

# RDA in RDF

## Abstract

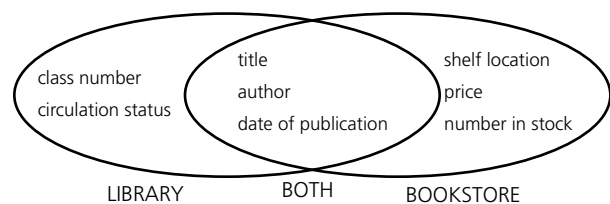
*The development of new cataloging rules that are based on the domain model provided by FRBR affords an opportunity to “data-fy” the underlying elements of the cataloging activity. In conjunction with members of the Dublin Core Metadata Initiative, the data elements identified in RDA have been defined using current Semantic Web standards. The elements now exist in an openly accessible registry on the Web where they can be downloaded and used by anyone wishing to describe bibliographic data. This work dovetails with similar efforts at the Library of Congress to define its key vocabularies in another Semantic Web format, SKOS. Together, these registered data elements can form the basis of a new generation of library data that can interact in the larger information space of linked data on the Web.*

There is a tendency today for different communities to create different metadata sets for similar, but not identical, needs. One has little choice when the metadata set, as defined, must be used as a whole or not at all. This is the case when the metadata is defined as a particular record structure, and the data elements are neither extendible nor reusable outside of that structure.

Once data elements are defined independently of a particular record standard, however, it becomes possible to create different applications using some of the same data elements. In theory, a bookstore and a library could use the same data elements where their interests are the same: title, publisher, year of publication. They could each also have different data elements for areas where they have different needs. Thus a library would have classification

numbers and circulation information, while a bookstore would have shelf location and pricing (see figure 18).

It is only by defining our data elements independently of a record structure that this kind of sharing will become possible. In the Semantic Web world, the definition of data elements is the creation of an *ontology*, which is an expression of the vocabulary of a particular domain. It so happens that at the time that the creation of an ontology for library data started to be of interest to some in the library community, the community was also undergoing a major change in its approach to the creation of metadata, first because of FRBR, and next because of the development of RDA.<sup>1</sup> FRBR uses an entity-relation (ER) model for the description of the bibliographic domain of interest to libraries, and RDA consciously incorporates the FRBR entities and relationships into the cataloging rules. While a Semantic Web-based vocabulary could be created for the current cataloging rules, AACR2, there is an advantage to making this first effort with rules that have an explicit ER model as their basis.<sup>2</sup>



**Figure 18**  
Overlap and differences in required metadata.

## RDA Background

The cataloging rules issued in a final draft in 2009 under the name Resource Description and Access (RDA) are the result of nearly ten years of study and are the culmination of nearly 150 years of thought about catalogs and cataloging. RDA is the first major revision of the rules governing library cataloging practices since the development of FRBR and was consciously aligned with the entity-relationship model of FRBR.

Like cataloging rules before it, RDA serves multiple functions. It is a set of rules that guide catalogers in the decisions that they must make in the course of creating a catalog entry. It is also implicitly a statement of the data elements that make up the bibliographic description. What RDA is *not* is as important as what it is. It is not a prescription for a machine-readable record format. RDA defines in some detail the strings that must be created to represent elements of the description, such as the recording of titles of works and the creation of access points. Although RDA states in its Prospectus that “it establishes a clear line of separation between the recording of data and the presentation of data,”<sup>3</sup> the descriptions and examples are recorded primarily as text strings. Some of those strings are necessarily of the nature of free text because they are transcriptions of data from the resource itself. Other strings may be entries from controlled vocabularies, including forms of names in authority-controlled entities such as the names of persons or corporate bodies.

RDA, as conceived by the Joint Steering Committee charged with its development, also includes a data element set. Each data element described in RDA is associated with one or more FRBR entities and has one or more possible value types. This is detailed in the element analysis of the final RDA draft.<sup>4</sup>

As a document, however, the elements are essentially inert; they exist on paper but not in a machine-actionable form. There is no direct path from the documentation to anything that could be used in a computer application.

A group of metadata developers active in the Dublin Core community recognized that RDA was on

the threshold of making the transition from *conceptual* to *actionable* metadata. What it needed, though, was the creation of a machine-actionable ontology from the documented RDA data elements. In a meeting funded by ALA Publishing and held at the British Library on April 30 and May 1, 2007, representatives of DCMI met with members of the JSC and offered to collaborate on the creation of an RDF-compatible expression of the RDA element set, including the association with FRBR entities and relationships. This work was carried out by Metadata Management Associates and volunteers in the metadata community, with funding from the British Library and Siderean Software. The result is an online registry of RDA in RDF, the first definition of library cataloging data in a Semantic Web format.

## RDA in RDF

The definition of RDA in RDF uses three basic components: the FRBR entities (Groups 1, 2, and 3); the RDA-defined properties from the RDA element analysis, including the relationships between entities as defined in RDA; and the many lists of terms that are sprinkled throughout the RDA document itself. These latter are called “value vocabularies,” using the Dublin Core Abstract Model terminology.

*Dublin Core Abstract Model*

<http://dublincore.org/documents/abstract-model>

## FRBR Entities

The FRBR entities serve as what RDF defines as *classes*. A class is a way to gather together like things so that we can say that both *Hamlet* and *Moby Dick* are members of the class Work, and “William Shakespeare” and “Herman Melville” are members of the class Person. Classes have particular attributes, known in RDF as *properties*. The properties of a Work, as defined by RDA, include a title and a form, while properties of a Person include name and dates of birth and death. In this way, the FRBR entities are the general organizing principle of the RDA element description.

## RDA Properties

Each data element defined by RDA is considered an RDF *property*. There are over 1,300 properties in the registered version of RDA, some of which are subproperties of other properties. The formal definition of a property follows conventions established in the Semantic Web world, including the extensions developed by the Dublin Core community.

*British Library*  
[www.bl.uk](http://www.bl.uk)

*Siderean Software*  
[www.siderean.com](http://www.siderean.com)

*Registry of RDA in RDF*  
<http://metadataregistry.org/rdabrowse.htm>

Element Sets: RDA Group 1 Elements	
Elements: Title proper	
Detail	Statements
<b>Metadata +</b>	
<b>Detail</b>	
Label:	Title proper
Name:	titleProper
URI:	http://rdvocab.info/Elements/titleProper (RDF)
Description:	The chief name of a resource (i.e., the title normally used when citing the resource).
Comment:	
Type:	subproperty
Parent:	Title
Domain:	
Range:	
Status:	New-Proposed
Language:	English
Note:	

**Figure 19**  
Overview of a property.

The high number of elements is due in part to the need by the developers of RDA to have a specific entry for each element with its pairing to a FRBR Group1 entity. It is this paired element that connects directly to the text of RDA and to the list of elements and definitions in the RDA documentation. The registrars chose to encode an entry for an element independent of its FRBR entity and an entry (or entries in the case of elements that can be associated with more than one FRBR entity) for the element and its associated FRBR entity to allow for extensibility. Registered properties are available online in a human-readable display with both an overview and a display of statements (see figures 19 and 20).

When accessed by a program rather than a browser, the registry entry is returned in a machine-readable format—RDF/XML in the case shown in figure 21.

The same registry data serves machine-processing needs as well as a useful display for metadata creators and any metadata applications that have user-oriented displays. It is not necessary to maintain two separate versions of the same information in order to serve both human users and programmatic needs.

The elements of the registry entry for properties are as follows:

- **Identifier (URI)**—a Semantic Web-compatible identifier that begins with “http://rdvocab.info/” identifying each term.
- **Name**—a machine-friendly form of the name of the element, generally in “camel case”: titleProper.
- **Label**—a human-display label for the element: “Title proper.” Labels are language-specific. RDA provides labels in English, but labels can be added in any language.
- **Description**—a human-readable definition of the ele-

Element Sets: RDA Group 1 Elements						
Elements: Title proper						
Detail	Statements	History				
Profile property	Show object	Lang	Status	Updated	Updated by	Actions
name	titleProper	English	New-Proposed	13 July 2008 18:17	kcoylenet	
label	Title proper	English	New-Proposed	13 July 2008 18:17	kcoylenet	
type	subproperty	English	New-Proposed	13 July 2008 18:17	kcoylenet	
subpropertyOf	Title	English	New-Proposed	13 July 2008 18:17	kcoylenet	
description	The chief name of a resourc...	English	New-Proposed	1 December 2008 17:07	kcoylenet	
hasSubproperty	Title proper (Manifestation)	English	New-Proposed	9 August 2009 15:01	DianeH	
<b>6 results</b>						

**Figure 20**  
Statement view of a property.

ment or term. The descriptions in the registry are those supplied in the RDA Glossary. For example, the description for Title proper is “The chief name of a resource (i.e., the title normally used when citing the resource).” Like labels, descriptions are language-specific, and others could be added in other languages.

- **Domain**—the class or classes to which the element belongs. The class is the FRBR entity with which the property is associated: “FRBR Manifestation.” Each element is entered into the registry in two forms: one that specifies the domain as defined in RDA, and one that presents the element without a domain designation. This latter can then be used by communities not adhering to FRBR or by those who wish to make a different decision in terms of the binding of elements to FRBR classes.
- **Range**—the value types that can be input as the element contents. Because RDA generally allows both controlled and uncontrolled values, this will be defined most often in the application profile rather than in the element definition.
- **Type**—the type of element, either property or subproperty, class or subclass.
- **subPropertyOf**—for properties that have a hierarchically superordinate property, such as “Variant title,” which is a subproperty of “Title.”
- **hasSubproperty**—For properties with subproperties associated with them, all subproperties are linked to the registry entry for the property. For example, the property “Title” has subproperties “Title proper,” “Key title,” and “Abbreviated title,” among others.

It may seem counterintuitive, but the relationships between FRBR entities are also coded as properties, as

## RDA Glossary

www.rdaonline.org/constituencyreview/  
Phase1Gloss\_10\_21\_08.pdf

```

<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xml:base="http://RDVocab.info/Elements"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dct="http://purl.org/dc/terms/"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:reg="http://metadataregistry.org/uri/profile/RegAp/">
<!-- Element Set: RDA Group 1 Elements -->
<rdf:Description rdf:about="http://RDVocab.info/Elements">
<dc:title xml:lang="en">RDA Group 1 Elements</dc:title>
<skos:note xml:lang="en">This is the provisional registration of the RDA
Group 1 Element Vocabulary, managed by the DCMI/RDA Task Group.</skos:note>
<foaf:homepage rdf:resource="http://dublincore.org/dcmirdataskgroup/">
</rdf:Description>
<!--Property: Title proper-->
<rdf:Description rdf:about="http://RDVocab.info/Elements/titleProper">
<rdfs:isDefinedBy rdf:resource="http://RDVocab.info/Elements" />
<reg:status rdf:resource="http://metadataregistry.org/uri/RegStatus/1002" />
<reg:name xml:lang="en">titleProper</reg:name>
<rdfs:label xml:lang="en">Title proper</rdfs:label>
<rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"
/>
<skos:definition xml:lang="en">The chief name of a resource (i.e., the title
normally used when citing the resource).</skos:definition>
</rdf:Description>
<!-- Status properties used in this document -->
<skos:Concept rdf:about="http://metadataregistry.org/uri/RegStatus/1002">
<skos:prefLabel xml:lang="en">New-Proposed</skos:prefLabel>
</skos:Concept>
</rdf:RDF>

```

**Figure 21**  
Registry entry in machine-readable format.

are the creator and contributor roles. This is an appropriate treatment of relationships in RDF, where all statements are reduced to the subject-predicate-object form (see figure 22).

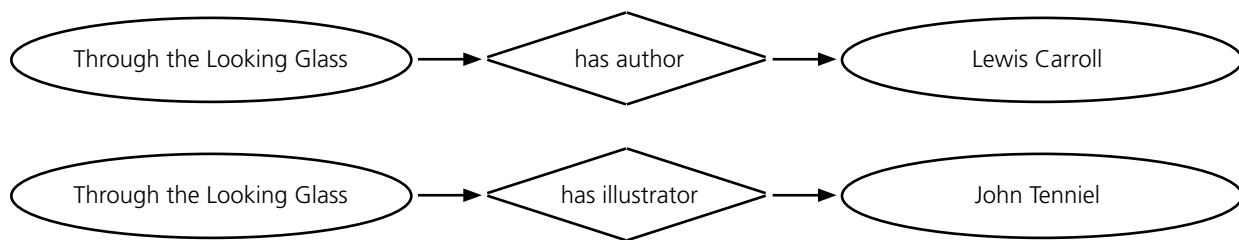
Each of the properties would actually be represented by a URI in a machine-readable triple. This would look something like the diagram in figure 23.

In figure 23, the persons are identified using the identifier for the Library of Congress Name Authority record (although this form of the LCNA is not yet available on the id.loc.gov site). The book is identified using its LC Catalog Number in a standard format provided by LC. This number identifies the manifestation, not the Work, so this illustration is not quite accurate from a strict FRBR point of view, but it satisfies RDF requirements. The relationship property uses the identifiers from the RDA Registry for author and illustrator. As ungainly as the diagram is in this form, this is the preferred way to represent data for applications using Semantic Web technology. Human users of the data should not have to interact with this view, and the data could readily be displayed in any one of many familiar formats:

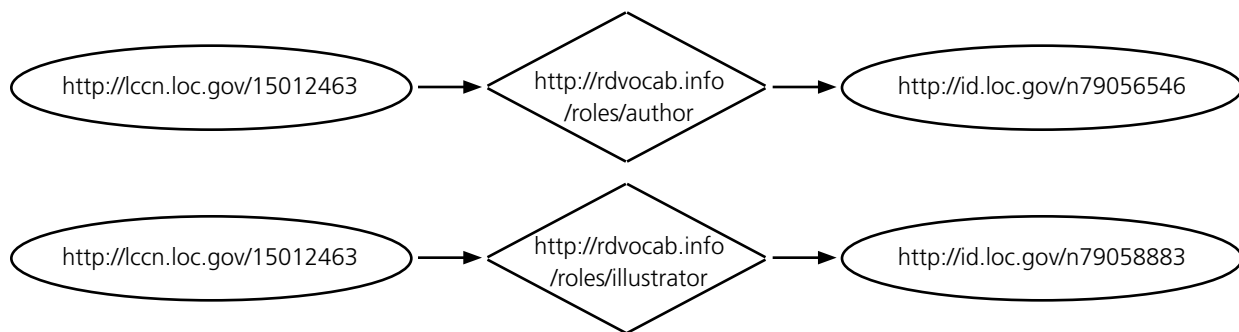
1.  
Through the looking-glass, and what Alice found there  
By Lewis Carroll  
Illustrated by John Tenniel

2.  
Title: Through the looking-glass, and what Alice found there  
Author: Carroll, Lewis  
Illustrator: Tenniel, John

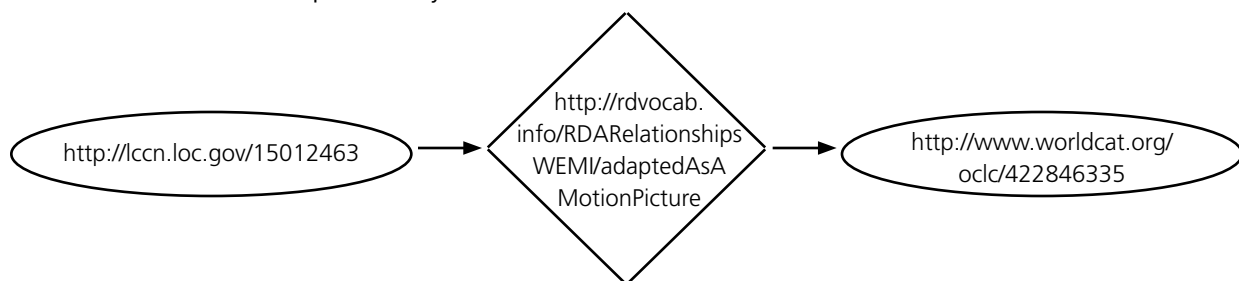
Relationships can be between any entities, such as between Works, between Expressions, or between Persons. These, too, are defined in the registry as properties and can be used in RDF-compatible statements. Figure 24 shows a triple that states that the 1933 film *Alice in Wonderland* was based on the book *Through the Looking-Glass*. The permalink from OCLC WorldCat is used in this case to identify the film.



**Figure 22**  
An author and a contributor, in triple form.



**Figure 23**  
An author and a contributor represented by URIs.



**Figure 24**  
A work/work relationship between the book and the motion picture.

### RDA Value Vocabularies

There are numerous areas in the instructions in the text of RDA where the cataloguer is instructed to make a selection from a limited list of values. These controlled lists are called *vocabularies* in the registry and often referred to as *value vocabularies* in Dublin Core documentation because entries in these lists are used as the value of a property. For example, one would say that the value of a particular instance of RDA Content type is “text,” which is taken from the list of content types defined in RDA. When the value does not come from a value vocabulary, it is simply a character string. When it does come from a value vocabulary and that vocabulary itself has been defined in RDF, the value then has a unique identifier, in this case “<http://RDVocab.info/>

[termList/RDAContentType/1020](http://RDVocab.info/termList/RDAContentType/1020),” which is the URI for the RDA Content Type “text.”

The value vocabularies are defined using the Simple Knowledge Organization System, SKOS, which is an RDF-compliant language specifically designed for term lists and thesauri. As mentioned in chapter 2, SKOS permits the creation of a group of concepts with relationships, such as broader and narrower concepts. Many of the vocabularies in RDA are simple lists of terms, but SKOS allows for the presentation of both preferred and alternate display and entry vocabularies, as well as human-readable definitions of the terms. SKOS can be used to provide vocabularies in more than one language.

An example of a simple list is that for *RDA base material*. This is a list of terms with no broader or narrower relationships (see figure 25).

Vocabulary: Show detail for RDA Base Material				
Detail	Concepts	History	Versions	Maintainers
Preferred Label	URI	Status	Updated	
bristol board	..termList/RDABaseMaterial/1001	New-Proposed	2008-05-23 11:27	
canvas	..termList/RDABaseMaterial/1002	New-Proposed	2008-05-23 11:27	
cardboard	..termList/RDABaseMaterial/1003	New-Proposed	2008-05-23 11:27	
ceramic	..termList/RDABaseMaterial/1004	New-Proposed	2008-05-23 11:27	
glass	..termList/RDABaseMaterial/1005	New-Proposed	2008-05-23 11:27	
hard board	..termList/RDABaseMaterial/1006	New-Proposed	2008-05-29 9:43	
illustration board	..termList/RDABaseMaterial/1007	New-Proposed	2008-05-23 11:28	
ivory	..termList/RDABaseMaterial/1008	New-Proposed	2008-05-23 11:28	
leather	..termList/RDABaseMaterial/1009	New-Proposed	2008-05-23 11:28	
metal	..termList/RDABaseMaterial/1010	New-Proposed	2008-05-23 11:28	
paper	..termList/RDABaseMaterial/1011	New-Proposed	2008-05-23 11:28	
parchment	..termList/RDABaseMaterial/1012	New-Proposed	2008-05-23 11:28	
plaster	..termList/RDABaseMaterial/1013	New-Proposed	2008-05-23 11:28	
plastic	..termList/RDABaseMaterial/1014	New-Proposed	2008-05-23 11:29	
porcelain	..termList/RDABaseMaterial/1015	New-Proposed	2008-05-23 11:29	
shellac	..termList/RDABaseMaterial/1016	New-Proposed	2008-05-23 11:29	
skin	..termList/RDABaseMaterial/1017	New-Proposed	2008-05-23 11:29	
stone	..termList/RDABaseMaterial/1018	New-Proposed	2008-05-23 11:29	
synthetic	..termList/RDABaseMaterial/1019	New-Proposed	2008-05-23 11:29	
textile	..termList/RDABaseMaterial/1020	New-Proposed	2008-05-23 11:30	
24 results				

**Figure 25**  
The registered vocabulary for RDA Base Material.

Vocabulary: Show detail for RDA Standard Combinations of Instruments				
Detail	Concepts	History	Versions	Maintainers
Preferred Label	URI	Status	Updated	
Piano quartet	..termList/StandCombinstr/1010	New-Proposed	2009-03-25 13:13	
Piano quintet	..termList/StandCombinstr/1011	New-Proposed	2009-03-25 13:14	
Piano strings	..termList/StandCombinstr/1004	New-Proposed	2009-03-25 13:14	
Piano trio	..termList/StandCombinstr/1009	New-Proposed	2009-03-25 13:15	
String quartet	..termList/StandCombinstr/1006	New-Proposed	2009-03-25 13:15	
String trio	..termList/StandCombinstr/1005	New-Proposed	2009-03-25 13:16	
Strings	..termList/StandCombinstr/1001	New-Proposed	2009-03-25 13:16	
Wind quintet	..termList/StandCombinstr/1008	New-Proposed	2009-03-25 13:17	
Winds	..termList/StandCombinstr/1003	New-Proposed	2009-03-25 13:17	
Woodwind quartet	..termList/StandCombinstr/1007	New-Proposed	2009-03-25 13:18	
Woodwinds	..termList/StandCombinstr/1002	New-Proposed	2009-03-25 13:18	
11 results				

**Figure 26**  
The registered vocabulary for RDA Standard Combinations of Instruments showing top level terms.

The vocabulary *RDA standard combinations of instruments* does have structure. The top level terms in that structure are noted in figure 26 with a check mark.

The detailed view of a top-level term shows the narrower terms (see figure 27).

Reciprocally, the narrower term's entry records the relationship with the top level term (see figure 28).

RDA defines nearly seventy such vocabulary lists, but this is not by any means an exhaustive treatment of the vocabularies that may be used in bibliographic records. The Library of Congress is working to provide the bibliographic vocabularies under its control in Semantic Web-compatible formats,<sup>5</sup> and the National Library of Medicine has available a version of MeSH in SKOS format.<sup>6</sup> In addition, the library community will make use of standard lists defined by authoritative organizations like the International Standards Organization. Specialist communities from medicine and law to art and architecture often have term lists specific to their interests. Many of these are not yet available in a Semantic Web format, but the trend to provide this data for reuse in Semantic Web environments is beginning.

Vocabulary: RDA Standard Combinations of Instruments			
Concepts: Piano strings			
Detail	Properties	History	
Detail			
Preferred Label:	Piano strings		
Language:	English		
URI:	http://rdvocab.info/termList/StandCombinstr/1004		
Top Concept(?):	<input checked="" type="checkbox"/>		
Status:	New-Proposed		
Properties			
preferred label	Piano strings	English	New-Proposed
has narrower	Piano quartet		New-Proposed
has narrower	Piano quintet		New-Proposed
has narrower	Piano trio		New-Proposed

**Figure 27**  
A detailed view of the term "Piano strings" showing related narrower terms.

Vocabulary: RDA Standard Combinations of Instruments			
Concepts: Piano quintet			
Detail	Properties	History	
Detail			
Preferred Label:	Piano quintet		
Language:	English		
URI:	http://rdvocab.info/termList/StandCombinstr/1011		
Top Concept(?):			
Status:	New-Proposed		
Properties			
preferred label	Piano quintet	English	New-Proposed
definition	Piano, 2 violins, viola, violoncello	English	New-Proposed
has broader	Piano strings		New-Proposed

**Figure 28**  
A view of the term "Piano quintet" with a reference to broader term "Piano strings."

## Maintenance of the Metadata Standard

One of the big issues for any standard is that of maintenance. Maintenance means either constant or periodic revision of the standard to make sure it keeps up to date with the needs of its users. The maintenance activity must also engage the community in decision making and inform all relevant parties of proposed changes and timelines. In the past, the library community has been hindered by very slow update cycles for its standards. Updates to cataloging rules have been years, and even decades, apart, making it impossible for library data creation to keep pace with the rapid evolution of information resources. Bulletins were issued in the time periods between major revisions, but systems were slow to make changes, in part because changes were almost always disruptive in nature.

The definition of elements and vocabularies in a machine-actionable format has the potential to make maintenance of the elements of the cataloging standard easier, faster, and more visible to the community. It also

*RDA/MARC Working Group of the Joint Steering Committee for the Development of RDA*  
www.rda-jsc.org/rdamarcwg.html

could facilitate the update of systems that use the elements and vocabularies.

In the past it was necessary to modify the cataloging standard in order to add new elements or vocabulary terms. This was also true of the standard machine-readable record used by library systems, MARC21. The lengthy process to add a new vocabulary entry to the standard has meant that often years could pass between an initial proposal and the approval of a change. Minor changes, such as adding a new value to an established term list, would go through the same process as major changes, such as adding or modifying significant data elements.

The metadata registry holding the RDA vocabularies has been designed to allow terms and elements to be added on a provisional basis for the purposes of development and testing. Provisional terms are marked so they could be selected for use by systems developers only when they are prepared to perform tests. Having provisional entry of new terms in the standard registry also allows for the time needed for upgrades to user interfaces and training materials.

Note that each value vocabulary could be maintained separately, and changes to one list do not affect other lists or the defined properties. Potentially, maintenance of specialist lists, such as those for music, film, or government documents, could be assigned to the interested community to manage.

With elements and vocabularies in a downloadable machine-readable format, systems can receive changes on a schedule or on an ad hoc basis, as desired. Registry entries can contain the display forms and definitions that will be needed for cataloging functions and used in the user interface so all of the information needed to incorporate a new term into an application is readily available in one place.

Version control is a key element of standards maintenance, and each entry in the registry is given a version stamp. Older versions can be retained, much like older forms of entries are retained commonly in wikis. This allows users to see how a term has changed over time—a feature that is missing in today’s standards and one that makes the combination of current records and older files of records extremely difficult.

## RDA Vocabularies and the Bibliographic Record

The purpose of creating RDF-defined vocabularies is to establish compatibility between applications at a data level rather than at a record level. Among the advantages of well-defined metadata elements is that metadata from different sources and residing in different records can be compatible, even if the record formats themselves are not.

Linked data relies on data in a statement-level format, the *triple*, which serves as a universal microformat that nearly all Semantic Web-compatible applications should be able to provide.

How the data elements are combined into a record format is still up for discussion. The MARC21 community is investigating to what extent RDA can be expressed in that existing format, but it seems clear that the full flexibility and extensibility of RDA goes beyond what can be done in a record format that is already experiencing difficulties in keeping up with needed changes.

There are some (probably valid) assumptions that RDA will be expressed in an XML format. How this will be structured is not known. The eXtensible Catalog project (XC) provides an example of RDF-compliant and FRBR-compliant records. The record examples in figures 29 and 30, received via correspondence with J. Bowen, use only a few RDA vocabulary elements to fill in where Dublin Core, which forms the basis of the XC metadata, is lacking. While not fully expressive of RDA, the XC metadata record does make use of the FRBR Group 1 entities in its record structure, creating separate records for each Group 1 entity, such as these two records for a Work and an Expression.

Note that each of the described entities has a unique identifier and that the two records are linked through the statement in the expression record:

```
<xc:workExpressed>oai:mst.
rochester.edu:MST/
MARCToXCTransformation/10081</
xc:workExpressed>
```

Schematically, this could be diagrammed as a standard RDF triple (see figure 31).

If one wishes to participate in the linked data community, then the data must be expressed as triples rather than XML records. Triples may represent the same data as an XML record, but they don’t constitute a record per se. Triples form a linked set of data that has no defined boundaries. Triples are hard to show because they are not very human-readable. I present them in a somewhat schematic way in figure 32, but remember that each property is either a character string (shown here in quotes) or a URI in URL format.

This “triple” form of RDF statements is awkward from a human standpoint because each statement contains only one relationship. Natural language expresses the same information in a much more compact form, such as “Akira Kurosawa was the director of *Shichinin no samurai* (also known as the *Seven Samurai*), which was adapted as *The Magnificent 7*.” However, the triples logically form a kind of machine-readable sentence, as shown in figure 33.

Both the XML record format and the RDF triple for-

```

<xc:frbr xmlns:xc="http://www.extensiblecatalog.info/Elements"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:rdvocab="http://rdvocab.info/Elements"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:rdarole="http://rdvocab.info/roles">
<xc:entity type="work"
id="oai:mst.rochester.edu:MST/MARCToXCTransformation/10081">
  <dcterms:subject xsi:type="dcterms:LCC">PS3505.U334</dcterms:subject>
  <dcterms:subject xsi:type="dcterms:DDC">811/.52</dcterms:subject>
  <dcterms:subject xsi:type="dcterms:DDC">B</dcterms:subject>
  <rdarole:author>Sawyer-Lauc<U+0327>anno, Christopher, 1951-
</rdarole:author>
  <rdvocab:titleOfWork>E.E. Cummings :</rdvocab:titleOfWork>
  <xc:subject xsi:type="dcterms:LCSH">Cummings, E. E. (Edward Estlin),
1894-1962.</xc:subject>
  <xc:subject xsi:type="dcterms:LCSH">Poets, American-20th century-
Biography.</xc:subject>
</xc:entity>
</xc:frbr>

```

**Figure 29**  
XC XML record for a Work.

```

<?xml version="1.0" encoding="UTF-8"?>
<xc:frbr xmlns:xc="http://www.extensiblecatalog.info/Elements"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:rdvocab="http://rdvocab.info/Elements"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:rdarole="http://rdvocab.info/roles">
<xc:entity type="expression"
id="oai:mst.rochester.edu:MST/MARCToXCTransformation/10082">
  <xc:titleOfExpression>E.E. Cummings :</xc:titleOfExpression>
  <rdvocab:illustrationContent>ill. ;</rdvocab:illustrationContent>
  <xc:workExpressed>oai:mst.rochester.edu:MST/MARCToXCTransformation/1008
1</xc:workExpressed>
</xc:entity>
</xc:frbr>
<?xml version="1.0" encoding="UTF-8"?>

```

**Figure 30**  
XC XML record for an Expression.

mat of data are valid to use. The record format creates a kind of container that can keep one set of data elements together for an application's purposes. The triple format allows the individual statements within the data to interact with other statements and form a constantly growing web of information.

One of the big questions relating to the creation of RDF data is how identifiers will be created for all of the metadata instances created in libraries. In some ideal universe where everything is perfectly neat—obviously not the one we occupy—there would be a single, universal identifier for each Work, each Person, each Place, and so on. This is unlikely to happen, although any sharing of identifiers increases the interoperability of data. The reality will

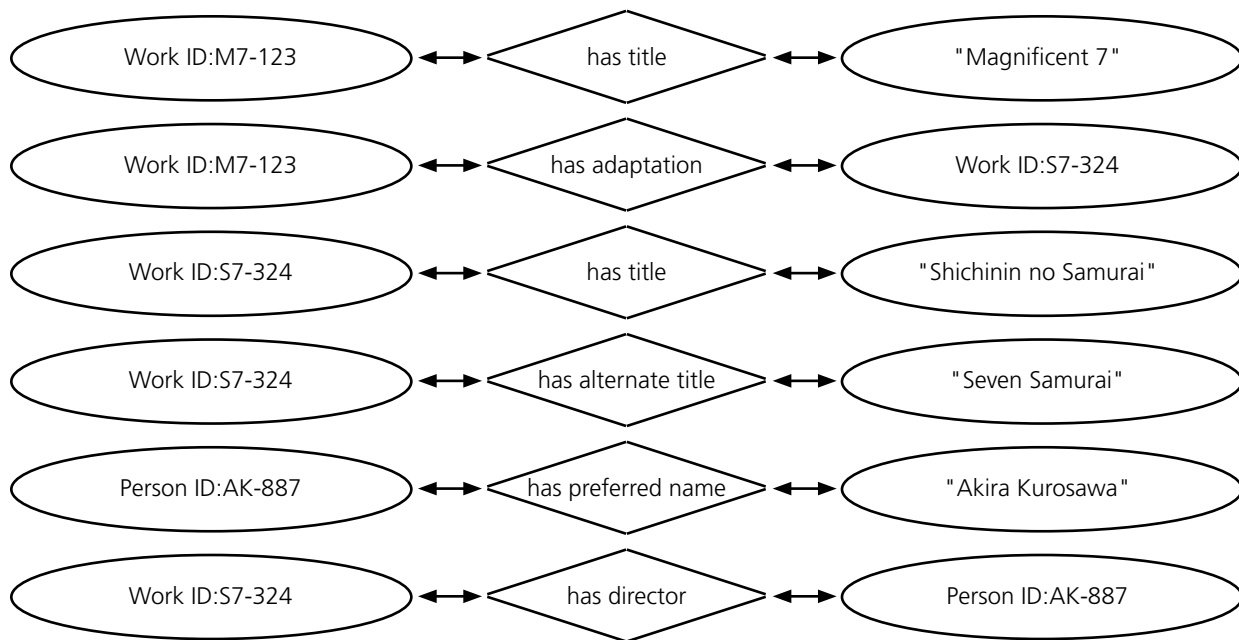
undoubtedly be that, as in the examples above, some if not all of the identifiers assigned will be only locally meaningful. There could be aggregation services that perform a similar matching that OCLC provides for library MARC records, bringing together data from different systems and associating that data with a shared identity. For this, the bibliographic data itself will be used, as it is today, to infer that two separately created bibliographic descriptions are describing the same bibliographic resource.

With data in a Web-compatible format, there is also the possibility of creating Web-based data-creation tools, with broad sharing of identified elements such as Works, Persons, and Places as well as relationships. The more that identifiers are shared, the more accurate any state-





**Figure 31**  
RDF triple of the Expression-to-Work relationships.



**Figure 32**  
Complex set of triples about "Magnificent 7."

ment that "A is the same as B" can be, whether that is for a Work, a Person, a Place, or any other instance of an entity or property.

## Application Profiles

The over six hundred pages of the *Anglo-American Cataloguing Rules* (2nd edition) and the many hundreds of properties defined for the new cataloging rules, Resource Description and Access (RDA), are all the proof we need that the library cataloging rules attempt to cover the widest possible range of cataloging situations. Perhaps only the largest and most varied of libraries will have a need for all of the rules and data elements, and in fact, studies of MARC data show that the majority of data elements defined in that standard are seldom used out of a body of millions of cataloging examples.<sup>7</sup>

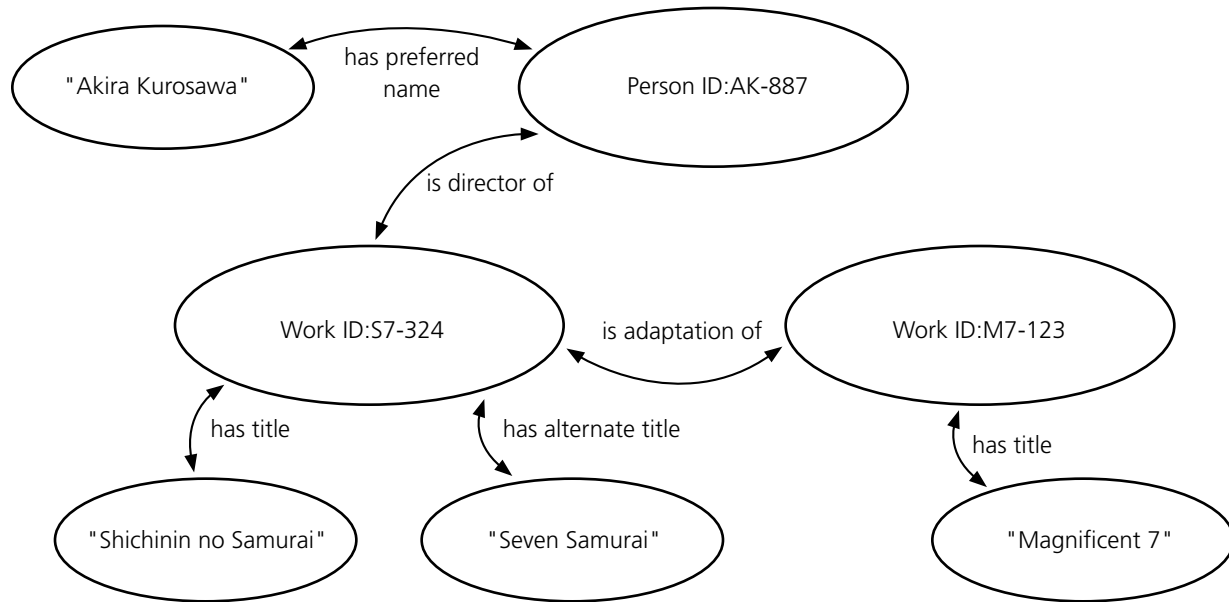
Libraries often find a need to create custom versions of the cataloging rules that are tailored to their specific needs. The RDA Online product being prepared by the publishing office of the American Library Association includes a customization function called "workflows" precisely because of this need. These workflows allow one

to select from the RDA chapters and sections that are pertinent to the library's cataloging activity.

The information technology world has a similar concept for the customization of applications call "application profiles" (APs). Application profiles are a selection of data elements from a larger universe. The Dublin Core Metadata Initiative has developed an RDF-compliant, machine-readable expression of application profiles. Called the Dublin Core Description Set Profile (DSP), it provides a standard format that facilitates the creation of applications from the selected data elements.<sup>8</sup> The AP consists of a selection of RDF-compliant elements, and a definition of constraints related to those properties. Constraints consist of the declaration of repeatability, whether the elements are mandatory or optional, and any requirements for the types of values the elements will allow (plain text, controlled vocabularies, and so on).<sup>9</sup>

In an RDF-compliant application profile, elements and vocabularies can be taken from any suitable defined

*eXtensible Catalog: Metadata*  
[www.extensiblecatalog.org/Metadata](http://www.extensiblecatalog.org/Metadata)



**Figure 33**  
 "Akira Kurosawa was the director of *Shichinin no samurai* (also known as the *Seven Samurai*), which was adapted as *The Magnificent 7*."

set, and many Semantic Web applications work with a mix of elements from numerous sources. There is a conscious effort in that community to reuse rather than reinvent as part of the goal of interoperability over the entire Web. An application profile would therefore describe the particular mix of elements that had been chosen for a particular application.

Library community members could create any number of application profiles to meet their needs. There could be profiles for specialist communities, like visual resources or law collections. There could be profiles based on the languages of the collection that don't include rules for languages not needed. There could be simplified rules for minimal cataloging. The key, however, is that all of these customized profiles would be compatible with each other because they all would make use of the same defined and registered metadata properties. Undoubtedly some core properties will be used by all or at least most of the profiles, while other, more specialized properties will be needed only by a few members of the community.

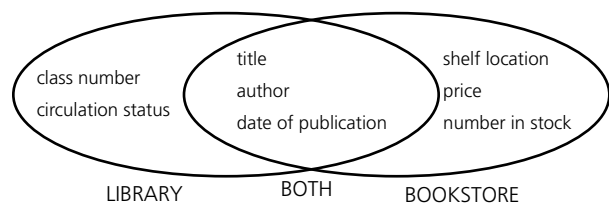
*RDA Online*  
<http://rdaonline.org>

While RDA intends to be as complete a set of metadata as possible, the adoption of application profiles would allow any community that wishes to use RDA to extend the vocabulary for local or specialist needs. It would no

longer be necessary to entirely recreate a metadata set if RDA is found to only partially fulfill an institution's needs. Application profiles are the technical mechanism that support the data sharing that was introduced at the beginning of this chapter (see figure 34).

## A Word about the Future

RDF is not a magic spell that will make library data perfect. It is today's technology wave, arising out of the current capabilities of networked information resources. It will, somewhere down the line, be replaced by another technology. Where RDF differs most from the present system of bibliographic records is in allowing bibliographic descriptions to interact, extend, and influence each other and to interact at a statement level with other data from



**Figure 34**  
 In application profiles, differences are accommodated with disrupting the advantages of shared data.

library and nonlibrary sources. The advantages to the library community are unmistakable.

It may be useful here to remember that when the MARC record was first developed, it was intended solely as a better way to issue printed card sets from the Library of Congress. Yet the machine-readable format made possible the creation of online catalogs, something that previously had been unthinkable. We cannot know today what innovations could be fostered through the transformation of library data to a new technology, but the possibilities are intriguing, not so much for how this could change the act of cataloging but for the new user services that could be built with a more flexible data carrier.

## Notes

1. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, Sept. 1997, as amended and corrected through Feb. 2009, [http://archive.ifa.org/VII/s13/frbr/frbr\\_2008.pdf](http://archive.ifa.org/VII/s13/frbr/frbr_2008.pdf) (accessed Dec. 14, 2009); Joint Steering Committee for Development of RDA, "RDA: Resource Description and Access," [www.rda-jsc.org/rda.html](http://www.rda-jsc.org/rda.html) (accessed Dec. 14, 2009).
2. American Library Association, *Anglo-American Cataloguing Rules*, 2nd ed. (London: Library Association, 1978).
3. "RDA—Resource Description and Access: A Prospectus," rev. 7, July 1, 2009, p. 2, Joint Steering Committee for the Development of RDA website, [www.rda-jsc.org/docs/5rda-prospectusrev7.pdf](http://www.rda-jsc.org/docs/5rda-prospectusrev7.pdf) (accessed Dec. 18, 2009).
4. "RDA Element Analysis," rev. 2, Oct. 26, 2008, Joint Steering Committee for the Development of RDA website, [www.rda-jsc.org/docs/5rda-elementanalysisrev2.pdf](http://www.rda-jsc.org/docs/5rda-elementanalysisrev2.pdf) (accessed Dec. 14, 2009).
5. See "Authorities and Vocabularies" on the Library of Congress website, <http://id.loc.gov/authorities>.
6. "Neurocommons Alpha" is available online to those with a username and password at <http://sw.neurocommons.org/2007/kb-sources/medline/medline-mesh.tgz>.
7. MARC Content Designation Utilization: Inquiry and Analysis blog, [www.mcd.unt.edu](http://www.mcd.unt.edu) (accessed Dec. 14, 2009).
8. Mikael Nilsson, "Description Set Profiles: A Constraint Language for Dublin Core Application Profiles," March 31, 2008, <http://dublincore.org/documents/2008/03/31/dc-dsp> (accessed Dec. 14, 2009).
9. Karen Coyle and Thomas Baker, "Guidelines for Dublin Core Application Profiles," May 18, 2009, <http://dublincore.org/documents/profile-guidelines> (accessed Dec. 14, 2009).

---

## Metadata Models of the World Wide Web, continued from page 19

### Notes

1. Tim Berners-Lee, James Hendler, and Ora Lassila, "The Semantic Web," *Scientific American*, May 2001: 34–43.
2. Tim Berners-Lee, "Linked Data," July 27, 2006, last updated June 18, 2009, World Wide Web Consortium (W3C) website, [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html) (accessed Dec. 16, 2009).
3. "DCMI Metadata Terms," Jan. 14, 2008, on the Dublin Core Metadata Initiative website, <http://dublincore.org/documents/dcmi-terms> (accessed Dec. 16, 2009).
4. Mikael Nilsson, Thomas Baker, and Pete Johnston, "The Singapore Framework for Dublin Core Application Profiles," Jan. 14, 2008, on the Dublin Core Metadata Initiative website, <http://dublincore.org/documents/singapore-framework> (accessed Dec. 16, 2009).