Abstract

This is the second part of an analysis of the methods that have been used to achieve or improve interoperability among metadata schemas and their applications in order to facilitate the conversion and exchange of metadata and to enable cross-domain metadata harvesting and federated searches. From a methodological point of view, implementing interoperability may be considered at different levels of operation: schema level (discussed in Part I of the article), record level (discussed in Part II of this article), and repository level (also discussed in Part II). The results of efforts to improve interoperability may be observed from different perspectives as well, including element-based and value-based approaches.

1. Introduction

As discussed in Part I of this study, the results of efforts to improve interoperability can be observed at different levels:

1. Schema level – Efforts are focused on the elements of the schemas, being independent of any applications. The results usually appear as derived element sets or encoded schemas, crosswalks, application profiles, and element registries.

2. Record level – Efforts are intended to integrate the metadata records through the mapping of the elements according to the semantic meanings of these elements. Common results include converted records and new records resulting from combining values of existing records.

3. Repository level – With harvested or integrated records from varying sources, efforts at this level focus on mapping value strings associated with particular elements (e.g., terms associated with
In the following sections, we will continue to analyze interoperability efforts and methodologies, focusing on the record level and the repository level. It should be noted that the models to be discussed in this article are not always mutually exclusive. Sometimes, within a particular project, more than one method may be used.

2. Achieving Interoperability at the Record Level

In some cases, a project may have started within a community without being aware of other projects that had similar material types, audiences, or subject domains. Often, a particular metadata schema had been developed or adopted for a project, and metadata records had already been created before the issue of interoperability was carefully considered. As a result, when different projects needed to be integrated or mapped, it was too late for the projects to consider any interoperability approach at the schema level (see discussions at Part I of this article). Converting metadata records becomes one of the few options for integrating established metadata databases. More recent projects also have attempted to reuse existing metadata records and combine them (or their components) with other types of metadata records (or their components) to create new records. Activities at the record level focus on integrating or converting data values associated with specific elements/fields, with or without a normalization process. The two common methods of achieving interoperability at the record level are conversion and data integration.

2.1 Conversion of Metadata Records

The major challenge in converting records prepared according to a particular metadata scheme into records based on another schema (see Figure 1) is how to minimize loss or distortion of data. Various tools have been created to facilitate such conversion. For example, the Library of Congress provides tools (available at <http://www.loc.gov/standards/mods/> to convert between the MARC record and the MODS record, and between the DC record and the MODS record.

The Picture Australia project serves as a good example of data conversion. It is a digital library project encompassing a variety of institutions, including libraries, the National Archives, and the Australian War Memorial, many of which came with legacy metadata records prepared under different standards. Records from participants are collected in a central location (the National Library of Australia) and then translated into a "common record format," with fields based on the Dublin Core [Tennant, 2001].

Another good example is the National Science Digital Library (NSDL) Metadata Repository where metadata records from various collections were harvested. The following figures show the result of an
ADL (Alexandria Digital Library) metadata record (Figure 2) being converted into a Dublin Core record (Figure 3) when these records were harvested by the NSDL Metadata Repository.

<table>
<thead>
<tr>
<th><strong>Identifier for this record:</strong></th>
<th>doqq_bw_ca:O3511664.SWS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>titles:</strong></td>
<td>[FGDC] Citation - Title</td>
</tr>
<tr>
<td></td>
<td>[DOQQ, Digital orthophoto quarter quadrangle], Soda Lake South SW, California.</td>
</tr>
<tr>
<td><strong>geographic-locations:</strong></td>
<td>[FGDC] Bounding Coordinates</td>
</tr>
<tr>
<td></td>
<td>Lat/Lon Bounding Box: North=35.0625 South=35.0 East=-116.0625 West=-116.125</td>
</tr>
<tr>
<td><strong>dates:</strong></td>
<td>[FGDC] Time Period of Content - Date Range</td>
</tr>
<tr>
<td></td>
<td>Begin:1997-02-04 End:1997-02-04</td>
</tr>
<tr>
<td><strong>types:</strong></td>
<td>[FGDC] Citation - Geospatial Data Presentation Form</td>
</tr>
<tr>
<td></td>
<td>aerial photographs</td>
</tr>
<tr>
<td><strong>formats:</strong></td>
<td>[FGDC] Format Name</td>
</tr>
<tr>
<td></td>
<td>DOQQ</td>
</tr>
<tr>
<td><strong>subject-related-text:</strong></td>
<td>[ADL] subject-related-text</td>
</tr>
<tr>
<td></td>
<td>DOQQ; quad; Band interleaved by line; BIL</td>
</tr>
<tr>
<td><strong>assigned-terms:</strong></td>
<td>[FGDC] Theme Keyword</td>
</tr>
<tr>
<td></td>
<td>DOQ; DOQQ; digital orthophoto; digital orthophotoquad; digital image map; aerial photograph; paper map; map; Universal Transverse Mercator (UTM); digital image</td>
</tr>
<tr>
<td></td>
<td>[FGDC] Place Keyword</td>
</tr>
<tr>
<td></td>
<td>United States; California</td>
</tr>
<tr>
<td><strong>originators:</strong></td>
<td>[FGDC] Citation - Originator</td>
</tr>
<tr>
<td></td>
<td>U.S. Geological Survey (USGS)</td>
</tr>
<tr>
<td></td>
<td>[DOQQ] PRODUCER</td>
</tr>
<tr>
<td></td>
<td>Analytical Surveys, Inc.</td>
</tr>
<tr>
<td></td>
<td>[DOQQ] AGENCY</td>
</tr>
</tbody>
</table>
When converting an ADL record into a DC-based record for display, value strings in the ADL elements are displayed in equivalent DC-elements. For example, coordinates recorded in ADL **Bounding Coordinates** now appear in DC **Coverage** and **Producer** becomes **Source**.

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>[DOQQ, Digital orthophoto quarter quadrangle], Soda Lake South SW, California.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creator</strong></td>
<td>United States Geological survey</td>
</tr>
<tr>
<td><strong>Creator</strong></td>
<td>Analytical Surveys, Inc.</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>Aerial photographs digital raster; California</td>
</tr>
<tr>
<td><strong>Subject</strong></td>
<td>DOQQ; quad; Band interleaved by line; BIL</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Digital Orthophoto Quarter Quadrangles (black &amp; white) cover the state of California, each quarter quadrangle covering an area 3.75 degrees by 3.75 degrees. Some quadrangles are still in process as of March 2002</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>U.S. Geological Survey Western Mapping Center</td>
</tr>
<tr>
<td><strong>Contributor</strong></td>
<td>Alexandria Digital Library</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>19970204</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>19940528</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>19971030</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>Image</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>remote-sensing images</td>
</tr>
<tr>
<td><strong>Type</strong></td>
<td>aerial photographs</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>49189680 bytes</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>BIL</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>DOQQ</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>Digital Orthophotographic Quarter Quadrangle</td>
</tr>
</tbody>
</table>
In their study, Zeng and Xiao (2001) found that when data values were involved, mapping or converting became even more complicated. When the target format is more inclusive and has defined elements and sub-elements in greater detail than the source format, the value in a source metadata record may need to be broken down into smaller units (for example, from DC or VRA Core’s elements to MARC records’ subfields). The result is that data values may be lost when converting from a rich structure to a simpler structure. Other complicated situations include converting value strings associated with certain elements that require the use of controlled vocabularies.

The reality is that crosswalks constructed based on the real data conversion might be very different from those based on metadata specifications. Additional instructions and detailed explanations need to be provided for different situations. Unfortunately, most crosswalks are focused only on mappings based on metadata specifications, not on real data conversion results. A recent study on metadata quality (to be released in August 2006) by one of the authors (Zeng) has provided strong evidence about the impact of crosswalks on quality when converting a large amount of real data.

### 2.2 Data Reuse and Integration

It is necessary to revisit the metadata principles, which are simplicity, modularity, reusability, extensibility, and interoperability. Duval et al. consider metadata modularity to be "a key organizing principle for environments characterized by vastly diverse sources of content, styles of content management, and approaches to resource description. It allows designers of metadata schemas to create new assemblies based on established metadata schemas and benefit from observed best practice, rather than reinventing elements anew" [Duval et al., 2002]. In a modular metadata environment, different types of metadata elements (descriptive, administrative, technical, use, and preservation) from different schemas, vocabularies, applications, and other building blocks can be combined in an interoperable way.
If we consider modularity, the components of a metadata record can be regarded as various pieces of a puzzle. They could be put together by combining pieces of metadata sources coming from different processes (by human or machine). They could also be used and reused piece by piece when new records need to be generated by human or machine. For a long time, libraries have been creating rich descriptive metadata. In newer, non-library catalog applications (such as digital images or assets collections), these descriptive metadata components can be reused or combined when other pieces of metadata are generated.

The Metadata Encoding and Transmission Standard (METS) provides a framework for incorporating various components from various sources under one structure and also makes it possible to "glue" the pieces together in a record. METS is a standard for packaging descriptive, administrative, and structural metadata into one XML document for interactions with digital repositories. It thus provides a framework for combining several internal metadata structures with external schemas. It is "a standard that provides a method to encapsulate all the information about an object – whether digital or not" [Tennant, 2001].

![Figure 4. The METS architecture. Source: McCallum (2003)](image)

As illustrated in Figure 4, the descriptive metadata section in a METS record may point to descriptive metadata external to the METS document such as a MARC record in an Online Public Access Catalog (OPAC) or an Encoded Archival Description (EAD) finding aid maintained on a WWW server. Or, it may contain internally embedded descriptive metadata. With the METS schema, it is possible to encode different types of metadata (descriptive, administrative, and structural) for any simple or complex digital library object and to express the complex links between these various types of metadata. It can therefore provide a useful standard for the exchange of digital library objects between collections or repositories [Library of Congress, 2005].

The Resource Description Framework (RDF) of the World Wide Web Consortium (W3C) is another model that "provides a mechanism for integrating multiple metadata schemes" [NISO, 2004] for the description of Web resources. It is a data model that provides a framework within which independent communities can develop vocabularies that suit their specific needs and share vocabularies with other communities. It utilizes the XML namespace to effectively allow RDF statements to reference a particular RDF vocabulary or schema. Each metadata element set constitutes a namespace bound by the rules and conventions determined by its maintenance agency. The metadata schema designer, by declaring the namespace, will be able "to define the context for a particular term, thereby assuring that the term has a unique definition within the bounds of the declared namespace", as stated by Duval et al. (2002). And, by declaring various namespaces within a block of metadata, the elements within that metadata will be able "to be identified as belonging to one or another element set" [Duval et al., 2002]. Thus, multiple namespaces expressed in XML may be defined to allow elements from different schemas to be combined into a single resource description. An RDF record links multiple descriptions, which may have been created at different times for different purposes, to one another. The following
example shows how different metadata schemas (as indicated by namespaces) can be packaged together in an RDF record [Iannella, 1999]:

```xml
<?xml version="1.0" ?>
<RDF xmlns="http://www.w3.org/TR/1999/PR-rdf-syntax-19990105#"
     xmlns:DC = "http://purl.org/DC#"
    <DC:Title> The Future of Metadata </DC:Title>
    <DC:Creator> Jacky Crystal </DC:Creator>
    <DC:Date> 1998-01-01 </DC:Date>
    <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
    <AGLS:Function> Information Management -- Internet </AGLS:Function>
  </Description>
</RDF>
```

There have been discussions about the theoretical possibilities of, and the practical barriers to, implementing RDF in metadata applications. DuCharme (2003) summarized evidence that "Conversations about the value of RDF often veer off on two distracting detours: debates about the architecture and syntax of RDF/XML and debates about the potential value of the Semantic Web". He suggested further: 'If you ignore the latter and let tools like RDFLib shield you from the less appealing details of the former, RDF’s value becomes much more readily apparent, and its increasing success in the metadata community starts to make more sense' [DuCharme, 2003].

### 3. Achieving Interoperability at the Repository Level

When multiple sources are searched through a single search engine, one of the major problems is that the retrieved results are rarely presented in a consistent, systematic, or reliable format. Another serious problem with distributed, independent metadata resources is that each original metadata source provider may have used different metadata schemas and/or applied them differently in the creation of metadata records. A metadata repository, which may be stored in a physical location or may consist of a virtual database in which metadata records have been drawn from separate sources, provides a viable solution to such interoperability problems by maintaining a consistent and reliable means of accessing data. The situation is illustrated in Figure 5, which shows records from multiple sources (which have applied different metadata schemas or application profiles) being integrated into a repository.

![Figure 5. Metadata records are integrated into a repository.](http://www.dlib.org/dlib/june06/zeng/06zeng.html)
Interesting processes or ideas related to ensuring interoperability at the repository level include metadata harvesting, supporting multiple formats, aggregation, crosswalking services, value-based mapping for cross-collection searching, and value-based co-occurrence mapping. One question a repository faces is whether to allow each original metadata source to keep its own format. If not, how would it convert / integrate all metadata records into a standardized format? If so, how would it support cross-collection search?

3.1 A Metadata Repository Based on the Open Archives Initiative (OAI) Protocol

A successful model of a metadata repository is the NSDL (National Science Digital Library) Metadata Repository (MR), a key component of the NSDL architecture. Its metadata repository management system is designed to collect, store, maintain and expose metadata that are contributed by NSDL projects that have used various metadata schemas. In addition to the thousands of item-level metadata records for individual items in the repository, the NSDL MR also holds collection-level metadata for each collection known to the NSDL [Arms et al., 2003]. The Metadata Repository has also facilitated the construction of an automated “ingestion” system, based on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Through this protocol, metadata flows into the Metadata Repository with a minimum of ongoing human intervention [Hillmann et al., 2004].

The goal of the OAI-PMH is to supply and promote an application-independent interoperability framework that can be used by a variety of communities engaged in publishing content on the Web [Lagoze, 2002]. The OAI approach enables searching for Web-accessible resources across different collections, databases, and repositories, based on the capability of metadata sharing, publishing, and archiving. Carpenter (2003) summarized the benefits of the OAI approach, which include:

1. Material can be accessed more widely;
2. Material can be exploited for purposes different from those that originally motivated the creation of the repositories;
3. New and improved services can be constructed because of the possibility of accessing multiple repositories; and
4. There is potential for cost-saving inherent in new models of the scholarly communication process that could be realized through an open archives approach.

3.2 A Metadata Repository Supporting Multiple Formats Without Record Conversion

A different approach that circumvents the need to convert metadata records in an integrated service is taken by the Digital Library for Earth System Education (DLESE). DLESE is a very successful distributed community effort involving educators, students, and scientists working together to build and provide access to high-quality resources consisting of earth system imagery and data sets for teaching and learning about Earth’s systems at all levels. The mechanism that resulted from this effort – the DLESE Collection System (DCS) – is a tool that allows participants to build their own collections of Earth system item-level metadata records, and to develop, manage, search, and share these collections, all without converting every metadata record into a uniform format. The metadata records for each collection are structured according to an XML schema that specifies required and optional metadata (and in some cases legal values) for a particular metadata field. The DCS framework used to manage metadata records has three main components: the schema, fields files, and framework-seAg1–8agings. Each collection may have a different XML schema [DLESE, 2005].

The DLESE Collection System currently supports the DLESE metadata frameworks of ADN (ADEPT/DLESE/NASA) for resources typically used in learning environments (discussed in section 4.5
Metadata Framework in Part I of this article). It also provides information on news and opportunities for events or time-sensitive resources, on collections for groups of metadata records as a whole entity, as well as providing annotations for additional information about resources or information not directly found in a resource. Other XML schema-based metadata frameworks can be supported by configuring the DLESE Collection System to point to the XML schema file. When searching over DLESE metadata formats, both keyword searches and fielded searches are supported across and within collections. When searching over non-DLESE metadata formats or formats that the tool has not been configured to interpret, only keyword searches are supported [DLESE, 2005].

3.3 Aggregation

As described in section 3.1 above, the NSDL Metadata Repository employs an automated "ingestion" system based on OAI-PMH, whereby metadata flows into the Metadata Repository with a minimum of ongoing human intervention. The NSDL, from this perspective, functions essentially as a metadata aggregator [Hillmann et al., 2004].

Four categories of problems that limit the usefulness of metadata have been identified by Dushay and Hillmann (2003) and Hillmann et al. (2004): (1) Missing data: elements not present, (2) Incorrect data: values not conforming to proper usage, (3) Confusing data: embedded html tags, improper separation of multiple elements, etc., and (4) Insufficient data: no indication of controlled vocabularies, formats, etc. It is possible that these problems could be partially eliminated through a process called aggregation in a metadata repository. The notion behind this process is that each metadata record contains a series of statements about a particular resource, and therefore metadata from different sources can be aggregated to build a more complete profile of that resource.

Figure 6 illustrates how several providers might contribute to an augmented metadata record.

![Figure 6. Sources, storage, and redistribution of augmented metadata in the metadata repository. Source: Hillmann et al. (2004)](http://www.dlib.org/dlib/june06/zeng/06zeng.html)

First, the Metadata Repository harvests a simple DC record from Provider A, which contains unqualified title, identifier, creator and type elements. The NSDL Safe Transforms at the repository recognizes the identifier as a valid URI and the type value as a valid DCMIType; therefore, the encoding schemes URI and DCMIType are added elements. Second, an ENC (The Eisenhower National
Clearinghouse) record for the same resource provides two additional pieces of information in the *audience* and *education level* elements; while third, the INFOMINE iVia service provides *subject* information based on three different vocabularies: GEM (Gateway to Educational Materials) Vocabulary, LCSH (*Library of Congress Subject Headings*), and LCC (*Library of Congress Classification*). As a result, the new NSDL normalized/augmented record contains information from sources of Provider A, MR, iVia, and ENC. These enhancements are exposed via OAI-PMH, and the MR can then harvest them. This process of enriching metadata records can be illustrated by the following figure:

![Figure 7. Enriched metadata records.](image)

Thus a new role for "metadata aggregators" is to provide enhanced metadata for other services to reuse, including the following: integrating fragmentary metadata created by automated services, improving metadata in standard ways, and exposing all relevant data in ways that allow consumers to evaluate quality and usefulness [

Hillmann et al., 2004]

### 3.4 Element-based and Value-based Crosswalking Services

While presently crosswalks have paved a way to the relatively effective exchange and sharing of schema and data, there is a further need for effective crosswalks to solve the everyday problem of ensuring consistency in large databases that are built of records from multiple sources. Efforts to establish a crosswalking service at OCLC (Godby et al., 2004) indicated that what are needed are robust systems that can handle validation, enhancement, and multiple character encodings and allow human guidance of the translation process. The OCLC researchers have developed a model that associates three pieces of information: the crosswalk, the source metadata standard, and the target metadata standard. The work proceeded from the hypothesis that "usable crosswalks must have the following characteristics:

- A set of mappings between metadata standards that is endorsed by a stakeholder community.
- A machine-processable encoding.
- A well-defined relationship to source and target metadata standards, which must make reference to particular versions and syntactic encodings" [Godby et al., 2004].

It is worth noting that the second characteristic, "a machine-processable encoding", is not usually assumed in many of the crosswalks that have been developed by various communities or applications.

In the OCLC project, crosswalks are encoded as METS records. (Please refer to section 2.2 above on Data Reuse and Integration for the architecture of METS.) The OCLC researchers also predicted an outcome, i.e., "a repository that collects publicly accessible metadata into a repository that can be
harvested using standard XML protocols and provides tools for creating sample services, such as customizable views of the data” [Godby et al., 2004].

Such a model of element-based crosswalking services could also be used for value-based crosswalking services. Researchers at NSDL [Phipps et al., 2005] have also included a crosswalking service in their sequence of metadata enhancement services. These crosswalking services are a type of metadata augmentation operation that generates new fielded metadata values that are based on crosswalking from a source (schema or vocabulary) to a target (schema or vocabulary). The operation can be performed on either controlled or uncontrolled vocabulary value strings associated with specific elements. Both element-based and value-based crosswalking services improve the reusability of metadata in a variety of knowledge domains. They help to present attributes of resources in terms familiar to particular groups of users in the filtering of their search results [Phipps et al., 2005].

3.5 Value-Based Mapping for Cross-Database Searching

The Multilingual Access to Subjects (MACS) project illustrates another value-based mapping approach to achieving interoperability among existing metadata databases. MACS is a European project designed to allow users to search across library cataloging databases of partner libraries in different languages, which currently include English, French, and German. Specifically, the project aims to provide multilingual subject access to library catalogs by establishing equivalence links among three lists of subject headings: SWD/RSWK (Schlagwortnormdatei / Regeln für den Schlagwortkatalog) for German, Rameau (Répertoire d’autorité-matière encyclopédique et alphabétique unifié) for French, and LCSH (Library of Congress Subject Headings) for English.

Instead of integrating records or mapping the data of every field, MACS chose to only map the values in the subject field. The method employed for mapping is to compare subject headings in three monolingual lists and check the consistency of bibliographic records retrieved with these headings. The links were analyzed on three levels: terminological level (subject heading), semantic level (authority record), and syntactic level (application) [Freyre & Naudi, 2003]. Using any of the three languages involved in the MACS project, a user can search and find bibliographic information of items hosted at the four national libraries of France, Germany, Switzerland, and the United Kingdom.

For example, when a search for "all terrain cycling" is initiated, the equivalents of the English heading in SWD (German) and RAMEAU (French) will also be displayed. When there is no exact mapping, a combination of two or more headings with the AND or OR Boolean operator will be provided. See examples of "All terrain cycling" and "All terrain cycling – Training" in Figure 8:
When the searcher chooses "all terrain cycling" search results from all participating national library catalogs are displayed (Figure 9):

When an item on the search result list is selected, a full record from that particular catalog will be displayed (Figure 10). Note particularly the values in the "Subjects" field of each record. They include subject headings in different languages or notations. Without a mapping process among the three lists of subject headings, cross-database searches based on these strings would be impossible.
3.6 Value-Based Co-Occurrence Mapping

With regard to searching, co-occurrence mapping (Figure 11) is similar to what is done in the MACS project discussed above. However, this approach uses the values already present in the subject fields and regards the subject terms in different languages registered in the subject fields of the same record as equivalent. When a metadata record includes terms from multiple controlled vocabularies, the co-occurrence of subject terms enables an automatic, loose mapping between vocabularies. As a group, these loosely-mapped terms can answer a particular search query or a group of questions.

![Co-occurrence mapping diagram](source)

Existing metadata standards and best practice guides have provided an opportunity to use the co-occurrence mapping method. For instance, the art- and image-related metadata standard VRA Core Categories version 3.0 requires the use of the Art and Architecture Thesaurus (AAT) for the Type, Material, and Style/Period elements; and, for the Culture and Subject elements, recommends the use of AAT, LCSH, Thesaurus of Graphic Materials (TGM), ICONCLASS (an international classification system for iconographic research and the documentation of images), and Sears Subject Headings (Figure 12).

<table>
<thead>
<tr>
<th>Type</th>
<th>Data Values (controlled):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>recommend AAT</td>
</tr>
<tr>
<td>Material</td>
<td>AAT</td>
</tr>
<tr>
<td>Technique</td>
<td>AAT</td>
</tr>
<tr>
<td>Style/Period</td>
<td>Data Values (controlled): recommend AAT</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Culture</td>
<td>Data Values: recommend AAT, LCSH</td>
</tr>
<tr>
<td>Subject</td>
<td>Data Values (controlled): recommend AAT, TGM, ICONVCLASS, Sears Subject Headings</td>
</tr>
</tbody>
</table>

Figure 12. Controlled vocabularies required or recommended for use in VRA Core metadata records, compiled based on VRA Core 3.0.

Another example of co-occurrence mapping source is the Gazetteer Standard Report of the Alexandria Digital Library. Under Feature Class, terms from two controlled vocabularies are recorded.

![Gazetteer Standard Report](image)

Figure 13. An ADL record showing assigned values from two vocabulary sources.

Source: ADL Digital Gazetteer

Metadata records often include both controlled terms and uncontrolled keywords. Mapped subject terms can be used as access points that lead to full metadata records. With the progress of the aggregation services described in section 3.3 Aggregation above, more "fielded-in" value strings associated with multiple sources may be integrated to enrich metadata records. As more co-occurrence-types of application become widely used, loosely-mapped values can become very useful in searching.

4. Conclusion

As early as 1999, Ercegovac observed: "In today's digital libraries environment, in which individual collections of massive heterogeneous objects need to be unified and linked in a single resource, we are witnessing both the growth of different metadata and the attempts to reconcile the common attributes in the existing overlapping standards" [Ercegovac, 1999]. Depending on the point at which interoperability efforts are initiated – at the beginning of forming a collection or repository before metadata records have been created or when participants already have existing records – different
approaches and methods have been tested and employed at different levels of operation: schema level, record level, and repository level. The results of interoperability efforts may be observed as element-based and value-based approaches. Currently, mapping metadata schemas for the sake of interoperability still requires enormous intellectual effort even with all the assistance computer technology can provide. Nevertheless, it is clear that technology has provided many possibilities for facilitating and enhancing metadata semantic interoperability efforts. What are still needed are the intellectual preparations to support the goal of interoperability and the practical tools to carry out such activities. If the information community is to provide optimal access to all the information available across digital libraries and repositories, information professionals must give high priority to the task of creating and maintaining the highest feasible level of interoperability among extant and new information services.

Acknowledgements

The authors gratefully thank the help and support of Dr. Theodora Hodges (Berkeley, CA), Katy Ginger (DLESE), Dr. Athena Salaba (Kent State University), and Samantha Nicholson (Kent State University).

Sources and References

Sources

(Metadata schemas, application profiles, and registries mentioned in Part II of the article)

ADL Digital Gazetteer
<http://webclient.alexandria.ucsb.edu/client/gaz/adl/index.jsp>

ADN (ADEPT/DLESE/NASA) Framework
<http://www.dlese.org/Metadata/adn-item/>
Created by DLESE (Digital Library for Earth System Education).
The current framework used in the DLESE Discovery System. It contains information used by the Earth system education community for resource discovery in learning environments (e.g., classroom activities, curriculum, virtual field trips, etc.).

Encoded Archival Description (EAD)
<http://lcweb.loc.gov/ead/>
Maintained by the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists.
A standard for encoding archival finding aids using Extensible Markup Language (XML).

MACS: Multilingual Access to Subjects
<https://ilmacs.uvt.nl/pub/>
A product of the Swiss National Library (SNL), the Bibliothèque nationale de France (BnF), The British Library (BL) and Die Deutsche Bibliothek (DDB).
A subject interoperability project that enables users to simultaneously search the catalogues of the project’s partner libraries in the language of their choice (English, French, German). This multilingual search is made possible through the equivalence links created among the three indexing languages used in these libraries: SWD (for German), RAMEAU (for French) and LCSH (for English).

MARC (MAchine-Readable Cataloging) Formats
<http://lcweb.loc.gov/marc/>
Developed and maintained by the Library of Congress Network Development and MARC Standards...
Office.
A set of standards for encoding cataloging data in online library catalogs. MARC provides the mechanism by which computers exchange, use, and interpret bibliographic information. The LC MARC became known as USMARC in the 1980s, and in 1999, it was merged with CANMARC (Canadian MARC) to become MARC 21.

METS (Metadata Encoding and Transmission Standard)
<http://www.loc.gov/mets>
A standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language.

MODS (Metadata Object Description Schema)
<http://www.loc.gov/standards/mods/>
Developed and maintained by the Library of Congress Network Development and MARC Standards Office.
A subset of MARC fields using language-based tags rather than numeric ones, in some cases regrouping elements from the MARC 21 bibliographic format. MODS is expressed using the XML.

VRA (Visual Resources Association) Core Categories, 3.0
<http://www.vraweb.org/vracore3.htm>
Developed and maintained by Visual Resources Association Data Standards Committee.
A metadata element set for creating records describing works of visual culture as well as the images that document them.

References


McIlwaine (Ed), *Subject retrieval in a networked environment, proceedings of the IFLA satellite meeting held in Dublin, Ohio, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC* (pp.3-10). München: K.G. Saur.


