Information Services & Use 25 (2005) 35–45
IOS Press

Metadata for electronic information resources: From variety to interoperability

Gail Hodge
Information International Associates, Inc., 312 Walnut Place, Havertown, PA 19083, USA
E-mail: ghodge@iiaweb.com

Abstract. Metadata serves several purposes. It supports resource discovery, locates the actual digital resource by inclusion of a digital identifier, organizes electronic resources bringing similar resources together and distinguishing dissimilar resources, provides administrative information for controlling the digital library, and provides technical, preservation and rights management information needed to support immediate and long-term permanent access. There are a variety of metadata schemes that serve different purposes for different object types, subjects and audiences. With disparate metadata schemes, ensuring that information collected in a specific scheme by one organization for a particular purpose can be exchanged, transferred or used by another organization for a different purpose becomes an issue. Metadata frameworks, crosswalks, and registries are ways to achieve interoperability. Controlled terminologies add more precise meaning to metadata. The integration of controlled terminologies and metadata schemes is key to the development of the Semantic Web.

1. The purpose of metadata

Similar to traditional cataloging and indexing, metadata performs three main functions. It facilitates discovery of relevant information by describing aspects of the original electronic resource in which the designated user community may be interested. Metadata, such as titles, subject terms and abstracts or descriptions, are particularly important for electronic resources, such as datasets or photographs, that have little if any text content on which current text-based Web searching can be performed. Metadata can describe the resource at any level of aggregation – a single resource; a part of a larger resource, for example, a photograph in an article; or a collection of resources, such as a digital library. The level at which metadata is applied depends on the type of data and the anticipated access needs. Datasets are generally cataloged at the file or collection level. Electronic journal articles may be cataloged individually, sometimes with no concern for metadata at the issue or journal title levels. Generally, the metadata for Web sites is applied to one or more pages that make up a cohesive resource with informational value.

Once a resource has been discovered via the metadata, the resource must be located. Metadata supports the location of the actual digital resource on the network. Most metadata schemes include an element that is defined as the unique identifier needed to locate the resource. In practice, most metadata schemes continue to use the URL (Uniform Resource Locator) as the unique identifier. However, there are other schemes that may provide more persistence by resolving the URL to a permanent identifier. These schemes include the Persistent URL, the Handle (also implemented as the Digital Object Identifier) and the Archive Resource Key (ARK).

In addition to the discovery of specific resources, metadata brings similar resources together and distinguishes dissimilar resources. As the number of Web-based resources grows exponentially, aggregate

0167-5265/05/$17.00 © 2005 – IOS Press and the authors. All rights reserved

sites, portals, or subject gateways are increasingly useful in organizing resources based on audience or topic. Originally, these resources were built as static Web pages with the names and locations of the resources "hard coded" in the HTML. However, it is more efficient and increasingly more common to build these pages dynamically from metadata stored in databases. Content management systems support the development of such portals by managing individual digital objects and the associated metadata. Metadata information is also matched against user profiles to create customized MyLibrary or MyPortal Web sites.

Administrative metadata is used by the digital library or data center to manage the digital object and its metadata. The elements included as administrative metadata depend in part on the workflow for the creation, capture and long-term use of the digital object that is being archived and preserved. They include control elements such as the date created, the date captured, and the date last migrated.

Technical metadata is the overall term used for metadata elements that describe the computer hardware and software needed to reproduce the digital object, including file formats such as pdf and video formats such as mpeg that are used by viewers or browsers that allow the user to access the object. Technical metadata elements are often considered part of the preservation metadata set because it is critical to rendering the digital object in new technical environments in the future or when using emulators of obsolete technologies. Technical metadata schemes are often quite large and detailed, since they are usually intended for use by technicians or for computer to computer communication.

In addition to these common functions, metadata elements can cut across functions, which may result in specific element sets. Digital rights elements, including security classifications or distribution limitations, indicate who owns the object and what rights various groups have to use or reuse that object. There are several schemes that have been developed particularly in the music and learning objects communities. In systems, the rights management elements must be matched against profiles of the user (following proper authentication) in order to ensure that the material is being properly distributed and in some cases the proper payments are being made to the rights holders. The variety of systems, the potential economic impacts, and the variety of materials requiring rights management have led to the concept of a Digital Rights Expression Language (DREL) that is of broad applicability and that can be used by a variety of automated systems in e-commerce. IEEE, MPEG21 and others have been working on rights elements and expression languages. In addition, preservation metadata records the provenance of an object and the long-term access. The current work in this area is discussed in the paper on Preservation and Permanent Access.

2. The variety of metadata schemes

Metadata schemes (also called "schema") have been developed and defined by a variety of communities, for different purposes, and for different types of electronic resources. Arms [2] points out that there are good reasons why different metadata schemes have been developed for different formats and for different subject matter. Images without text require different types and levels of metadata than a digital document that can be searched using a full-text search engine. Arms goes on to point out that in some cases images can only be distinguished through the use of thumbnails (essentially metadata in image form) rather than by textual metadata. Software, datasets, audio recording and Web sites each call for different metadata practices. Even though practitioners may dream of a single unifying metadata scheme, this is unlikely and would not serve the users well. Therefore, we are left to deal with the variety.

This section describes some common metadata schemes. In addition, some lesser known schemes have been selected to show the range of electronic resources and purposes for which schemes have been developed. While the focus here is on electronic library resources, it should be noted that many other metadata schemes have been developed in support of e-commerce and electronic data exchange. Additional metadata schemes can be located through the Metadata Schema Registry (metadata.net), the UKOLN Metadata web page (http://www.ukoln.ac.uk/metadata/), and MetaMap (http://mapageweb.umontreal.ca/turner/meta/english/index.html).

2.1. Dublin core

The Dublin Core is perhaps the most well known metadata element set (purl.oclc.org/metadata/dublin_core/). The original objective of the Dublin Core was to define a set of elements that could be used by authors to describe their own Web resources. Fifteen relevant elements and simple rules were defined so that non-catalogers could provide basic information for resource discovery. In part because of its simplicity, the Dublin Core has been used with other types of materials and for applications demanding increased complexity. The desire to be able to specify more detail resulted in the development of qualified Dublin Core, in which qualifiers are used to refine the meaning of an element or to specify the domain values or rules for representing an element. The element "Date", for example, can be used with the qualifier "created" to narrow the meaning of the element to the date the resource was created. A qualifier can also be used in the element "Date" to specify the ISO 8601 standard as the required format for representing date. There are perhaps thousands of projects worldwide that use the Dublin Core for cataloging or to collect data from the Internet. Dublin Core is also the minimum shareable metadata set in the Open Archive Initiative-Protocol for Metadata Harvesting. While other sets can be used based on mutual agreement between the data provider and the harvester, every OAI-compliant provider must provide unqualified Dublin Core metadata.

2.2. Metadata Object Description Schema

The Metadata Object Description Schema (MODS) is a schema for bibliographic elements to support the interoperability of MARC records (especially MARC21) with other bibliographic metadata schemes (www.loc.gov/standards/mods/). It was developed by the Library of Congress for a variety of applications, particularly those related to library catalogs. MODS includes a subset of MARC fields, but it uses language-based tags rather than the traditional numeric tags used by MARC. MODS includes 19 top level elements which in some cases regroup the MARC elements. MODS is expressed in XML and is often used in conjunction with the Metadata Encoding and Transmission Standard (METS) (see Section 4.1) as a transfer format. MODS 3.0 was released in March 2004.

2.3. Global Information Locator Service

The Global Information Locator Service (GILS) was developed by the U.S. government as a tool for enhancing public access to government information (www.gils.net). International implementers of GILS include Australia, Germany, Singapore, and Hong Kong. GILS is also widely used with spatial and environmental clearinghouses implemented by countries and international organizations. GILS specifies a profile of the Z39.50 protocol for distributed search and retrieval which is a common standard used in online library catalogs. It specifies the attributes (or the elements) that must be able to be searched in order for a system to be GILS compliant. However, organizations have specifically defined GILS elements for their own communities. Since the purpose of GILS is to act as a locator service, GILS elements emphasize availability and distribution rather than description, including elements such as the name and address of the distributor and information on ordering. GILS records were intended to describe aggregates or collections such as catalogs, publishing services and databases, but some organizations use GILS at the individual item (journal article or technical report) level.

2.4. Encoded Archival Description

The Encoded Archival Description (EAD) is also used to describe collections (lcweb.loc.gov/ead/). The EAD was developed as a way of marking up the data contained in finding aids, so that it could be searched and displayed online. The EAD is particularly popular in academic libraries with large special collections and in national archives. The EAD begins with a header section that describes the finding aid itself (for example, who created it), and then it describes the whole collection or record series, with successively more detailed information about the contents of the collection.

2.5. ONIX International

ONIX (Online Information Exchange) International (www.editeur.org/) is a metadata scheme developed by a number of book industry trade groups in the United States and Europe to support e-commerce. ONIX has elements for basic bibliographic, trade, evaluation and promotional information for books and e-books. This metadata standard is particularly valuable for Internet-based booksellers, such as Amazon.com. It supports the display of such online features as pictures of book covers, book review "snippets", and links to author biographies. Although initially focused on books, ONIX has been adapted to serial publications.

2.6. ISO Standard for Digital Geospatial Metadata

Metadata schemes also exist for a variety of other format and object types. Metadata for datasets is particularly important in disciplines, such as genetics or demographics, where numeric and statistical data are primary resources. One of the most well developed element sets and content standards for data is the ISO Standard for Digital Geospatial Metadata (ISO 19115:2003). Geospatial datasets link data for a specific purpose to the latitude and longitude coordinates on the earth. These datasets are used in a wide variety of applications, including soil and land use studies, climatology and global change monitoring, remote sensing, and demographic and social science research. Many national and local governments use the content standard, and it has become deeply embedded in Geospatial Information Systems (GIS) forming the basis for the work of the Open GIS Consortium to provide for better interoperability among GIS applications.

2.7. Technical metadata for digital still images

In the area of technical metadata, the National Information Standards Organization developed a data dictionary of technical elements for digital still images (www.niso.org/committees/committee.au.html). A draft was released for comment in February 2001. NISO realized that the focus of most cultural institutions was on descriptive metadata, without any emphasis on the technical aspects of digital images that would be needed to adequately store and preserve them. The purpose of the standard is to facilitate the "development of applications to validate, manage, migrate and process images of enduring value". The emphasis is not only on current use of still images, but on the long-term provenance, preservation, and assessment for use and re-use.

3. Metadata extensions and profiles

In addition to metadata schemes that are individually specified through standards or common practice, specific implementations or the needs of a certain community can result in modifications to a metadata scheme. Since it is often difficult to anticipate the ways in which a scheme might be used, schemes that can easily be modified are preferred over those that are more restrictive. An extension is the addition of elements to an already developed scheme to support the description of a particular resource type, to handle material on a particular subject, or to address the needs of a particular user community. Profiles are subsets of a larger scheme that are implemented by a particular user community. Extensions generally increase the number of elements that can be used; profiles constrain the number of elements, refine or narrow the definitions of certain elements, or specify the rules for completing the content of certain elements.

In practice, many applications use both extensions and profiles of base metadata schemes. The metadata scheme for the U.S. Department of Education's Gateway to Educational Materials (GEM) Project is based on the Dublin Core (www.geminfo.org/Workbench/Metadata/index.html). However, GEM limits the elements to be used (for example, Contributor is not used). It also extends the Dublin Core element set by adding elements that are important to the educational community when describing and using educational resources. These fields include audience (teacher versus student), grade level, and relevant educational standards. Similarly, the VRA Core Category scheme (www.vraweb.org/vracore3.htm), a profile and extension of the Dublin Core, consists of 17 optional metadata elements needed to describe visual resources. Managers of visual resource collections hope that use of the VRA Core Categories will allow them to better describe materials in their own collections and to share descriptions across collections.

4. Metadata interoperability

The preceding section describes only a small number of the existing metadata schemes. With so many metadata schemes, how will chaos be avoided? How can we ensure that systems that use different metadata schemes will be interoperable, in other words that information collected by one organization for a particular purpose can be searched, exchanged, transferred, used and understood by another organization for a different purpose. Practitioners cite metadata frameworks, crosswalks, and metadata registries as tools to support this interoperability.

4.1. Metadata frameworks

A metadata framework is a reference model that provides a high-level, conceptual structure into which other metadata schemes can be placed. It also gives designers and developers a consistent, cross cutting terminology around which to discuss metadata for a particular discipline.

An example of a metadata framework is the Metadata Encoding and Transmission Standard (METS) (www.loc.gov/standards/mets/). METS was developed by the Digital Library Federation and the Library of Congress for the management of digital library objects. METS uses a framework which defines metadata as descriptive, administrative or structural. The most significant contribution of METS is its emphasis on structural metadata. METS also adds a fourth component, a list of the files in the digital library object. The structural component of the METS scheme indicates how these files work together to form the digital library object. This structure information supports the management of the object by a digital library, and it facilitates the exchange of these objects among digital libraries. METS provides an XML DTD that can point to metadata in other schemes by declaring the scheme that is being used. For example within the METS framework, Dublin Core elements could be used to describe a digital still image for resource discovery, and the technical elements from NISO's Draft Standard for Digital Still Images could be used to document the structural aspects of the image.

The Interoperability of Data in E-Commerce Systems (<indecs>) Framework (www.indecs.org/) is an international collaborative effort originally supported by the European Commission. It has developed a metadata framework, or a reference model, that supports the sharing of information about intellectual property rights in electronic commerce. In the basic model, people make "stuff", people use "stuff", and people make deals about "stuff". Rather than develop a new metadata standard, <indecs> provides a framework for the various existing schemes to interact. For example, transactions related to music, journal articles or books could interchange information with one another. This framework has also been discussed as a way to allow the various groups (publishers, libraries and users) involved in access to electronic journal subscriptions to work within a consistent framework for interchange while maintaining the original metadata for their local applications. In a significant move, the <indecs> framework has been adopted by the MP3 standards group working on standards for multimedia including intellectual property. In the MP3 context, the <indecs> framework is known as Contecs:DD.

4.2. Metadata crosswalks

In addition to metadata frameworks, crosswalks are often developed to map the elements, semantics and syntax from one metadata scheme to those of another. A crosswalk allows metadata created by one community to be used by another community with a different metadata standard. The degree to which crosswalks are successful depends on the similarity of the two schemes. The mapping of schemes with fewer elements (less granular or atomic) to those with more elements (more granular or atomic) is problematic. Despite similarity at the semantic level, the crosswalk can be difficult if the content rules differ from the original scheme to the target scheme even if the definitions of the elements are similar.

Crosswalks are key to the Open Archive Initiative (OAI) (www.openarchives.org/). To achieve OAI-compliance, an archive exposes the OAI metadata set by crosswalking its native metadata format to the simple Dublin Core. This file is exposed and then harvested into a central repository.

Many common schemes such as MARC and Dublin Core have been mapped and made available on the Web [3]. A graphic representation of various mappings, called MetaMap, has been developed by students at the University of Montreal (http://mapageweb.umontreal.ca/turner/meta/english/index.html).

4.3. Metadata registries

Registries are another tool for exchanging metadata. They provide information about the definition, origin, source, and location of the scheme, usage profile, element set, and/or authority files for element values. A registry maps one scheme to another so that both humans and computers can understand how they might integrate. Registries can also document rules for transforming content for an element in one system to the content required for an equivalent element in another. The DESIRE (Development of a European Service for Information on Research and Education) Project (desire.ukoln.ac.uk/registry/) funded by the European Commission has developed a prototype of such a registry based on the ISO standard for defining data elements (ISO/IEC 11179). The Dublin Core Metadata Initiative has also developed a registry for the Dublin Core elements (http://dublincore.org/dcregistry/). The JISC IE Metadata Schema Registry is focused on interoperability of metadata within the education community [6].

Registries are particularly useful in specific disciplines or industries such as health care, aeronautics, or environmental science, where they can be used to make the contents of resources more easily integrated. A good example is the U.S. Environmental Protection Agency's Environmental Data Registry which provides information about thousands of data elements used in current and legacy EPA databases. The EDR metadata registry provides an integrating resource for legacy data, acts as a look-up tool for designers of new databases, and documents each data element (www.epa.gov/edr/). The European Environment Agency has developed a similar registry which is available as open source software on the Web.

5. Interoperability through controlled terminology

The use of controlled terminology is becoming increasingly important as a tool for metadata creation and access. This is particularly true as more information managers realize the problems that arise from free text searching or the use of uncontrolled keywords in an unorganized information space such as the Web. Controlled terminology can be used with metadata to restrict value domains and to categorize metadata elements in a metadata registry.

5.1. Limiting metadata domain values

The most prevalent use of controlled terminology is to limit the values that can be entered into a particular metadata element. While most metadata schemes do not dictate the use of a particular controlled terminology when entering the contents of elements that describe what the resource is about, use of controlled vocabularies is encouraged, and many metadata schemes allow controlled terminologies to be defined or referenced within the syntax of the metadata. For example, the Dublin Core description recommends the use of controlled values for fields where they are appropriate (for example, controlled terms from a thesaurus for the Subject field or the use of the ISO language names and abbreviations for the Language field). The content rules are determined by the particular implementation, but the adoption of profiles that define domain-specific rules is encouraged.

An individual project may specify the controlled terminology to be used. For example, the National Biological Information Infrastructure, which uses the biological profile for the FGDC Geospatial Content Standard (www.fgdc.gov/metadata/contstan.html), specifies the controlled terminology to be used. Cambridge Scientific Abstracts, as a partner of the National Biological Information Infrastructure (NBII), has developed a Biocomplexity Thesaurus (http://thesaurus.nbii.gov). The terms in the thesaurus must be used in the NBII metadata to tag electronic resources across the NBII subject and geographic nodes. The thesaurus also is used to select terms for the more traditional bibliographic indexing in CSA's Biocomplexity database, which is searchable through the NBII Web site. The NBII portal will use the terms to create collections of information based on a user's personal preferences. The NBII's biological profile of the FGDC Metadata Content Standard also specifies the use of the Integrated Taxonomic Information System (ITIS) (www.itis.usda.gov) as the authority file for completing the biological taxonomic classification elements within the metadata record.

Similarly, controlled term lists have been developed by many of the U.S. states using the Global Information Locator Service to describe government resources. These include terms that describe the major services and products provided by states to their citizens, to state employees, or to other governments, whether state, local or national.

A variety of controlled terminology systems are being used for indexing electronic resources. These include traditional library schemes, such as the Library of Congress Subject Headings and the Dewey Decimal Classification; specific domain-oriented thesauri or classification schemes such as the Medical Subject Headings (MeSH); and locally created lists of frequently used or important terms. The tools that are required to use existing controlled terminology schemes in the Internet environment is a major research area for OCLC.

5.2. Categorizing metadata elements in registries

The second use of controlled terminologies is to classify the metadata elements when the scheme is documented in a metadata registry. Categorizing or classifying elements is recommended by the ISO 11179 standard for describing data elements. In addition to definitions and documented value domains, applying a controlled terminology scheme to the documentation of metadata elements in a registry can support the identification and appropriate re-use of elements, particularly if the registry reflects data elements from legacy systems.

The National Cancer Institute (NCI) has successfully used the NCI Thesaurus to categorize the metadata elements in its 11179-compliant registry [10]. By doing so, the use and re-use of metadata elements in the registry is enhanced. Users of the registry (including computer programs) can use the terms from the NCI Thesaurus that have been assigned to metadata concepts to locate specific metadata elements or to disambiguate metadata elements with similar names but different meanings. By applying a rigorous workflow and a close relationship between the NCI staff who maintain the thesaurus and those documenting datasets, NCI has developed a very robust metadata registry that supports the integration of diverse datasets across the enterprise.

5.3. Metadata and the Semantic Web

Metadata is seen as key to the development of the Semantic Web, a Web that is more "organized" and able to convey meaning. Metadata creation is increasingly integrated with the software applications people routinely use to do their jobs. For example, Adobe's Extensible Metadata Platform (XMP) is an open standard that supports more extensive and flexible embedding of metadata in PDF documents [1]. XMP-encoded metadata can be captured and modified during the document's workflow and, subsequently, used by various applications, such as content management systems. However, metadata is only a first step in moving toward the Semantic Web, because simply tagging an object cannot convey meaning. Meaning requires shared understanding within a community or across communities.

Terminology and metadata communities believe that it is the combination of metadata and controlled terminologies that will allow the Web to better convey meaning. Heery and Wagner [7] and Fitzwater and Spencer [5] believe that metadata registries are key to the development of the Semantic Web. Providing metadata registries that will support the Semantic Web requires wider availability of terminology resources on the Web and the integration of knowledge organization structures, such as ontologies that convey concepts, terms, and their relationships, with metadata registries.

One approach to making knowledge organization structures more generally available as Internet tools is the development of terminology Web services. A Web service uses specific standardized protocols to create modules that can be used and re-used in a variety of applications over the Web. Web services and other networked applications of controlled terminologies help to support the development of a Semantic Web, a major activity of the World Wide Web Consortium. The goal of this initiative is to provide the

Web with an "understanding" of concepts in order to result in better machine to machine processing of text and provision of services. The basis for the Semantic Web is a knowledge representation that is much richer than that reflected in standard thesauri or classification schemes. These ontologies, semantic networks or topic maps encode more specific relationships between concepts. For example, instead of simply labeling "leg" as a narrower term to body part, the relationship would be specifically identified as "whole-part". In a knowledge organization system concerned with the environment and human health, the relationship between mosquito and West Nile Virus might be "carrier (or vector)-disease". The same term, mosquito would have another relationship to the term Insect as its higher level biological taxonomy relationship. Having more explicit relationships allows for better disambiguation of results and the building of rules and assumptions into information retrieval systems, resulting in more robust retrieval.

Many examples of terminology Web services are under development. For example, the U.S. National Agricultural Library has developed its Web-enabled Agricultural Thesaurus as a Web service [4]. Functionality includes locating a term, navigating the thesaurus, and selecting the term and various types of related terms. This Web service can be called by any other system wanting to perform these functions against the Agricultural Thesaurus as the basis for searching or browsing its content from its own applications. Similar implementations have been developed by others in the environmental and agricultural domain, including the European Environment Agency, the U.S. Environmental Protection Agency, and the Food and Agriculture Organization. Through the Ecoterm Working Group of the Ecoinformatics Initiative work is underway to ensure some commonality among these Web services.

In a similar initiative, Zthes, a Z39.50 profile for thesaurus navigation (www.loc.gov/z3950/agency/profiles/zthes-04.html) has been developed. The profile provides a high level, abstract representation for navigating a thesaurus. In addition to providing thesaurus search capabilities within the realm of Z39.50 (which includes the GILS initiatives and many of the initiatives that use the FGDC content standard) an appendix to the profile provides an XML DTD for thesauri that could be used by other protocols.

Both Web services and Zthes require terminology system owners to prepare their terminology systems for this type of access. This has proven to be a major impediment to the use of existing terminology resources, many of which have been developed over decades. However, starting from scratch and developing richer structures such as ontologies is also a daunting task. Many owners of terminology systems have made their systems searchable/browsable on the Web and systems are being used to limit domain values. However, these initiatives do not significantly improve access to terminologies across the breadth of Internet resources or when the user is searching outside his or her area of expertise.

However, several groups involving both metadata and terminology experts are addressing this challenge. These groups are approaching the problem from various angles on both an international and project-level scale. There is significant overlap in the members of these groups which should lead to cross-fertilization of ideas, the development of appropriate practices, and eventual standards development.

A group called the Networked Knowledge Organization Systems/Services (NKOS), an ad hoc group from public and private sector organizations in more than ten countries, has been discussing the issues related to the challenge of providing generally applicable knowledge organization services (KOS) via the Internet (nkos.slis.kent.edu/). The group defines KOSs to include authority files, thesauri, gazetteers, ontologies, topic maps, taxonomies, subject headings, and any other types of schemes intended to organize digital objects. NKOS has been discussing protocols for the use of KOSs via the Internet, and has developed a set of metadata elements to describe KOSs and their characteristics and behavior. This metadata could be used as part of a registry of KOSs or as metatag information embedded in header information for a Web-based KOS.

The SWAD-E group in Europe has developed SKOS Core 1.0, an RDF schema for thesauri and other similar knowledge organization systems [9]. It is intended to serve as the basis for moving traditional knowledge organization systems into formats that are more appropriate for the Semantic Web even though the rich semantic relationships may need to be enhanced manually or through additional programming.

The eXtended Metadata Registries (XMDR) Project (www.xmdr.org), sponsored by the U.S. Department of Defense, the U.S. Environmental Protection Agency and others, is developing a prototype to help determine how ISO 11179 and other standards must be modified in order to provide more semantic capabilities in 11179 metadata registries [8]. The current standard can accommodate only very simple semantics (e.g., the classification and categorization described above). However, in order to make registries more useful, particularly in the Semantic Web environment, more complex knowledge organization systems, such as thesauri and ontologies are being included in the testbed for the prototype which focuses on human health and the environment.

6. Conclusions

Metadata schemes have been developed for a variety of purposes – resource discovery, location, collection organization and management, administration, rights management, technical reproducibility and preservation. However, because the needs of resource types and user communities differ, a variety of schemes have been developed, along with specific extensions and profiles. Metadata frameworks, crosswalks and registries can help to bridge the various metadata communities. Metadata standards and interoperability remain key issues. The use of controlled terminologies can improve the precision of domain values within the implementation of a metadata scheme. Controlled terminologies can also support the documentation of metadata elements in registries, facilitating the discovery and re-use of metadata elements, when communities need to develop new schemes. Ultimately, the integration of controlled terminology systems and metadata systems will provide the infrastructure needed by the Semantic Web to convey meaning based on shared understanding.

References

[1] Adobe, Extensible Metadata Platform (2005). (Online.) Available: http://www.adobe.com/products/xmp/main.html (13 May 2005).

[2] C. Arms and W. Arms, Mixed content and mixed metadata: Information discovery in a messy world, in: *Metadata in Practice*, D. Hillmann and E. Westbrooks, eds, American Library Association, Chicago, 2004, pp. 223–237.

[3] M. Day, Metadata: Mapping between metadata formats (2001). (Online.) Available: www.ukoln.ac.uk/metadata/interoperability/ (13 May 2005).

[4] L. Finch and T. Lynch, National Agricultural Library Agriculture Thesaurus: Thesaurus as a Web Service. Presented at Knowledge Organization Systems: Current and Future Uses Among CENDI Agencies, September 16, 2004. (Online.) Available: http://cendi.dtic.mil/presentations/KOS_NAL_Finch.ppt (13 May 2005).

[5] L. Fitzwater and L. Spencer, EDR and management of meaning. Presented at the EPA System of Registries Conference, May 5, 2005.

[6] R. Heery, JISC IE metadata schema registry. Presented at the OpenForum 2005, Berlin, Germany, April 13, 2005. (Online.) Available: http://www.berlinopenforum.de./download.html (13 May 2005).

[7] Heery, R. and H. Wagner, A metadata registry for the Semantic Web, *D-Lib Magazine* **8**(5) (2002). (Online.) Available: http://www.dlib.org/dlib/may02/wagner/05wagner.html (13 May 2005).

[8] K. Keck, XMDR Project to extend MDR for registration of complex concept structures. Presented at the OpenForum 2005, Berlin, Germany, April 12, 2005. (Online.) Available: http://www.berlinopenforum.de./download.html (13 May 2005).

[9] A. Miles and D. Brickley, SKOS Core guide: Editor's draft (2005). (Online.) Available: http://www.w3.org/2004/02/skos/core/guide/ (13 May 2005).

[10] D. Warzel, National Cancer Institute's Cancer Data Standards Repository (caDSR). Presented at the OpenForum 2005, Berlin, Germany, April 14, 2005. (Online.) Available: http://www.berlinopenforum.de/download.html (13 May 2005).

Copyright of Information Services & Use is the property of IOS Press. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.