> <role> <roleTerm type="code" authority="marcrelator">aut</roleTerm> <roleTerm type="text" authority="marcrelator">author</roleTerm> </role> </name>
Dublin Core	<dc:creator>Last name, First name. Date.</dc:creator>

Table 1. Describing <name> information in MARC, TEI, Dublin Core, and MODS

¹ MARC data field 100 has 19 subfields that can be used for additional information about both the name and work.

² MODS <name> element has three other additional sub-elements other than <namePart> and <role>; <displayForm> <affiliation>, and <description>. <http://www.loc.gov/standards/mods/userguide/name.html>

One way to solve these semantic limitations is through extensibility of the element. Depending on the metadata standards, the schema allows users to extend the element set to meet local needs, i.e., Dublin Core encourages users to extend local elements according to its Status Report published in March 2011.³

Granularity of content values

Discussion on how to provide content values to metadata records is another aspect of the metadata granularity issue, especially with regards to the subject of the resources they describe. When adding subject terms, the Library of Congress recommends the *twenty percent rule*, i.e., assigning subject headings only for topics that comprise at least twenty percent of the work (Library of Congress 2008). The Library of Congress also recommends the *Rule of three* and the *Rule of four*. The *Rule of three* recommends that if a general topic includes in its scope more than three subtopics, but the work only discusses two or three of these topics, assign the appropriate two or three headings rather than the broader heading. The rule goes on to say that if more than three of the subtopics are discussed in the work, assign the broad heading instead, unless the *Rule of four* applies. The *Rule of four* states that subtopics cannot be exceeded more than four, which could cause the absence of subject headings that might help users discover important resources for their research or pedagogy (Library of Congress 2008). These rules can be applied well when the unit of granularity is at book or journal title level, and the descriptive metadata is created for this specific unit of granularity, in which case it is nearly impossible to include subject headings that describe only a small portion of the resource. However, in digital library environments, it is required to have more granular levels of description to improve the access and discovery of resources. According to the CONTENTdm Metadata Best Practices (2009), when describing digital resources, especially manuscripts or book collections, the Library of Congress' *twenty percent rule* should be compromised to promote the discovery of resources with specific topics of subject. To make digital resources more easily searchable and retrievable by users, libraries are now adding more specific subject headings into metadata although the heading applies to only a small part of the work it describes (p. 27), i.e., libraries add granular levels of subject terms to increase the discoverability of their digitized resources.

³ <http://dublincore.org/news/communications/statusreports/2011/03/index.shtml>

New Units of Granularity

New units of granularity have emerged as libraries purchase more resources in digital formats, and actively produce digital resources by creating digital collections and participating in massive book digitization projects which expand the corpus of the digital resources that humanities scholars can access and use. In addition, the active development of digital humanities projects and the increasing use of digital resources in scholars' research and teaching have made libraries reexamine the level of description and access services.

The traditional unit of granularity – volume-on-the-shelf – is no longer the only form of granularity. Libraries can now provide access to resources in different granularities, such as items included in digital collections, articles contained in a journal, chapters of a book, or a table of contents. All of these have become new units of granularity. The granular levels of access are also what users want from libraries' discovery services. In their usability test on next generation cataloging discovery systems, Yang and Hoffman (2011) found that users like to have article and chapter level access available from the new discovery systems in addition to other services, such as faceted navigating services and federated search functions.

As interest in and development of digital humanities grow, new units of granularity are also emerging. Although mass book digitization projects make it possible for these digitized books to be viewable in a page-by-page manner in a web environment as surrogates for printed books, humanities scholars want more granular levels of access to specialized collections of books, such as Renaissance Emblem Books or period plays, rather than only book level access. They want the digital surrogates to be curated in a certain way, i.e., with a new unit of granularity. These new units obviously require structural and contextual analysis by scholars, and must have the right levels of descriptive metadata. To meet the scholars' needs and users' expectations, libraries should find a sustainable method of creating metadata that would contain descriptions for the new units of granularity. This requires metadata and cataloging librarians to develop a new schema with a new element set whenever a new need arises. The question is how to work with scholars and users, and develop a new metadata schema with a new element set and a content standard.

Creating or Extending Metadata Schemas

Granular levels of description need a new container, a schema, which defines the elements, attributes, entity, encodings schemes allowed in a metadata record, and relationships between them, to make a metadata record that stores granular levels of information. In addition, the schema can specify constraints on the types of data an element may contain, which results in a structured and valid metadata record. The new elements required for granular levels of description can be added into pre-established metadata schemas with the most common sets of elements and allow metadata extensibility, such as Dublin Core. A new schema can be created by mixing and matching semantics from different metadata standards as the method used for creating application profiles (Heery and Patel 2000). If not, a completely new metadata schema can be created with new semantics and syntax to meet the specific needs. Regardless of the methods, the schema approach is highly desired because it ensures the consistency of metadata quality over time and increases the interoperability and shareability of metadata with a wider user group. SPINE is an example of a schema specially designed for describing a resource with the images and text needed for granular levels of access.

Mass book digitization projects produced digital surrogates of the many emblem books published between the 16th and 18th century in Europe, which until now weren't available to emblem scholars unless they visited the institutions where the printed copies were housed. A new metadata schema called SPINE was developed for describing the digital emblem books and the emblems contained in the books by mixing and matching semantics from different metadata schemas. The emblems contained in a book are the main content that emblem scholars are interested in finding and using. An emblem consists of three different parts - motto (title), pictura (image), and subscriptio (associated text). In digitized emblem books, scholars want emblem level description that would enable emblem level access in addition to book level access. In order to incorporate scholars' needs, the schema borrowed the element sets that describe an emblem and related parts published by Stephen Rawles (2004) titled *A Spine of Information Headings for Emblem-Related Electronic Resources*. The XML schema for SPINE version 1.0⁴, developed in 2007, was based on these emblem specific elements and semantics from TEI for book level

⁴ <http://diglib.hab.de/rules/schema/emblem/emblem.xsd>

description (TEI Header) and the texts of the book (other TEI elements such as <p> for paragraph). The current SPINE schema has been updated to version 1.2 with additional elements, and incorporates linked data technologies (Stacker 2012). Table 2 shows how the SPINE schema uses the different namespaces from different metadata standards to accomplish granular levels of description.

Book Level Description	Digitized Copy Description	Emblem Level Description
MODS TEI Header (*Both are transformed from MARC format bibliographic records)	SPINE of Information <copyDesc> <i>Sub elements:</i> <copyID> <owner> <digDesc> <i>Attributes:</i> countryCode, comp, scope, xml:id, globalID, pageImages	SPINE of Information <emblem> <i>Sub elements:</i> <motto> <transcription> <pictura> <iconclass> SKOS/RDF <skos:notation> <skos:prefLabel> <subscriptio> <commentatio> <keyword> ... <i>Attributes:</i> xmlns:xlink, xml:id, citeNo, xlink:href, globalID, xml:lang, rdf:about

Table 2. Elements and metadata schemas used in SPINE schema

As mentioned above, the SPINE schema borrows semantics from other already established schemas, i.e., it supports book level description in either MODS or TEI Header. It also uses semantics from Resource Description Framework (RDF) and Simple Knowledge Organization System (SKOS) to enhance the user experience in searching and browsing controlled terms. Because SPINE uses pre-established and widely used metadata schemas, any institutions that have a collection of digitized emblem books can use the SPINE schema for book level description even if the emblem level descriptions are not ready yet, which makes book level access possible through any service portals specializing in emblem books or Renaissance literature with images and texts.


```

<biblioDesc>
  <mods>
    <mods:titleInfo>
      <mods:title>XL [i.e. Quadraginta] emblemata miscella nova</mods:title>
    </mods:titleInfo>
    <mods:physicalDescription>
      <mods:digitalOrigin>reformatted digital</mods:digitalOrigin>
      <mods:form authority="marcform">print</mods:form>
      <mods:extent>[8], xxx p. : 41 ill. ; 20 cm.</mods:extent>
    </mods:physicalDescription>
    ...
  </mods>
  <copyDesc>
    <copyID>uiu2895515</copyID>
    <owner countryCode="US">University of Illinois</owner>
    <digDesc comp="complete" scope="all" xml:id="xliequadragintae00mure"
      globalID="http://hdl.handle.net/10111/UIUCOCA:xliequadragintae00mure">
      <copyID>10111/UIUCOCA:xliequadragintae00mure</copyID>
      <owner countryCode="US">University of Illinois</owner>
    </digDesc>
    ...
  </copyDesc>
  <emblem xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#" xml:id="E000944" citeNo="1."
    globalID="http://hdl.handle.net/10111/EmblemRegistry:E000944">
    <motto><transcription xml:lang="de">
      Alchimisterey:<normalisation>xml:lang="de">Alchemie:</normalisation></transcription>
    </motto>
    <pictura xml:id="E000944_P1">
      <iconclass rdf:about="http://www.iconclass.org/rkd/31A247">
        <skos:notation>31A247</skos:notation>
        <skos:prefLabel>looking over the shoulder</skos:prefLabel>
      </iconclass>
      ...
    </picture>
  </emblem>
</biblioDesc>

```

Figure 1. SPINE metadata record has granular levels of descriptions for a book and emblems contained in the book titled, *XL [i.e. Quadraginta] Emblemata Miscella Nova*

In addition, since there are established crosswalks and transformation stylesheets from MARC to TEI Header and from MARC to MODS⁵, adding book level descriptions to SPINE metadata records is a relatively simple task, which is a benefit of using established metadata schemas when creating a new schema. Figure 1 is an example of a SPINE metadata record with different levels of granularity, describing a book, a digitized copy, and an emblem contained in the book.

⁵ <http://www.loc.gov/standards/marcxml/>

Implementing linked data technologies, e.g., RDF and SKOS namespaces provides yet another layer of granular access service using a controlled vocabulary called Iconclass⁶, a multilingual classification system for cultural heritage art and images. Iconclass has a hierarchical structure and its vocabularies are currently available in five different languages. By using RDF and SKOS, emblem scholars can browse Iconclass vocabularies in different languages and view hierarchical relationships between vocabularies. RDF and SKOS also work as tools to enhance metadata in a more granular way. As discussed above, since Iconclass vocabularies are available in multiple languages and RDF includes this information, as shown in figure 2, metadata can now have terms in different languages, in addition to the vocabularies in the original language.

```
<rdf:RDF>
  <rdf:Description rdf:about="http://iconclass.org/25F33%28EAGLE%29">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="fr">oiseaux de proie : aigle</skos:prefLabel>
    <skos:prefLabel xml:lang="en">predatory birds: eagle</skos:prefLabel>
    <skos:prefLabel xml:lang="de">Greifvögel: Adler</skos:prefLabel>
    <skos:prefLabel xml:lang="it">uccelli rapaci: aquila</skos:prefLabel>
    <skos:prefLabel xml:lang="fi">petolinnut: kotka</skos:prefLabel>
    ...
  </rdf:Description>
</rdf:RDF>
```

Figure 2. Iconclass vocabularies are available in multiple languages

The SPINE schema development example introduces two changes that catalogers and metadata librarians must work with. First, XML has become a driving force in cataloging and metadata workflows. A new metadata schema for granular levels of description can be developed with XML, and XML can help enforce and assure metadata quality and consistency in cataloging workflows. For this reason, knowledge of XML is one of the most commonly requested skills of catalogers and metadata librarians (Han and Hswe 2010). Additionally, other XML related technologies and emerging trends such as RDF and semantic web technologies are also integrated into common cataloging and metadata workflows. Second, the SPINE example provides clear evidence that the metadata design and creation processes require a highly collaborative effort between librarians, users (scholars and students), and domain specialists (or

⁶ <http://www.iconclass.nl/home>

subject specialists), which in turn requires good communication skills and outreach effort from catalogers and metadata librarians.

Metadata Design and Creation as a Collaborative Effort

Designing a metadata schema for granular levels of description can (only) be possible through collaboration among catalogers and metadata librarians, users, and domain specialist. Analysis of metadata content and data structure and user behavior are essential components of designing a new metadata schema, because these factors determine the granularity of the access levels. In addition, quality metadata can be created only with subject knowledge that comes from domain specialists. Catalogers and metadata librarians can contribute the foundation for information organization, users can provide the way information can be used for discovering resources as well as desired levels of granularity, and domain specialists can create metadata that requires specific subject knowledge as shown in Figure 3 below.

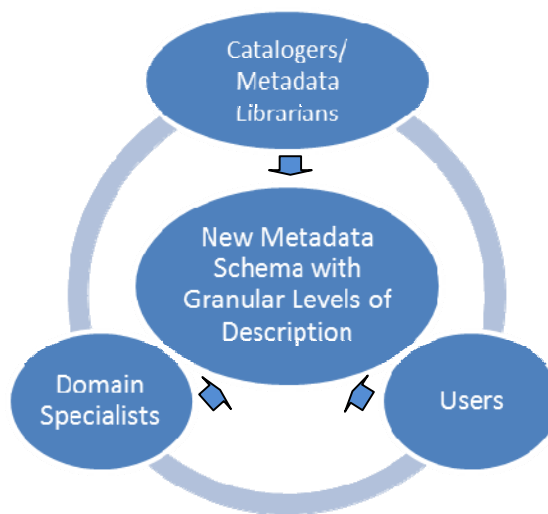


Figure 3. Developing a new metadata schema with granular levels of description

Conclusion

Metadata is an integral part of designing a library’s access service. Depending on the granularity of the descriptions, the granularity of access services can be determined. At the same time, a metadata design must be determined to best serve users’ needs. As libraries collect and create more resources in digital formats, the granularity of access services that libraries provide for

users has changed since users' expectations and needs have evolved over time. Users want digital surrogates to be curated and described in a specific way. Depending on the content and structure of the resources, libraries should develop a new metadata model to accommodate the new levels of granularity in access services. The new metadata schema can be created by mixing and matching metadata schemas or with completely new semantics and syntaxes in XML. However, more importantly, metadata schema design and development and the creation of metadata with granular levels of description have become a highly collaborative effort among three groups: users, domains specialists, and librarians, each of which contributes in their respective area of expertise. As libraries try to better serve their users in a digital environment, catalogers and metadata librarians are required to work with and develop new metadata schemas to enable more granular levels of descriptions for curated digital resources, which also require catalogers and metadata librarians to be equipped with not only information technologies, but also good communication skills to collaborate with scholars and domain specialists.

References

- Cole, T. W., Han, M-J, & Vannoy, J. (2012). Descriptive Metadata, Iconclass, and Digitized Emblem Literature. *Proceedings of Joint Congress of Digital Libraries 2012*, 120-120.
- CONTENTdm Metadata Best Practices. (2009). Online. Available at http://contentdmmwg.wikispaces.com/file/view/Best_Practices_2+0.pdf
- Han, M. J., & Hswe, P. (2010). The Evolving Role of the Metadata Librarian: Competencies Found in Job Descriptions from 2000 to 2008. *Library Resources and Technical Services*, 54/3, 129-141.
- Library of Congress. (2008). *Subject Headings Manual* (2008 ed.). Washington, D.C: Library of Congress, Cataloging Policy and Support Office.
- Ma, J., & Association of Research Libraries. (2007). *Metadata*. SPEC kit. Washington, D.C: Association of Research Libraries.
- Heery, R., & Patel, M. (2000). Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne*, Issue 25. Online. Available at <http://www.ariadne.ac.uk/issue25/app-profiles/>
- Rawles, S. (2004) *A Spine of Information Headings for Emblem-Related Electronic Resources*. In: Wade, M.R.(ed.) *Digital Collections and the Management of Knowledge*:

Renaissance Emblem Literature as a Case Study for the Digitization of Rare Texts and Images. DigiCULT, Salzburg, Austria.

Riley, J. (2010). *Seeing Standards: A Visualization of the Metadata Landscape*. Available from <http://www.dlib.indiana.edu/~jenrile/metadatamap/>

Tennant, R. (2002). The Importance of Being Granular. *Library Journal*, 127(9), 32.

Tennant, R. (2004). A Bibliographic Metadata Infrastructure for the Twenty-First Century. *Library Hi Tech*, 22(2), 175-181.