Metadata Elements

Abstract

Chapter 3 discusses metadata elements, the building blocks of any metadata scheme. Metadata elements have been defined using Semantic Web standards, which form a universal pool from which anyone creating metadata can make choices. This chapter describes a range of sources for metadata elements with explanations on their use.

The building blocks for any metadata scheme are the elements that will be used to define the information that is provided. These are often called *data elements*, although in Semantic Web terminology they are referred to as *properties*. I will call them elements when discussing them generally in this document because that is the term most familiar to the metadata developers and creators in the library world. I will use the terms *properties* and *classes* in the Semantic Web sense when describing particular element sets since that is what they will be called in the documentation where they are defined in RDF or OWL. In addition, groups of Semantic Web elements that have been defined are called either *vocabularies* or *ontologies*, and these terms are used imprecisely and interchangeably. To make things worse, the term *vocabularies* is also applied to controlled lists of terms that are used as data, not as elements. As much as I dislike the term *ontology* for metadata element sets (an -ology should be a study of something, and *ontology* in its original definition in philosophy means the study of reality), I will use it here for metadata term sets so that I can reserve the term *vocabulary* for the controlled lists (see table 3.1).

Elements can be as simple or complex as the metadata task warrants. There can be one single data element for the title of the resource being described, or there can be distinct elements for primary and secondary titles, translated titles, titles of articles, titles of books, and so forth. It all depends on the purpose of the metadata and the anticipated uses.

Finding Linked Data Elements

As stated in chapter 2 in the section on the Semantic Web standards, in this environment the preference is to reuse elements that have already been defined. Metadata elements that have been defined using Semantic Web standards form a universal pool from which anyone creating metadata can make choices. This is different from previous generations of metadata, where each metadata need resulted in a separate definition of data elements that were valid only internal to the local application that used the metadata. The question then becomes, "How can I find elements to use?" There is no one place to go on the Web to learn about the existence of elements; a certain amount of hunting and observing is needed. However, you needn't worry overly about missing an element that you might have

Traditional Terms	Semantic Web Terms	Coyle's Term
data elements	classes properties	elements
metadata schema	vocabulary ontology	ontology
data	values	data
controlled list	vocabulary	vocabulary

Table 3.1
Summary of terminology used in this document

used: you will define your own elements whenever you do not find one that you can use, and you can later add to your definition links to any equivalent elements that you discover or that are later defined elsewhere. While it would be ideal if there were a limited number of metadata element sets that everyone could reuse, the Semantic Web standards are designed to work in an imperfect world where the same concept may get defined in more than one environment. Still, reuse is preferred to the creation of new, redundant terms, so looking for such terms is advised.

The following are some places you might look for previously defined elements.

Swoogle

- **Name**: Swoogle
- **Creator**: eBiquity (University of Maryland, Baltimore County)
- **URL**: http://swoogle.umbc.edu
- **Created**: 2007
- **Updated**: daily

Swoogle is a Google for the Semantic Web, although it is in early stages of development. It crawls the Web looking for Semantic Web documents and performs keyword indexing on them. It has three search modes:

- **ontology** searches the full text of ontology documents
- **data** searches actual instance data
- **term** searches only terms that have been defined as classes or properties

Swoogle also archives copies of the Semantic Web documents that it finds, so it works as a kind of archive for the Semantic Web.

The Swoogle service was designed by the eBiquity Research Group of the University of Maryland, Baltimore County. It was developed as a research project but continues to be updated daily. It should not be considered to be complete by any means. If you have a dataset or an ontology that you would like to see listed in Swoogle, the site has forms where you can enter a starting URL so that your data will be indexed.

vocab.org

- **Name**: vocab.org
- **Creator**: Ian Davis
- **URL**: http://vocab.org
- **Created**: 2004
- **Updated**: 2006

The site vocab.org hosts about two dozen ontologies, including two of the FRBR-based ones, FRBR Core and FRBR Extended. Most sets of metadata terms that we encounter have been designed for a particular application. Even Dublin Core was developed for the singular purpose of describing Web resources. The terms at vocab.org do not serve a particular application, however. For the most part, they are generalized vocabularies, each covering a narrow area. For example, there is BIO, a vocabulary for biographical events, like birth, death, and marriage, that might be useful in a genealogical application or metadata for historical information. There are also lists of terms that could be used in administering sets of metadata terms, like Changeset, which contains terms relating to changes in descriptions, like additions and deletions from a dataset.

Vocab.org was developed by Ian Davis and appears to have been superseded by the site open.vocab.org. Even so, the vocab.org version of FRBR, called FRBR Core, created in 2005, is currently the most used vocabulary for expressing FRBR in linked data, in part because it was the first expression of FRBR in RDF, but also because it is a relatively simple and therefore easy to understand implementation.

Open Metadata Registry

- **Name:** Open Metadata Registry
- **Creators:** Diane Hillmann, Jon Phipps
- **URL:** http://metadataregistry.org
- **Created:** 2005
- **Updated:** 2012

A number of library-related ontologies (and vocabularies) can be found in the Open Metadata Registry (OMR). It is currently being used by the Joint Steering Committee for Development of RDA as well as IFLA. With all of the terms used in RDA, ISBD, and members of the FR family (FRBR, FRAD, and FRSAD), there is a wealth of terms to reuse in bibliographic metadata.

Libraries are not the only community using the OMR. Numerous entries are registered in the OMR sandbox (an area of the site for experimentation). While many of these are not production-ready, browsing the sandbox (from the link on the OMR homepage), you can see a number of element sets and vocabularies in development.

Linked Data Cloud

- **Name:** Linked Data cloud
- **Creators:** Richard Cyganiak, Anja Jentzsch
- **URL:** http://linkeddata.org
- **Created:** 2007
- **Updated:** 2011

As of November 2011, there are 313 datasets in the Linked Data cloud. Wherever there are datasets, there are also data elements. Most of these datasets

have documentation pages that describe their ontology. Clicking on a circle in the cloud diagram will take you to an information page for the dataset with links to the primary website. One particular advantage of reusing terms from linked data cloud–based sets is that you can be sure that your data will link to that set on the data element in question.

General Use Data Elements

While the combination of elements of a library bibliographic record may be specific to the library application, a great deal of data in library catalogs is hardly limited to library use: time periods, geographic places, scientific names, and other elements are shared with the wider world. These areas are ones where library data can find points of overlap on the Web. This section highlights some of the elements that have been defined for Semantic Web use that may be of particular interest in the development of library linked data. Some of these elements could be used directly, and others may be suitable for mapping and interoperability.

Describing Web Resources

Dublin Core: The Mother of All Metadata

- **Name:** DCMI Metadata Terms
- **Creator:** Dublin Core Metadata Initiative
- **URL:** http://dublincore.org/documents/dcmi-terms
- **Created:** October 10, 2010

The use case stated for the development of the Dublin Core Metadata Element Set (DC) at the 1995 meeting in Dublin, Ohio, was the perceived need for a simple set of metadata elements that could be used to define electronic resources, mainly Web documents. This metadata had to be usable by noncatalogers and would help make Web documents more visible to search engines. It grew out of the awareness that traditional cataloging practices would not be able to keep up with the massive growth of information resources that the Internet was making possible. The original fifteen data elements are still what most people think of when they think of Dublin Core, although beginning in 1997, the original set was extended by the introduction of extensions to the core elements. Extending Dublin Core also meant developing a philosophy of extension and methods to extend the vocabulary without breaking any uses of the original set of data elements.

While Dublin Core could be used as stand-alone metadata, it could also be embedded in Web documents using HTML `meta` tags, as shown in this example from the W3C HTML4 document:

```
<META name="DC.identifier"
content="http://www.ietf.org/rfc/
rfc1866.txt">[1]
```

In this way Dublin Core metadata could be included in the documents it was describing. This method of describing documents unfortunately fell out of favor due to the use of false meta-tagged data designed to improve placement in search results.

Dublin Core was being actively worked on at the same time as the early Semantic Web work that resulted in the RDF standard was taking place. The Dublin Core Metadata Initiative (DCMI) decided to model Dublin Core as an RDF element set, and this was completed in 2008. This set of fifty-five RDF properties includes both the original fifteen, the extensions to those, and some added elements that are appropriate to the linked data environment. In addition, there are twenty-two classes defined that provide context for the properties.

Why is Dublin Core the mother of all metadata? Not only was Dublin Core the first bibliographic metadata to be inspired by library practices, but it continues to have a key role as a core for the description of resources. Dublin Core's `dc:` namespace is the second most commonly used namespace on the Semantic Web and is probably at least that popular on the Web in general. Dublin Core is used in such common applications as Creative Commons licenses and the MusicBrainz project. Because the Dublin Core elements are not confined to a particular record format, one finds at least a few Dublin Core elements in a wide variety of metadata. It is also used in library-related applications like OCLC's CONTENTdm digital collection software and DSpace institutional repository software.

Dublin Core metadata gains a new role in the linked data operational space. What makes linked data work is linking, and what makes linking work is having commonality in your metadata elements and data. In such an environment a core vocabulary becomes a kind of link-glue that helps hold data together. Any metadata set using Dublin Core terms is essentially guaranteed to link widely. Most communities creating metadata will need to use elements that are of greater detail than Dublin Core's fifty-five elements, but it is likely that they can define their specific elements as subordinate to the Dublin Core terms (see figure 3.1). In this way, they gain compatibility with any other metadata that also defines itself as subordinate with Dublin Core terms.

The Dublin Core terms provide a foundation for resource description metadata and allow commonalities to be found through the linking process.

There are other frequently used element sets, and we'll see some below, but Dublin Core remains the pioneer in the metadata arena, being not only the first and most used of Web description metadata, but also
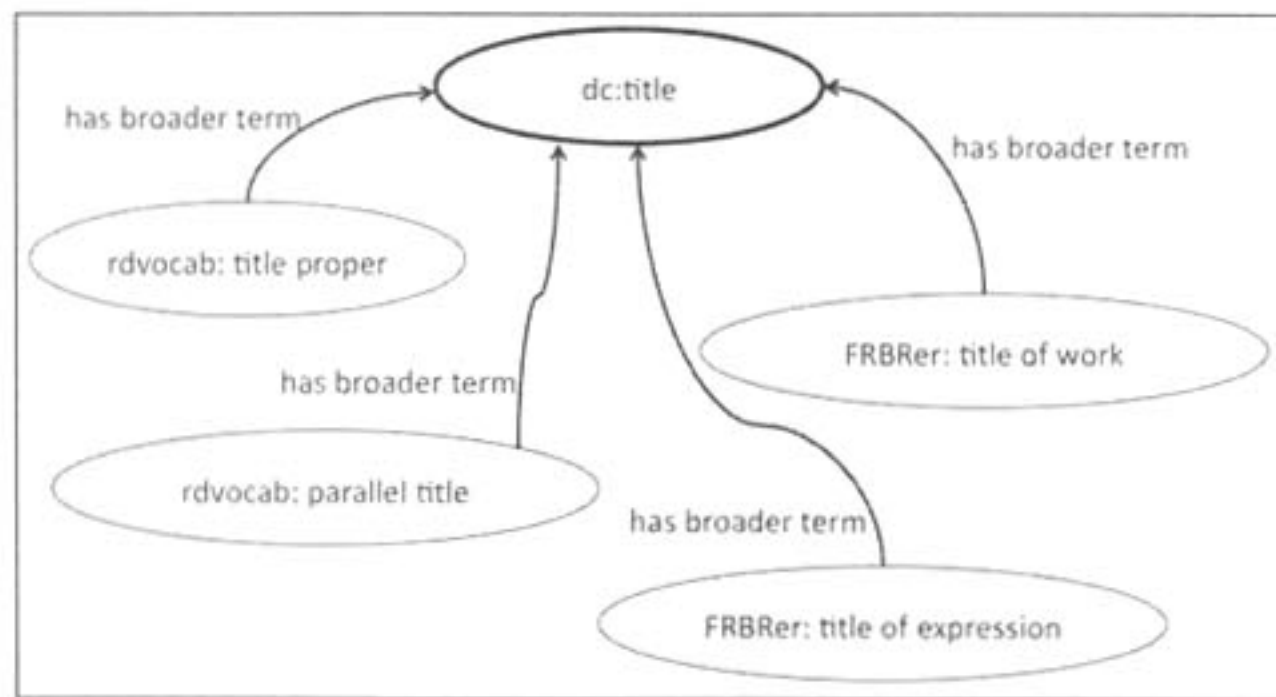
Figure 3.1
Specific titles from RDA and FRBR can link to Dublin Core for more compatibility with more generalized metadata.

the first core of metadata to be redefined for use as linked data.

RDFa

- **Name:** RDFa
- **Creator:** W3C
- **URL:** http://www.w3.org/TR/rdfa-syntax
- **Created:** 2007
- **Updated:** October 14, 2008

The initial goal of the Semantic Web was not a web of datasets but a web of data embedded in HTML documents that, unseen to the human user, enhances the meaning of the information in the text of the document. This would allow search engines to index the contents of documents meaningfully. It would also perform the self-documentation function that the HTML meta tags were designed for, allowing authors of HTML pages to add titles, author names and contact information, and any other relevant page information. It could also be used to code what a document is about by marking up information within the text of the document, like this example from the RDFa documentation, which marks up the visible text strings with calendar information:

```
<html>
  <head>
    <title>Jo's Friends and Family
    Blog</title>
    <link rel="foaf:primaryTopic"
    href="#bbq" />
    <meta property="dc:creator"
    content="Jo" />
  </head>
  <body>
  <p about="#bbq"
    typeof="cal:Vevent">
    I'm holding
    <span property='cal:summary'>
    one last summer barbecue
    </span>,
```

```
  on
    <span property="cal:dtstart"
    content="2007-09-
    16T16:00:00-05:00"
    datatype="xsd:dateTime">
    September 16th at 4pm
    </span>.
  </p>
  </body>
</html>
```

Which displays on the screen simply as:

I'm holding one last summer barbecue, on September 16th at 4pm.

Some vendors have experimented with using RDFa to mark up product pages. RDFa is based on RDF and uses the RDF rules on URIs and triples, and the complexity of RDF has made acceptance of RDFa difficult. The need for a way to mark up the content of webpages for better searching still exists, however, and in particular for the commercial websites that depend for their business on their placement in search engine results. Enter Schema.org and Schema.RDFS.org.

Schema.RDFS.org

- **Name:** Schema.RDFS.org
- **Creators:** Michael Hausenblas, Richard Cyganiak, Deri Centre
- **URL:** http://schema.rdfs.org
- **Created:** June 2011

Schema.RDFS.org is an RDF version of the microformat Schema.org. Schema.org is a microformat developed cooperatively by Google, Bing, and Yahoo! and announced to great interest in June 2011. It provides a markup of the content of webpages that is beneficial to those seeking search engine optimization. This example from the Schema.org *Getting Started* documentation shows markup that might be used on a website describing a movie:

```
<div itemscope itemtype ="http://
  schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <span>Director: <span
  itemprop="director">James
  Cameron</span> (born August
  16, 1954)</span>
  <span itemprop="genre">Science
  fiction</span>
  <a href="../movies/avatar
  -theatrical-trailer.html"
  itemprop="trailer">Trailer</a>
</div>[2]
```

Schema.org markup is simpler than that of RDFa, but it doesn't facilitate the linking that is the goal of the Semantic Web. To promote linking, members of the Semantic Web community created Schema.RDFS.org, a complementary site that repackages the Schema.org terms as RDF. The resulting markup is not noticeably more complex than that of Schema.org.

```
<div vocab="http://schema.org/"
 typeof="Movie">
 <h1 property="name">Pirates of
  the Caribbean: On Stranger
  Tides (2011)</h1>
 <span property="description">
  Jack Sparrow and Barbossa
  embark on a quest to Find the
  elusive fountain of youth, only
  to discover that Blackbeard and
  his daughter are after it too.
 </span>
  Director:
 <div rel="director">
  <div typeof="http://schema.org/
   Person">
   <span property="name">Rob
    Marshall</span>
  </div>
 </div>
 <div rel="author">
  Writers:
   <div typeof="Person">
    <span property="name">Ted
     Elliott</span>
   </div>
   <div typeof="Person">Terry
    Rossio</div>
    , and 7 more credits
    Stars:
   <div typeof="Person">
    <span property="name">Johnny
     Depp</span>,
   </div>
   <div typeof="Person">
    <span property="name">Penelope
     Cruz</span>,
   </div>
   <div typeof="Person">
    <span property="name">Ian
     McShane</span>
   </div>
  </div>
 </div>³
```

If there is a lesson here, it is more about marketing and user-friendliness than technology. The Semantic Web community failed to communicate the underlying simplicity of RDFa to potential users. Pushed by the triumvirate of search engines, RDFa is proving to be more user-friendly than was originally thought.

People

We are a species-centric species: much of our attention currency is spent paying attention to people. The incredible rise of social media is ample evidence of this. Libraries are also about people in the sense that they collect and organize the long, slow conversation that is human culture. It makes sense, then, that among the first linked data schemes we find ones for people and their relationships.

Friend of a Friend (FOAF)

- **Name**: Friend of a Friend (FOAF)
- **Creators**: Dan Brickley, Libby Miller
- **URL**: http://xmlns.com/foaf/spec
- **Created**: 2000
- **Updated**: August 9, 2010

Friend of a Friend (FOAF) was developed by Dan Brickley and Libby Miller as an early proof of concept for RDF. It began as a way to encode information about persons in a social networking context, but it has evolved into one of the primary ways that people create metadata for persons on the Semantic Web.

FOAF is a linked data–compatible way to record descriptions about people, their places, and their relationships to other people. The resulting description is called a FOAF profile, something like a vCard but with a much richer context. While FOAF can be used to describe persons other than oneself, it is particularly suitable for people on the Web keeping updated information about themselves and their ever-growing network of connections to people, institutions, and websites.

Work began on the FOAF concept in 2000, and the full set of elements was published in 2008. It continues to be updated in response to community needs and suggestions. Like many Web-based projects, the FOAF community is made up of volunteers who communicate through e-mail lists and do their work on a wiki.

Because FOAF was the first RDF project to define a class for `Person` with key elements like `name`, `mbox` (mail address), and `homepage`, it is used in many early linked data efforts. FOAF is divided into a core (`name`, `title`, `knows`, `age`, and others) and a set of social Web elements (`nick`, `mbox`, `homepage`, `weblog`, `interest`, `publications`, `schoolHomepage`, and more). FOAF covers some of the elements used for persons in library authority data, but does not include birth and death years or concepts like "flourished." In addition, there isn't the concept of preferred and alternate name forms, although any number of names can be used for an individual. Some experimentation

would be needed to see if it could accommodate some of the more complex name forms that library authority data must create. It also has a class for organizations, mainly because people have relationships to organizations like schools and workplaces. The organization class is relatively undeveloped in FOAF, but as we have discussed in relation to Dublin Core, it could serve as a general element that other communities refine with more detail.

FOAF's modern social Web roots are visible in its use of `birthday` (day and month) but not the year, since full birthdates are among the key identity theft data elements. It has a somewhat touching lack of death dates, perhaps because the developers were not yet anticipating how to handle death on the social Web.

FOAF is Web-savvy in many ways: for example, it can hide your e-mail address behind an SHA-1 hash key. If you take this option, the e-mail address acts as an identifier but not contact information. You can choose to have your e-mail address "in clear" in your FOAF profile if you wish.

You can create your own FOAF profile using FOAF-a-Matic (see chapter 5). There have also been some programs written to export FOAF data from common social networks, such as Facebook. These latter encode not only your information but the names of those you "know" in your social network.

Here is a FOAF profile I made for myself with FOAF-a-Matic. It has the basic contact information, plus a statement that I know Dan Brickley and gives a link to his FOAF profile. My FOAF profile can be found at `http://kcoyle.net/foaf.rdf`. This is a common location and filename for FOAF files on personal websites and is retrievable in a search engine with a search on `foaf.rdf`.

```
<rdf:RDF xmlns:rdf="http://www.
  w3.org/1999/02/22-rdf-syntax-
  ns#" xmlns:rdfs="http://www.
  w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/
  foaf/0.1/" xmlns:admin="http://
  webns.net/mvcb/">
<foaf:PersonalProfileDocument
  rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic
    rdf:resource="#me"/>
  <admin:generatorAgent
    rdf:resource="http://www.
    ldodds.com/foaf/foaf-a-matic"/>
    <admin:errorReportsTo
      rdf:resource="mailto:leigh@
      ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="me">
```

```
<foaf:name>Karen Coyle</
  foaf:name>
<foaf:title>Ms.</foaf:title>
<foaf:givenname>Karen</
  foaf:givenname>
<foaf:family_name>Coyle</
  foaf:family_name>
<foaf:nick>kc</foaf:nick>
<foaf:mbox_sha1sum>1cb4607d1d88
  aca847b1e1a0179383323032f50a</
  foaf:mbox_sha1sum>
<foaf:homepage rdf:resource=
  "http://kcoyle.net/"/>
<foaf:depiction rdf:resource=
  "http://kcoyle.net/img/kc_head.
  jpg"/>
<foaf:workInfoHomepage
  rdf:resource="http://kcoyle.
  net/reach.html/"/>
<foaf:knows>
<foaf:Person>
<foaf:name>Dan Brickley</
  foaf:name>
<foaf:mbox_sha1sum>748934f32135
  cfcf6f8c06e253c53442721e15e7</
  foaf:mbox_sha1sum>
<rdfs:seeAlso
  rdf:resource="http://danbri.
  livejournal.com/data/foaf"/></
  foaf:Person></foaf:knows></
  foaf:Person>
</rdf:RDF>
```

BIO and RELATIONSHIP

- **Name:** BIO
- **Creators:** Ian Davis, David Galbraith
- **URL:** http://vocab.org/bio/0.1/.html
- **Created:** March 7, 2003
- **Updated:** June 14, 2011

- **Name:** RELATIONSHIP
- **Creators:** Ian Davis, Eric Vitiello Jr.
- **URL:** http://vocab.org/relationship/.html
- **Created:** February 11, 2004
- **Updated:** March 19, 2010

Where FOAF is about contact information and activities, BIO ("A vocabulary for biographical information") is likely to appeal to genealogy buffs for its emphasis on the key events of a person's life—birth, graduation, marriage, employment, retirement—and the darker moments including divorce, death, murder, and imprisonment. BIO can be used to describe a person's life as a series of events, some biological, some social. Combined with the elements defined in the set called RELATIONSHIP, you have nearly everything

you can say about any group of people. RELATION-SHIP covers everything from "knows in passing" to "mentor of" and the full set of family relationships from "spouse of," "sibling to," and the very modern "life partner of."

Both of these element sets were developed by Ian Davis (with collaborators David Galbraith for BIO and Eric Vitiello Jr. for RELATIONSHIP). They were designed to be compatible with FOAF, and therefore they do not repeat any of the information in FOAF, such as name or gender.

Elements from BIO are used, for example, in the RDF output for author pages in the Open Library, which also use elements from FOAF and RDA.

```
<foaf:Person rdf:about="http://
  openlibrary.org/authors/
  OL22022A">
<foaf:name>Barbara Cartland</
  foaf:name>
<rdg2:variantNameForThePe
  rson>Mary Barbara Hamil
  ton Cartland McCorquodale</
  rdg2:variantNameForThePerson>
<rdg2:titleOfThePerson>Dame</
  rdg2:titleOfThePerson>
<bio:event>
 <bio:Birth>
  <dcterms:date>9 July 1901</
    dcterms:date>
 </bio:Birth>
</bio:event>
<bio:event>
 <bio:Death>
  <dcterms:date>21 May 2000</
    dcterms:date>
 </bio:Death>
</bio:event>
```

Open Graph

- **Name**: Open Graph
- **Creator**: Facebook
- **URL**: http://developers.facebook.com/docs/opengraph
- **Created**: 2010
- **Updated**: 2011

Behind the by now nearly ubiquitous Like buttons on the Web is a protocol called Open Graph that was developed by Facebook. Open Graph extends the Facebook links of people and relationships, adding a Like link from a Facebook identity to any webpage or resource on the Web. When a person clicks on Like, a link is posted to Facebook that makes the connection between that Facebook persona and the Web object. Open Graph's base design makes use of meta

tags within the HTML of the participating webpages, but the conceptual connection to the linked data is obvious, and the Facebook API is experimentally producing output in a linked data format. It makes sense that social relationship data will be part of the Web of data. The controversy around privacy will continue, of course, but of all possible linking the inherent relationships between persons seems too important to ignore.

Geography

Another very common type of information that we need data elements for is information about the geographical and geopolitical world we live in. The place where something is or occurs is a key element of understanding and searching. Maps are heavily used on the Web, in no small part due to the ubiquity of online mapping programs.

Geographical information is an obvious nexus for linking on the Web of data. Datasets that cover this information are available for use.

GeoNames Ontology

- **Name**: GeoNames Ontology
- **Creator**: GeoNames Community (Marc Wick, founder)
- **URL**: http://www.geonames.org/ontology/documentation.html
- **Updated**: February 2012 (version 3.01)

The GeoNames dataset (figure 3.2) is one of the most linked-to datasets in the linked data cloud. Its data element set, the GeoNames Ontology, has only a few data elements, but these cover the key linking concepts of "child" (for subordinate or narrower places), "nearby" (for physical proximity), and "neighbor" (for places sharing a common boundary). The categorization of places into types, such as city, lake, and continent, is covered by the extensive controlled vocabulary of nearly 700 terms.

The data is searchable through a Web interface as well as accessible in a machine-actionable linked data format.

FAO Geopolitical Ontology

- **Name**: FAO Geopolitical Ontology
- **Creator**: FAO
- **URL**: http://www.fao.org/countryprofiles/geoinfo/geopolitical/resource
- **Updated**: August 11, 2011 (version 1.1)

The United Nations Food and Agriculture Organization (FAO) has long been a leader in the development of data services in its functional area. FAO is an early adopter of Semantic Web principles. The FAO

2 (H)	Cold Lake	Canada, Alberta	lake	N 54° 33' 0"	W 110° 5' 6"
3 (P)	Cold Lake	Canada, Alberta	populated place population 11,595	N 54° 27' 0"	W 110° 12' 6"

Figure 3.2
An example from GeoNames showing the distinction between a city and a lake with the same name.

geopolitical information has an emphasis on providing key governance, geographical, demographic, and economic indicators. Beyond the primary geographical elements of name and coordinates (hasMaxLatitude, hasMaxLongitude, hasMinLatitude, hasMinLongitude), there are elements describing currency and population. Acknowledging the complexity of political reality, political units are coded as being self-governing, non–self-governing, or disputed. To enhance linking, the ontology has elements for information about economic group membership, such as the European Union or the Arab Maghreb Union.

Rights

Whether we make it explicit or not, all cultural expressions have a rights status, even if that status is unknown. The language of expression of these rights is that of the legal code or contractual document. As yet, creating a set of data elements for intellectual property rights has eluded data creators, and it may continue to do so because legal language is far from the algorithmic language of data. There is, however, one use of rights declarations online, the Creative Commons (CC) rights declarations, and it is Semantic Web–compatible. While designed for use with the CC licenses, which are limited in scope, the data elements themselves could be useful in other rights declarations.

Creative Commons Rights Expression Language

- **Name:** Creative Commons Rights Expression Language
- **Creator:** Creative Commons
- **URL:** http://creativecommons.org/ns
- **Created:** 2008

The Creative Commons Rights Expression Language (CC REL) has elements organized in five groups: permissions, requirements, prohibitions, license properties, and work properties. It is possible to specify rights like reproduction and distribution and requirements such as attribution and share-alike. Because CC REL is designed for use on the Web, it recognizes that rights are jurisdictional, and therefore has elements for jurisdiction and the relevant legal code. Descriptive information in CC license about the work and the

creator use Dublin Core properties, since these satisfy the minimal needs for description. One could alternatively use FOAF data elements to give more detailed information about the creator or rights holder, including contact information.

Some of these elements could be relevant to cultural heritage data that is not using the CC licenses themselves. Archival data in particular could use these elements to record what rights information it has for materials.

Citations

For an active researcher, managing citations is an important but tedious part of the research and writing process, and there are a number of applications that help researchers gather, organize, and use citations. Citing and being cited are important parts of the scholarly conversation that takes place over time as scholars build on the thoughts and discoveries of those who preceded them. Citations to your work by others is a measure of the importance of your work and can have an effect on career and even pay scale.

Citations are themselves links from one document to another, so it makes sense that one would want to integrate citations with the linked data Web. Combined with the increased access to full digital text online, it becomes possible to make those virtual links real—to actually follow citations from document to document, and perhaps even create maps that show interesting patterns of influence in a field of study. The two citation metadata schemes here are ones that have been designed as linked data.

Bibliographic Ontology (BIBO)

- **Name:** Bibliographic Ontology
- **Creators:** Frédérick Giasson, Bruce D'Arcus
- **URL:** http://bibliontology.com
- **Created:** 2008
- **Updated:** 2011

The Bibliographic Ontology, commonly shortened to BIBO, was created by Bruce D'Arcus and Frédérick Giasson in 2008. It can be used for citations or for other bibliographic metadata needs. This was the first significant set of bibliographic metadata for the

Semantic Web, and elements of it are used in many implementations, including the Library of Congress Newspaper Project and the British Library's National Bibliography.

BIBO defines about sixty document types that form the backbone of the scheme. Some type examples are Proceedings, Bill, Book section, and Film. Its academic roots are visible in many areas, such as its treatment of authors. Authors are presented as an ordered list. "Normally, this list is seen as a priority list that order[s] authors by importance."[4] This reflects the importance not only of publishing in the academic environment but also of where one appears in an authorship statement. Another area of academic interest is the status of the document, such as peer-reviewed, accepted, and rejected.

BIBO has elements for ISBN, ISSN, and even the OCLC number, and these are heavily used simply because they have not been defined elsewhere for use in linked data. There should be little need for metadata elements for identification numbers in linked data because the standard is to use Web-based identifiers, or URIs. These metadata elements are needed because the maintenance organizations have not provided a Web-standard format. An example of a standard identifier is the LCCN permalink that is found on every LC catalog record:

```
LCCN permalink: http://lccn.loc.
gov/2011020573
```

This identifier can be used in linked data without having to code it as an LC identifier, since the URI contains all of the information that is needed to be unique on the Web. It also serves as a link back to the thing it identifies, an entry in LC's catalog. One hopes that eventually the parties responsible for issuing and maintaining key identifiers will provide a Web-friendly format for them.

BIBO makes use of some elements from other linked data schemes, such as Dublin Core and PRISM, the latter a metadata scheme for publication and syndication.

Semantic Publishing and Referencing (SPAR)

- **Name:** Semantic Publishing and Referencing
- **Creators:** David Shotton, Silvio Peroni
- **URL:** http://purl.org/spar
- **Created:** 2010

Semantic Publishing and Referencing (SPAR) is a suite of eight Semantic Web metadata sets that encompass bibliographic citations, publishing status and workflow, citation types, and document formatting. This is an ambitious undertaking. It is especially interesting because it takes an event-driven workflow view rather than the generally static approach of bibliographic description.

SPAR makes use of metadata terms from FOAF, Dublin Core, and SWAN (Semantic Web Applications in Neuromedicine). Both SPAR and SWAN go beyond the simple concept of citation and categorize the intention or meaning of the citation, such as "in response to," "agrees with," or "disagrees with." SPAR also uses the FRBR model for its bibliographic description. At the time of this writing, four of the eight modules of SPAR have been developed:

- FaBiO—FRBR-aligned Bibliographic Ontology. This is the bibliographic description module. It is oriented toward academic texts, both print and digital. Its use of FRBR differs considerably from, for example, the use of FRBR in RDA. Where RDA has the concept of genre that is a list of terms, FaBiO treats its resource types as more specific types of FRBR:Expression. The FaBiO Expression types include things like "article," "book," "news item," "spreadsheet," and "dust jacket." Work is similarly subclassed by types of Works, which include "critical edition," "questionnaire," and "reference work."
- BiRO—Bibliographic Reference Ontology. BiRO is metadata for describing bibliographic records and references, as well as collections of these. It includes relationships between references with "references" and "is referenced." It has "annotation properties," which are administrative data elements that can provide information about the metadata, its creator, and date.
- CiTO—Citation Typing Ontology. The CiTO metadata allows one to characterize the nature of the citation, and in considerable detail. CiTO has nearly seventy different citation types, although this includes references in both directions, such as "corrects" and "is corrected by."
- C4O—Citation Counting and Context Characterization Ontology. C4O will be used in applications that do citation counting. It is designed not only to count that a document is cited but also to count the number of citations when a document is cited more than once in the citing document.

Bibliographic Description in Libraries

FRBR (FRBR Core, FRBR Extended, FRBRer Model, FRBRoo)

- **Name:** FRBR Core
- **Creators:** Ian Davis, Richard Newman, Bruce D'Arcus
- **URL:** http://vocab.org/frbr/core.html
- **Created:** 2005
- **Updated:** May 16, 2009

- **Name:** FRBR Extended
- **Creators:** Ian Davis, Richard Newman
- **URL:** http://vocab.org/frbr/extended.html
- **Created:** July 29, 2005
- **Updated:** August 10, 2005

- **Name:** FRBRer Model
- **Creator:** IFLA FRBR WG
- **URL:** http://iflastandards.info/ns/fr/frbr/frbrer
- **Created:** September 30, 2010

- **Name:** FRBRoo
- **Creator:** International Council of Museums (CIDOC)
- **URL:** http://www.cidoc-crm.org/frbr_inro.html
- **Created:** 2006
- **Updated:** January 2008

FRBR was one of the earliest library standards to be expressed as an RDF element set. It is also the one that has been created in the most versions and with significant differences between them. FRBR provides a lesson for any community that has data standards: sometimes being the official version is less important than being the first version. When new technologies are coming into being, enthusiastic early adopters cannot wait for organizations to catch on to the coming trends, and experimental versions by third parties can become focused in the consciousness of developers.

Discussed below are three well-known implementations of FRBR using linked data standards. There are others, such as in the SPAR suite of citation metadata. It is notable that each of the interpretations is different, and some are considerably so. FRBRer can be considered the "official" version, since it was developed by the IFLA FRBR Study Group. Each of the nonofficial versions makes some changes to the model. More than one of them adds an element for each of the FRBR groups that represents the group as a whole, something that is not valid in the IFLA version. There are also new links defined that allow, for example, a Manifestation to link to a Work, or an Item to link to an Expression, without the intervening FRBR entities. If these additional elements prove useful in implementations of FRBR, it would be advisable for the IFLA group to take these practical concerns into consideration.

Note that there is an even more reduced version of FRBR, including only the ten entities that were developed for the RDA element set work, since an IFLA-approved version of FRBR in Semantic Web format was not available when the RDA elements were being defined. This version can be found at the Open Metadata Registry, called FRBR Entities for RDA. This may be aligned with FRBRer (described below) at some point.

FRBR Core was created in 2005 by Ian Davis and Richard Newman and is available on the vocab.org site. FRBR Core is a limited rendering of FRBR describing only the FRBR entities and the relationships between

them. There is a lesser-known version of FRBR also on the vocab.org website called FRBR Extended that includes the FRBR attributes. In part because FRBR Core was the first, and perhaps in part because it is a relatively small and therefore easy-to-understand set of elements, FRBR Core has been reused in more Semantic Web element sets than any other version of FRBR. FRBR Core is based on the 1998 FRBR document and has not been updated since 2005; therefore, it may vary from the current version of FRBR. It does add some more abstract entities for each FRBR group that represent the entire group. This concept is not included in the version of FRBR produced by the IFLA Study Group. FRBR Core is used by FaBiO (discussed above) and by the UK legislation documentation managed by the National Archives. The latter produces a dataset that can be found in the linked data cloud (http://www.legislation.gov.uk/).

FRBRoo is an object-oriented interpretation of FRBR. It was developed by the International Consortium of Museums as an integration of FRBR concepts with its Conceptual Reference Model (CIDOC CRM). FRBRoo is being harmonized with the IFLA FRBR model. The CIDOC CRM is a standard that has been in progress for over a decade, and there are some early adopters. These institutions are using the XML DTD version of the terms, and it is likely that the data elements will be available in both an XML DTD and in Semantic Web format for the foreseeable future.

FRBRer is the "official" version of FRBR from the IFLA FRBR Study Group. It has been entered into the Open Metadata Registry by Gordon Dunsire, consultant to that group. FRBRer has ten classes and 206 elements in its ontology. The classes correspond to each of the FRBR entities. The elements include both attributes and relationships as defined in FRBR's 2008 document. This element set was produced in 2010 and approved by the study group in 2011.

ISBD elements

- **Name:** ISBD
- **Creator:** Gordon Dunsire, ISBD Review Group
- **URL:** http://iflastandards.info/ns/isbd/elements
- **Created:** 2009
- **Updated:** June 2011

In 2009, working with consultant Gordon Dunsire, the ISBD review group decided that ISBD terms should be expressed using the Semantic Web standard, RDF. These elements reflect the flat ISBN structure with eight bibliographic areas. There are a total of 181 classes and properties in the set of ISBD elements in the Open Metadata Registry. The British Library Data Model for its Semantic Web implementation makes use of a small number of the ISBD elements, as does the project of the Mannheim University Library. Both of these uses are evidence of the mix-and-match nature of linked data.

The British Library model also uses terms from Dublin Core, BIO, and the Bibliontology, among others.

RDA elements

- **Name**: RDA Elements
- **Creator**: DCMI/JSC Working Group
- **URL**: http://rdvocab.info
- **Created**: 2009
- **Updated**: 2011

It was an idea whose time had come: participants in the Dublin Core Metadata Initiative (DCMI), the Semantic Web effort, and libraries saw a unique opportunity coming out of the development of the cataloging rules called Resource Description and Access. The new rules could also foster a new era for library data. At a meeting hosted by the British Library in May 2007, DCMI and the Joint Steering Committee for Development of RDA (JSC) agreed to work together to create a standard Semantic Web implementation of the elements and controlled vocabularies of RDA.

This work took place over the 2008–2010 time frame, and today all of the elements and lists defined by the JSC are registered in the Open Metadata Registry. There are over 1,300 elements, but the elements are grouped based on their adherence to FRBR entities to make it a bit easier to navigate.

In keeping with the mix-and-match nature of linked data, some elements from RDA are being used in linked data projects. The British Library uses at least one RDA term, and the Open Library's linked data uses some RDA terms when describing works, authors, and editions (the last being similar to a FRBR Manifestation).

Creating the standard basis for linked data brings up some questions that are particular to the process of turning a thoughtful set of cataloging rules into something that not only can be manipulated by programs but that also will play well with other linked data on the Web. For example, in order to have a way to make relationships between RDA elements and other bibliographic elements being used in linked data on the Web, it is necessary to have a version of each RDA element that is not limited to its FRBR entity. This is because most other bibliographic data creators will not be following the FRBR model. For RDA data to connect to non-FRBR bibliographic data, it will need links to general bibliographic concepts like those in Dublin Core and BIBO that are not bound to FRBR entities.

Authority Data

Simple Knowledge Organization System (SKOS)

- **Name**: Simple Knowledge Organization System (SKOS)
- **Creator**: W3C

- **URL**: http://www.w3.org/2004/02/skos
- **Created**: 2005
- **Updated**: 2009

Simple Knowledge Organization System (SKOS) is a W3C standard element set for describing thesauri or other term lists. As the word simple in its name implies, SKOS has basic thesaurus functionality: broader term, narrower term, and related term. It also has the concepts of preferred label and alternate label that function much like the library authority concepts of authoritative and nonauthoritative terms. SKOS is much used for linked data, and not only for term lists: its elements for labeling, which also includes hidden labels, are frequently used to indicate display decisions in element sets that are not of the list or thesaurus type.

As a linked data standard, SKOS makes use of all of the basic concepts of the Semantic Web including the use of identifiers for each "thing" and built-in multilingual capability. Each language can have a single preferred term and an unlimited number of alternate or hidden terms. Some examples are given in the introduction to chapter 4.

MADS in RDF

- **Name**: MADS in RDF
- **Creator**: Library of Congress
- **URL**: http://www.loc.gov/standards/mads/rdf
- **Created**: November 19, 2010
- **Updated**: March 28, 2011

SKOS is suitable only for thesauri or vocabularies that have a simple structure, with each "node" being a single term. This is not sufficient to describe a precomposed faceted vocabulary like the Library of Congress Subject Headings. The Library of Congress has developed an authority element set based on MADS (Metadata Authority Description Schema), which is itself based on the MARC authority record. MADS does not use elements from SKOS but instead makes use of the more comprehensive element set of the Web Ontology Language, OWL.

MADS in RDF supports complex headings, like those in LCSH or the author/title authority entries. For those particularly complex situations where one heading breaks into two or multiple headings are combined into one, MADS is able to represent the complex cross-references that arise out of such a change, something that cannot be achieved in SKOS.

Functional Requirements for Authority Data (FRAD)

- **Name**: Functional Requirements for Authority Data (FRAD)
- **Creator**: Gordon Dunsire, IFLA Working Group on FRAD

- **URL**: http://iflastandards.info/ns/fr/frad
- **Created**: 2010
- **Updated**: 2011

In the FRBR family, Functional Requirements for Authority Data (FRAD) covers the area of name authority. The terms of this vocabulary have been provisionally registered in the Open Metadata Registry. This means that they are still under review by the group and are not considered official until their status has been changed to "published." Note that as with other members of the IFLA Functional Requirements family, the elements include descriptors as well as terms for relationships between described entities.

Functional Requirements for Subject Authority Data (FRSAD)

- **Name**: Functional Requirements for Subject Authority Data (FRSAD)
- **Creator**: Gordon Dunsire, IFLA Working Group on Functional Requirements for Subject Authority Records
- **URL**: http://iflastandards.info/ns/fr/frsad
- **Created**: 2010
- **Updated**: 2011

As its name indicates, FRSAD is a description of subject authority data. The FRSAD elements are registered in the Open Metadata Registry. Because of the FRSAD approach to subject authorities (it has only two entities, Thema and Nomen), the FRSAD element set consists only of nineteen elements and relationships, some of which are administrative in nature ("reference source," "script of nomen"). The elements of FRSAD have been given a status of "published," which means that they are currently available for use.

Preservation Metadata Elements

Preservation Metadata: Implementation Strategies (PREMIS)

- **Name**: PREMIS in RDF
- **Creators**: Rob Sanderson, Ed Summers, John Harrison
- **URL**: http://annotation.lanl.gov/premis
- **Created**: 2010

PREMIS development began in 2003 with a joint working group of OCLC and RLG, and a first version of the data dictionary for preservation metadata was issued in 2005. The PREMIS data dictionary contains detailed elements for all aspects of preservation, including an event-driven view that can be used to chronicle a history of preservation actions on a resource.

PREMIS is now available as an OWL ontology so that it can be used with linked data. Note that at the time of this writing, the PREMIS elements were considered preliminary, and therefore they may undergo modification before the element set is considered production-ready. PREMIS joins a growing list of standards that are made available in a variety of formats so that they can be used as widely as possible.

PRONOM Vocabulary

- **Name**: PRONOM
- **Creator**: UK National Archives
- **URL**: http://www.nationalarchives.gov.uk/PRONOM/Default.aspx
- **Created**: 2011

PRONOM is an information system developed by the Digital Preservation Department of the UK National Archives to support its digital preservation activities. PRONOM addresses the issue of describing file formats and the applications that can read and write them. A linked data expression of the PRONOM elements for describing digital file formats is available in a test version.

Notes

1. Dave Raggett, Arnaud Le Hors, and Ian Jacobs, eds., "7 The Global Structure of an HTML Document," in HTML 4.01 Specification, W3C website, December 24, 1999, www.w3.org/TR/html4/struct/global.html.
2. "Getting Started with Schema.org," last modified June 29, 2011, http://schema.org/docs/gs.html.
3. "Schema.org Movie," Structured Data Linter, accessed March 7, 2012, http://linter.structured-data.org/examples/schema.org/Movie.
4. Bruce D'Arcus and Frédérick Giasson, "Bibliographic Ontology Specification," revision 1.3, November 4, 2009, http://purl.org/ontology/bibo.