Title
Can Bibliographic Data Be Put Directly Onto the Semantic Web?

Journal
Information Technology and Libraries, 28(2)

Author
Yee, Martha M

Publication Date
2009-06-01

Peer reviewed

Can Bibliographic Data be Put Directly onto the Semantic Web?

Martha M. Yee

This paper is a think piece about the possible future of bibliographic control; it provides a brief introduction to the Semantic Web and defines related terms, and it discusses granularity and structure issues and the lack of standards for the efficient display and indexing of bibliographic data. It is also a report on a work in progress—an experiment in building a Resource Description Framework (RDF) model of more FRBRized cataloging rules than those about to be introduced to the library community (Resource Description and Access) and in creating an RDF data model for the rules. I am now in the process of trying to model my cataloging rules in the form of an RDF model, which can also be inspected at http://myee.bol.ucla.edu/. In the process of doing this, I have discovered a number of areas in which I am not sure that RDF is sophisticated enough yet to deal with our data. This article is an attempt to identify some of those areas and explore whether or not the problems I have encountered are soluble—in other words, whether or not our data might be able to live on the Semantic Web. In this paper, I am focusing on raising the questions about the suitability of RDF to our data that have come up in the course of my work.

This paper is a think piece about the possible future of bibliographic control; as such, it raises more complex questions than it answers. It is also a report on a work in progress—an experiment in building a Resource Description Framework (RDF) model of FRBRized descriptive and subject-cataloging rules. Here my focus will be on the data model rather than on the FRBRized cataloging rules for gathering data to put in the model, although I hope to have more to say about the latter in the future. The intent is not to present you with conclusions but to present some questions about data modeling that have arisen in the course of the experiment. My premise is that decisions about the data model we follow in the future should be made openly and as a community rather than in a small, closed group of insiders. If we are to move toward the creation of metadata that is more interoperable with metadata being created outside our community, as is called for by many in our profession, we will need to address these complex questions as a community following a period of deep thinking, clever experimentation, and astute political strategizing.

Martha M. Yee (myee@ucla.edu) is Cataloging Supervisor at the University of California, Los Angeles Film and Television Archive.

The vision

The Semantic Web is still a bewitching midsummer night's dream. It is the idea that we might be able to replace the existing HTML–based Web consisting of marked-up documents—or pages—with a new RDF–based Web consisting of data encoded as classes, class properties, and class relationships (semantic linkages), allowing the Web to become a huge shared database. Some call this Web 3.0, with hyperdata replacing hypertext. Embracing the Semantic Web might allow us to do a better job of integrating our content and services with the wider Internet, thereby satisfying the desire for greater data interoperability that seems to be widespread in our field. It also might free our data from the proprietary prisons in which it is currently held and allow us to cooperate in developing open-source software to index and display the data in much better ways than we have managed to achieve so far in vendor-developed ILS OPACs or in giant, bureaucratic bibliographic empires such as OCLC WorldCat.

The Semantic Web also holds the promise of allowing us to make our work more efficient. In this bewitching vision, we would share in the creation of Uniform Resource Identifiers (URIs) for works, expressions, manifestations, persons, corporate bodies, places, subjects, and so on. At the URI would be found all of the data about that entity, including the preferred name and the variant names, but also including much more data about the entity than we currently put into our work (*name-title* and *title*), such as *personal name, corporate name, geographic,* and *subject authority records*. If any of that data needed to be changed, it would be changed only once, and the change would be immediately accessible to all users, libraries, and library staff by means of links down to local data such as circulation, acquisitions, and binding data. Each work would need to be described only once at one URI, each expression would need to be described only once at one URI, and so forth.

Very much up in the air is the question of what institutional structures would support the sharing of the creation of URIs for entities on the Semantic Web. For the data to be reliable, we would need to have a way to ensure that the system would be under the control of people who had been educated about the value of clean and accurate entity definition, the value of choosing "most commonly known" preferred forms (for display in lists of multiple different entities), and the value of providing access

under all variant forms likely to be sought. At the same time, we would need a mechanism to ensure that any interested members of the public could contribute to the effort of gathering variants or correcting entity definitions when we have had inadequate information. For example, it would be very valuable to have the input of a textual or descriptive bibliographer applied to difficult questions concerning particular editions, issues, and states of a significant literary work. It would also be very valuable to be able to solicit input from a subject expert in determining the bounds of a concept entity (subject heading) or class entity (classification).

The experiment (my project)

To explore these bewitching ideas, I have been conducting an experiment. As part of my experiment, I designed a set of cataloging rules that are more FRBRized than is RDA in the sense that they more clearly differentiate between data applying to expression and data applying to manifestation. Note that there is an underlying assumption in both FRBR (which defines expression quite differently from manifestation) and on my part, namely that catalogers always know whether a given piece of data applies at either the expression or the manifestation level. That assumption is open to questioning in the process of the experiment as well. My rules also call for creating a more hierarchical and degressive relationship between the FRBR entities *work*, *expression*, *manifestation*, and *item*, such that data pertaining to the work does not need to be repeated for every expression, data pertaining to the expression does not need to be repeated for every manifestation, and so forth. *Degressive* is an old term used by bibliographers for bibliographies that provide great detail about first editions and less detail for editions after the first. I have adapted this term to characterize my rules, according to which the cataloger begins by describing the work; any details that pertain to all expressions and manifestations of the work are not repeated in the expression and manifestation descriptions. This paper would be entirely too long if I spent any more time describing the rules I am developing, which can be inspected at http://myee.bol.ucla.edu. Here, I would like to focus on the data-modeling process and the questions about the suitability of RDF and the Semantic Web for encoding our data. (By the way, I don't seriously expect anyone to adopt my rules! They are radically different than the rules currently being applied and would represent a revolution in cataloging practice that we may not be up to undertaking in the current economic climate. Their value lies in their thought-experiment aspect and their ability to clarify what entities we can model and what entities we may not be able to model.)

I am now in the process of trying to model my cataloging rules in the form of an RDF model ("RDF" as used in this paper should be considered from now on to encompass RDF Schema [RDFS], Web Ontology Language [OWL], and Simple Knowledge Organization System [SKOS] unless otherwise stated); this model can also be inspected at http://myee.bol.ucla.edu. In the process of doing this, I have discovered a number of areas in which I am not sure that RDF is yet sophisticated enough to deal with our data. This article is an attempt to outline some of those areas and explore whether the problems I have encountered are soluble, in other words, whether or not our data might be able to live on the Semantic Web eventually. I have already heard from RDF experts Bruce D'Arcus (Miami University) and Rob Styles (developer of Talis, a Semantic Web technology company), whom I cite later, but through this article I hope to reach a larger community.

My research questions can be found later, but first some definitions.

Definition of terms

The *Semantic Web* is a way to represent knowledge; it is a knowledge-representation language that provides ways of expressing meaning that are amenable to computation; it is also a means of constructing knowledge-domain maps consisting of class and property axioms with a formal semantics.

RDF is a family of specifications for methods of modeling information that underpins the Semantic Web through a variety of syntax formats; an RDF metadata model is based on making statements about resources in the form of triples that consist of

1. the *subject* of the triple (e.g., "New York");
2. the *predicate* of the triple that links the subject and the object (e.g., "has the postal abbreviation"); and
3. the *object* of the triple (e.g., "NY").

XML is commonly used to express RDF, but it is not a necessity; it can also be expressed in Notation 3 or N3, for example.[1]

RDFS is an extensible knowledge-representation language that provides basic elements for the description of ontologies, also known as *RDF vocabularies*. Using RDFS, statements are made about resources in the form of

1. a *class* (or *entity*) as subject of the RDF triple (e.g., "New York");
2. a *relationship* (or *semantic linkage*) as predicate of the RDF triple that links the subject and the object (e.g.,

"has the postal abbreviation"); and
3. a *property* (or *attribute*) as object of the RDF triple (e.g., "NY").

OWL is a family of knowledge representation languages for authoring ontologies compatible with RDF.

SKOS is a family of formal languages built upon RDF and designed for representation of thesauri, classification schemes, taxonomies, or subject-heading systems.

Research questions

Actually, the full-blown Semantic Web may not be exactly what we need. Remember that the fundamental definition of the Semantic Web is "a way to represent knowledge." The Semantic Web is a direct descendant of the attempt to create artificial intelligence, that is, of the attempt to encode enough knowledge of the real world to allow a computer to reason about reality in a way indistinguishable from the way a human being reasons. One of the research questions should probably be whether or not the technology developed to support the Semantic Web can be used to represent information rather than knowledge. Fortunately, we do not need to represent all of human knowledge—we simply need to describe and index resources to facilitate their retrieval. We need to encode facts about the resources and what the resources discuss (what they are "about"), not facts about "reality." Based on our past experience, doing even this is not as simple as people think it is. The question is whether we could do what we need to do within the context of the Semantic Web. Sometimes things that sound simple do not turn out to be so simple in the doing.

My research questions are as follows:

1. Is it possible for catalogers to tell in all cases whether a piece of data pertains to the FRBR expression or the FRBR manifestation?
2. Is it possible to fit our data into RDF? Given that RDF was designed to encode knowledge rather than information, perhaps it is the wrong technology to use for our purposes?
3. If it is possible to fit our data into RDF, is it possible to use that data to design indexes and displays that meet the objectives of the catalog (i.e., providing an efficient instrument to allow a user to find a particular work of which the author and title are known, a particular expression of a work, all of the works of an author, all of the works in a given genre or form, or all of the works on a particular subject)?

As stated previously, I am not yet ready to answer these questions. I hope to find answers in the course of developing the rules and the model. In this paper, I am focusing on raising the questions about the suitability of RDF to our data that have come up in the course of my work.

Other relevant projects

Other relevant projects include the following:

1. *FRBR,* Functional Requirements for Authority Data (FRAD), Functional Requirements for Subject Authority Records (FRSAR), and FRBR-object-oriented (FRBRoo). All are attempts to create conceptual models of bibliographic entities using an entity-relationship model that is very similar to the class-property model used by RDF.[2]
2. *Various initiatives at the Library of Congress (LC),* such as LC Subject Headings (LCSH) in SKOS,[3] the LC Name Authority File in SKOS,[4] the LCCN Permalink project to create persistent URIs for bibliographic records,[5] and initiatives to provide SKOS representations for vocabularies and data elements used in MARC, PREMIS, and METS. These all represent attempts to convert our existing bibliographic data into URIs that stand for the bibliographic entities represented by bibliographic records and authority records; the URIs would then be available for experiments in putting our data directly onto the Semantic Web.
3. *The DC-RDA Task Group project to put RDA data elements into RDF.*[6] As noted previously and discussed further later, RDA is less FRBRized than my cataloging rules, but otherwise this project is very similar to mine.
4. *Dublin Core's (DC's) work on an RDF schema.*[7] Dublin Core is very focused on manifestation and does not deal with expressions and works, so it is less similar to my project than is the DC-RDA Task Group's project (see further discussion later).

Why my project?

One might legitimately ask why there is a need for a different model than the ones already provided by FRBR, FRAD, FRSAR, FRBRoo, RDA, and DC. The FRBR and RDA models are still tied to the model that is implicit in our current bibliographic data in which expression and manifestation are undifferentiated. This is because publishers publish and libraries acquire and shelve manifestations. In our current bibliographic practice, a new

bibliographic record is made for either a new manifestation or a new expression. Thus, in effect, there is no way for a computer to tell one from the other in our current data. Despite the fact that FRBR has good definitions of expression (change in content) and manifestation (mere change in carrier), it perpetuates the existing implicit model in its mapping of attributes to entities. For example, FRBR maps the following to manifestation: edition statements ("2nd rev. ed."); statements of responsibility that identify translators, editors, and illustrators; physical description statements that identify illustrated editions; and extent statements that differentiate expressions (the 102-minute version vs. the 89-minute version); etc. Thus the FRBR definition of expression recognizes that a 2nd revised edition is a new expression, but FRBR maps the edition statement to manifestation. In my model, I have tried to differentiate more cleanly data applying to expressions from data applying to manifestations.[8]

FRBR and RDA tend to assume that our current bibliographic data elements map to one and only one group 1 entity or class. There are exceptions, such as title, which FRBR and RDA define at work, expression, and manifestation levels. However, there is a lack of recognition that, to create an accurate model of the bibliographic universe, more data elements need to be applied at the work and expression level in addition to (or even instead of) the manifestation level. In the appendix I have tried to contrast the FRBR, FRAD, and RDA models with mine. In my model, many more data elements (properties and attributes) are linked to the work and expression level. After all, if the expression entity is defined as any change in work content, the work entity needs to be associated with all content elements that might change, such as the original extent of the work, the original statement of responsibility, whether illustrations were originally present, whether color was originally present in a visual work, whether sound was originally present in an audiovisual work, the original aspect ratio of a moving image work, and so on.

FRBR also tends to assume that our current data elements map to one and only one entity. In working on my model, I have come to the conclusion that this is not necessarily true. In some cases, a data element pertaining to a manifestation also pertains to the expression and the work. In other cases, the same data element is specific to that manifestation, and, in other cases, the same data element is specific to its expression. This is true of most of the elements of the bibliographic description.

FRAD, in attempting to deal with the fact that our current cataloging rules allow a single person to have several bibliographic identities (or pseudonyms), treats *person*, *name*, and *controlled access point* as three separate entities or classes. I have tried to keep my model simpler and more elegant by treating only *person* as an entity, with *preferred name* and *variant name* as attributes or properties of that entity.

FRBRoo is focused on the creation process for works, with special attention to the creation of unique works of art and other one-off items found in museums. Thus FRBRoo tends to neglect the collocation of the various expressions that develop in the history of a work that is reproduced and published, such as translations, abridged editions, editions with commentary, etc.

DC has concentrated exclusively on the description of manifestations and has neglected expression and work altogether.

One of the tenets of Semantic Web development is that, once an entity is defined by a community, other communities can reuse that entity without defining it themselves. The very different definitions of the work and expression entities in the different communities described above raise some serious questions about the viability of this tenet.

Assumptions

It should be noted that this entire experiment is based on two assumptions about the future of human intervention for information organization. These two assumptions are based on the even bigger assumption that, even though the Internet seems to be an economy based on free intellectual labor, and, even though human intervention for information organization is expensive (and therefore at more risk than ever), human intervention for information organization is worth the expense.

- *Assumption 1*: What we need is not artificial intelligence, but a better human–machine partnership such that humans can do all of the intellectual labor and machines can do all of the repetitive clerical labor. Currently, catalogers spend too much time on the latter because of the poor design of current systems for inputting data. The universal employment provided by paying humans to do the intellectual labor of building the Semantic Web might be just the stimulus our economy needs.
- *Assumption 2*: Those who need structured and granular data—and the precise retrieval that results from it—to carry out research and scholarship may constitute an elite minority rather than most of the people of the world (sadly), but that talented and intelligent minority is an important one for the cultural and technological advancement of humanity. It is even possible that, if we did a better job of providing access to such data, we might enable the enlargement of that minority.

Granularity and structure issues

As soon as one starts to create a data model, one encounters granularity or cataloger-data parsing issues. These issues have actually been with us all along as we developed the data model implicit in AACR2R and MARC 21. Those familiar with RDA, FRBR, and FRAD development will recognize that much of that development is directed at increasing structure and granularity in cataloger-produced data to prepare for moving it onto the Semantic Web. However, there are clear trade-offs in an increase in structure and granularity. More structure and more granularity make possible more powerful indexing and more sophisticated display, but more structure and more granularity are more complex and expensive to apply and less likely to be implemented in a standard fashion across all communities; that is, it is less likely that interoperable data would be produced. Any switching or mapping that was employed to create interoperable data would produce the lowest common denominator (the simplest and least granular data), and once rendered interoperable, it would not be possible for that data to swim back upstream to regain its lost granularity. Data with less structure and less granularity could be easier and cheaper to apply and might have the potential to be adopted in a more standard fashion across all communities, but that data would limit the degree to which powerful indexing and sophisticated display would be possible.

Take the example of a personal name: Currently, we demarcate surname from forename by putting the surname first, followed by a comma and then the forename. Even that amount of granularity can sometimes pose a problem for a cataloger who does not necessarily know which part of the name is surname and which part is forename in a culture unfamiliar to the cataloger. In other words, the more granularity you desire in your data, the more often the people collecting the data are going to encounter ambiguous situations. Another example: Currently, we do not collect information about gender self-identification; if we were to increase the granularity of our data to gather that information, we would surely encounter situations in which the cataloger would not necessarily know if a given creator was self-defined as a female or a male or of some other gender identity.

Presently, if we are adding a birth and death date, whatever dates we use are all together in a *$d* subfield without any separate coding to indicate which date is the birth date and which is the death date (although an occasional "b." or "d." will tell us this kind of information). We could certainly provide more granularity for dates, but that would make the MARC 21 format much more complex and difficult to learn. People who dislike the MARC 21 format already argue that it is too granular and therefore requires too much of a learning curve before people can use it. For example, Tennant claims that "there are only two kinds of people who believe themselves able to read a MARC record without referring to a stack of manuals: a handful of our top catalogers and those on serious drugs."[9] How much of the granularity already in MARC 21 is used either in existing records or, even if present, is used in indexing and display software? Granularity costs money, and libraries and archives are already starving for resources. Granularity can only be provided by people, and people are expensive.

Granularity and structure also exist in tension with each other. More granularity can lead to less structure (or more complexity to retain structure along with granularity). In the pursuit of more granularity of data than we have now, RDA, attempting to support RDF–compliant XML encoding, has been atomizing data to make it useful to computers, but this will not necessarily make the data more useful to humans. To be useful to humans, it must be possible to group and arrange (sort) the data meaningfully, both for indexing and for display. The developers of SKOS refer to the "vast amounts of unstructured (i.e., human readable) information in the web,"[10] yet labeling bits of data as to type and recording semantic relationships in a machine-actionable way do not necessarily provide the kind of structure necessary to make data readable by humans and therefore useful to the people the Web is ultimately supposed to serve. Consider the case of music instrumentation. If you have a piece of music for five guitars and one flute, and you simply code number and instrumentation without any way to link "five" with "guitars" and "one" with "flute," you will not be able to guarantee that a person looking for music for five flutes and one guitar will not be given this piece of music in their results (see figure 1).[11] The more granular the data, the less the cataloger can build order, sequencing, and linking into the data; the coding must be carefully designed to allow the desired order, sequencing, and linking for indexing and display to be possible, which might call for even more complex coding. It would be easy to lose information about order, sequencing, and linking inadvertently.

Actually, there are several different meanings for the term *structure*:

1. Structure is an *object of a record* (structure of document?); for example, Elings and Waibel refer to "data fields . . . also referred to as elements . . . which are organized into a record by a data structure."[12]
2. Structure is the *communications layer*, as opposed to the display layer or content designation.[13]
3. Structure is the *record, field, and subfield.*
4. Structure is the *linking of bits of data together in the*

form of various types of relationships.

5. Structure is the *display of data in a structured, ordered, and sequenced manner to facilitate human understanding.*
6. Data structure is a *way of storing data in a computer so that it can be used efficiently* (this is how computer programmers use the term).

I hasten to add that I am definitely in favor of adding more structure and granularity to our data when it is necessary to carry out the fundamental objectives of our profession and of our catalogs. I argued earlier that FRBR and RDA are not granular enough when it comes to the distinction between data elements that apply to expression and those that apply to manifestation. If we could just agree on how to differentiate data applying to the manifestation from data applying to the expression instead of our current practice of identifying works with headings and lumping all manifestation and expression data together, we could increase the level of service we are able to provide to users a thousandfold. However, if we are not going to commit to differentiating between

```
<rdfs:Property rdf:about="http://myee.bol.ucla.edu/ycrschema/elements/1.0/expinstr"/>
<rdfs:isDefinedBy rdf:resource="http://myee.bol.ucla.edu/ycrschema/elements/1.0/"/>
<rdfs:label xml:lang="en">instrumentation of musical expression</rdfs:label>
<rdfs:domain rdf:resource="http://myee.bol.ucla.edu/ycrschema#Expression"/>
<rdfs:range rdf:resource="www.w3.org/TR/rdf-schema#Literal"/>
<rdfs:subPropertyOf rdf:resource="http://myee.bol.ucla.edu/ycrschema#expdesc"/>

<rdfs:Property rdf:about="http://myee.bol.ucla.edu/ycrschema/elements/1.0/expinstrnumber"/>
<rdfs:isDefinedBy rdf:resource="http://myee.bol.ucla.edu/ycrschema/elements/1.0/"/>
<rdfs:label xml:lang="en">original instrumentation of musical expression—number of a particular instrument</
rdfs:label>
<rdfs:domain rdf:resource="http://myee.bol.ucla.edu/ycrschema#Expression"/>
<rdfs:range rdf:resource="www.w3.org/TR/rdf-schema#Literal"/>
<rdfs:subPropertyOf rdf:resource="http://myee.bol.ucla.edu/ycrschema#expinstr"/>

<rdfs:Property rdf:about="http://myee.bol.ucla.edu/ycrschema/elements/1.0/expinstrtype"/>
<rdfs:isDefinedBy rdf:resource="http://myee.bol.ucla.edu/ycrschema/elements/1.0/"/>
<rdfs:label xml:lang="en">original instrumentation of musical expression—type of instrument</rdfs:label>
<rdfs:domain rdf:resource="http://myee.bol.ucla.edu/ycrschema#Expression"/>
<rdfs:range rdf:resource="www.w3.org/TR/rdf-schema#Literal"/>
<rdfs:subPropertyOf rdf:resource="http://myee.bol.ucla.edu/ycrschema#expinstr"/>
```

Figure 1a. Extract from Yee RDF model that illustrates one technique for modeling musical instrumentation at the expression level (using a blank node to group repeated number and instrument type)

```
<ycr:expinstr>

<ycr:expinstrnumber>5</ycr:expinstrnumber>
<ycr:expinstrtype>guitars</ycr:expinstrtype>

</ycr:expinstr>

<ycr:expinstr>

<ycr:expinstrnumber>1</ycr:expinstrnumber>
<ycr:expinstrtype>flute</ycr:expinstrtype>

</ycr:expinstr>
```

Figure 1b. Example of encoding of musical instrumentation at the expression level based on the above model

expression and manifestation, it would be more intellectually honest for FRBR and RDA to take the less granular path of mapping all existing bibliographic data to manifestation and expression undifferentiated, that is, to use our current data model unchanged and state this openly. I am not in favor of adding granularity for granularity's sake or for the sake of vague conceptions of possible future use. Granularity is expensive and should be used only in support of clear and fundamental objectives.

The goal: efficient displays and indexes

My main concern is that we model and then structure the data in a way that allows us to build the complex displays that are necessary to make catalogs appear simple to use. I am aware that the current orthodoxy is that recording data should be kept completely separate from indexing and display ("the applications layer"). Because I have spent my career in a field in which catalog records are indexed and displayed badly by systems people who don't seem to understand the data contained in them, I am a skeptic. It is definitely possible to model and structure data in such a way that desired displays and indexes are impossible to construct. I have seen it happen!

The LC Working Group report states that "it will be recognized that human users and their needs for display and discovery do not represent the only use of bibliographic metadata; instead, to an increasing degree, machine applications are their primary users."[14] My fear is that the underlying assumption here is that users need to (and can) retrieve the single perfect record. This will never be true for bibliographic metadata. Users will always need to assemble all relevant records (of all kinds) as precisely as possible and then browse through them before making a decision about which resources to obtain. This is as true in the Semantic Web—where "records" can be conceived of as entity or class URIs—as it is in the world of MARC-encoded metadata.

Some of the problems that have arisen in the past in trying to index bibliographic metadata for humans are connected to the fact that existing systems do not group all of the data related to a particular entity effectively, such that a user can use any variant name or any combination of variant names for an entity and do a successful search. Currently, you can only look for a match among two or more keywords within the bounds of a single manifestation-based bibliographic record or within the bounds of a single heading, minus any variant terms for that entity. Thus, when you do a keyword search for two keywords, for example, "clemens" and "adventures," you will retrieve only those manifestations of Mark Twain's *Adventures of Tom Sawyer* that have his real name (Clemens) and the title word "Adventures" co-occurring within the bounded space created by a single manifestation-based bibliographic record. Instead, the preferred forms and the variant forms for a given entity need to be bounded for indexing such that the keywords the user employs to search for that entity can be matched using co-occurrence rules that look for matches within a single bounded space representing the entity desired. We will return to this problem in the discussion of issue 3 in the later section "RDF Problems Encountered."

The most complex indexing problem has always proven to be the grouping or bounding of data related to a work, since it requires pulling in all variants for the creator(s) of that work as well. Otherwise, a user who searches for a work using a variant of the author's name and a variant of the title will continue to fail (as they do in all current OPACs), even when the desired work exists in the catalog. If we could create a URI for the *Adventures of Tom Sawyer* that included all variant names for the author and all variant titles for the work (including the variant title *Tom Sawyer*), the same keyword search described above ("clemens" and "adventures") could be made to retrieve all manifestations and expressions of the *Adventures of Tom Sawyer*, instead of the few isolated manifestations that it would retrieve in current catalogs.

We need to make sure that we design and structure the data such that the following displays are possible:

- Display all works by this author in alphabetical order by title with the sorting element (*title*) appearing at the top of each work displayed.
- Display all works on this subject in alphabetical order by principal author and title (with *principal author* and *title* appearing at top of each work displayed), or title if there is no principal author (with *title* appearing at top of each work displayed).

We must ensure that we design and structure the data in such a way that our structure allows us to create subgroups of related data, such as *instrumentation* for a piece of music (consisting of a number associated with each particular instrument), *place* and *related publisher* for a certain span of dates on a serial title change record, and the like.

Which standards will carry out which functions?

Currently, we have a number of different standards to carry out a number of different functions; we can speculate about how those functions might be allocated in a new Semantic Web–based dispensation, as shown in table 1.

In table 1, *data structure* is taken to mean what a record represents or stands for; traditionally, a record has represented an expression (in the days of hand-

(same I above)

press books) or a manifestation (ever since reproduction mechanisms have become more sophisticated, allowing an explosion of reproductions of the same content in different formats and coming from different distributors). RDA is record-neutral; RDF would allow URIs to be established for any and all of the FRBR levels; that is, there would be a URI for a particular work, a URI for a particular expression, a URI for a particular manifestation, and a URI for a particular item. Note that I am not using data structure in the sense that a computer programmer does (as a way of storing data in a computer so that it can be used efficiently).

Currently, the encoding of facts about entity relationships (see table 1) is carried out by matching data-value character strings (headings or linking fields using ISSNs and the like) that are defined by the LC/NACO authority file (following AACR2R rules), LCSH (following rules in the Subject Cataloging Manual), etc. In the future, this function might be carried out by using RDF to link the URI for a resource to the URI for a data value.

Display rules (see table 1) are currently defined by ISBD and AACR2R but widely ignored by systems, which frequently truncate bibliographic records arbitrarily in displays, supply labels, and the like; RDA abdicates responsibility, pushing display out of the cataloging rules. The general principle on the Web is to divorce data from display and allow anyone to display the data any way they want. Display is the heart of the objects (or goals) of cataloging: The point is to display to the user the works of an author, the editions of a work, or the works on a subject. All of these goals only can be met if complex, high-quality displays can be built from the data created according to the data model.

Indexing rules (see table 1) were once under the control of catalogers (in book and card catalogs) in that users had to navigate through headings and cross-references to find

Table 1. Possible reallocation of current functions in a new Semantic Web–based dispensation

Function	Current	Future?
Data content, or content guidelines (rules for providing data in a particular element)	Defined by AACR2R and MARC 21	Defined by RDA and RDF/RDFS/OWL/SKOS
Data elements	Defined by ISBD–based AACR2R and MARC 21	Defined by RDA and RDF/RDFS/OWL/SKOS
Data values	Defined by LC/NACO authority file, LCSH, MARC 21 coded data values, etc.	Defined as ontologies using RDF/RDFS/OWL/SKOS
Encoding or labeling of data elements for machine manipulation; same as data format?	Defined by ISO 2709–based MARC 21	Defined by RDF/RDFS/XML
Data structure (i.e., what a record stands for)	Defined by AACR2R and MARC 21; also FRBR?	Defined by RDF/RDFS/OWL/SKOS
Schematization (constraint on structure and content)	MARC 21, MODS, DCMI abstract model	Defined by RDF/RDFS/OWL/SKOS
Encoding of facts about entity relationships	Carried out by matching data value strings (headings found in LC/NACO authority file and LCSH, ISSN's, and the like)	Carried out by RDF/RDFS/OWL/SKOS in the form of URI links
Display rules	ILS software, formerly ISBD–based AACR2R	("Application layer") or Yee rules
Indexing rules	ILS software	SPARQL, "application layer," or Yee rules

what they wanted; currently indexing is in the hands of system designers who prefer to provide keyword indexing of bibliographic (i.e., manifestation-based) records rather than provide users with access to the entities they are really interested in (works, authors and subjects), all represented currently by authority records for headings and cross-references. RDA abdicates responsibility, pushing indexing concerns completely out of the cataloging rules. The general principle on the Web is to allow resources to be indexed by any Web search engines that wish to index them. Current Web data is not structured at all for either indexing or display.

I would argue that our interest in the Semantic Web should be focused on whether or not it will support more data structure—as well as more logic in that data structure—to support better indexes and better displays than we have now in manifestation-based ILS OPACs. Crucial to better indexing than we have ever had before are the co-occurrence rules for keyword indexing, that is, the rules for when a co-occurrence of two or more keywords should produce a match. We need to be able to do a keyword search across all possible variant names for the entity of interest, and the entity of interest for the average catalog user is much more likely to be a particular work than to be a particular manifestation. Unfortunately, catalog-use studies only have studied so-called known-item searches without investigating whether a known-item searcher was looking for a particular edition or manifestation of a work or was simply looking for a particular work in order to make a choice as to edition or manifestation once the work was found. However, common sense tells us that it is a rare user who approaches the catalog with prior knowledge about all published editions of a given work. The more common situation is surely one in which a user desires to read a particular Shakespeare play or view a particular David Lean film and discovers that the desired work exists in more than one expression or manifestation only after searching the catalog. We need to have the keyword(s) in our search for a particular work co-occur within a bounded space that encompasses all possible keywords that might refer to that particular work entity, including both creator and title keywords.

Notice in table 1 the unifying effect that RDF could potentially have; it could free us from the use of multiple standards that can easily contradict each other, or at least not live peacefully together. Examples are not hard to find in the current environment. One that has cropped up in the course of RDA development concerns family names. Presently the rules for naming families are different depending on whether the family is the subject of a work (and established according to LCSH) or whether the family is responsible for a collection of papers (and established according to RDA).

Types of data

RDA has blurred the distinctions among certain types of data, apparently because there is a perception that on the Semantic Web the same piece of data needs to be coded only once, and all indexing and display needs can be supported from that one piece of data. I question that assumption on the basis of my experience with bibliographic cataloging. All of the following ways of encoding the same piece of data can still have value in certain circumstances:

- *Transcribed*; in RDF terms, a *literal* (i.e., any data that is not a URI, a constant value). Transcribed data is data copied from an item being cataloged. It is valuable for providing access to the form of the name used on a title page and is particularly useful for people who use pseudonyms, corporate bodies that change name, and so on. Transcribed data is an important part of the historical record and not just for off-line materials; it can be a historical record of changing data on notoriously fluid webpages.
- *Composed*; in RDF terms, also a *literal*. Composed data is information composed by a cataloger on the basis of observation of the item in hand; it can be valuable for historical purposes to know which data was composed.
- *Supplied*; in RDF terms, also a *literal*. Supplied data is information supplied by a cataloger from outside sources; it can be valuable for historical purposes to know which data was supplied and from which outside sources it came.
- *Coded*; in RDF, represented by a URI. Coded data would likely transform on the Semantic Web into links to ontologies that could provide normalized, human-readable identification strings on demand, thus causing coded and normalized data to merge into one type of data. Is it not possible, though, that the coded form of normalized data might continue to provide for more efficient searching for computers as opposed to humans? Coded data also has great cross-cultural value, since it is not as language-dependent as literals or normalized headings.
- *Normalized Headings (controlled headings)*; in RDF, represented by a URI. Normalized or controlled headings are still necessary to provide users with coherent, ordered displays of thousands of entities that all match the user's search for a particular entity (*work, author, subject*, etc.). The reason Google displays are so hideous is that, so far, the data searched lacks any normalized display data. If variant language forms of the name for an entity

are linked to an entity URI, it should be possible to supply headings in the language and script desired by a particular user.

The RDF model

Those who have become familiar with FRBR over the years will probably not find it too difficult to transition from the FRBR conceptual model to the RDF model. What FRBR calls an "entity," RDF calls a "subject" and RDFS calls a "class." What FRBR calls an "attribute," RDF calls an "object" and RDFS calls a "property." What FRBR calls a "relationship," RDF calls a "predicate" and RDFS calls a "relationship" or a "semantic linkage" (see table 2).

The difficulty in any data-modeling exercise lies in deciding what to treat as an entity or class and what to treat as an attribute or property. The authors of FRBR decided to create a class called *expression* to deal with any change in the content of a work. When FRBR is applied to serials, which change content with every issue, the model does not work well. In my model, I found it useful to create a new entity at the manifestation level, the *serial title*, to deal with the type of change that is more relevant to serials, the change in title. I also created another new entity at the manifestation level, *title-manifestation*, to deal with a change of title in a nonserial work that is not associated with a change in content. One hundred years ago, this entity would have been called *title-edition*. I am also in the process of developing an entity at the expression level—*surrogate*—to deal with reproductions of original artworks that need to inherit the qualities of the original artwork they reproduce without being treated as an edition of that original artwork, which *ipso facto* is unique. These are just examples of cases in which it is not that easy to decide on the classes or entities that are necessary to accurately model bibliographic information. See the appendix for a complete comparison of the classes and entities defined in four different models: FRBR, FRAD, RDA, and the Yee Cataloging Rules (YCR). The appendix also shows variation among these models concerning whether a given data element is treated as a class/entity or as an attribute/property. The most notable examples are name and preferred access point, which are treated as classes/entities in FRAD, as attributes in FRBR and YCR, and as both in RDA.

RDF problems encountered

My goal for this paper is to institute discussion with data modelers about which problems I observed are insoluble and which are soluble:

Table 2. The FRBR conceptual model translated into RDF and RDFS

FRBR	RDF	RDFS
Entity	Subject	Class
Attribute	Object	Property
Relationship	Predicate	Relationship/ semantic linkage

1. *Is there an assumption on the part of Semantic Web developers that a given data element, such as a publisher name, should be expressed as either a literal or using a URI (i.e., controlled), but never both?* Cataloging is rooted in humanistic practices that require careful recording of evidence. There will always be value in distinguishing and labeling the following types of data:

- Copied as is from an artifact (transcribed)
- Supplied by a cataloger
- Categorized by a cataloger (controlled)

Tim Berners-Lee (the father of the Internet and the Semantic Web) emphasizes the importance of recording not just data but also its provenance for the sake of authenticity.[15] For many data elements, therefore, it will be important to be able to record both a literal (transcribed or composed form or both) and a URI (controlled form). Is this a problem in RDF? As a corollary, if any data that can be given a URI cannot also be represented by a literal (transcribed and composed data, or one or the other), it may not be possible to design coherent, readable displays of the data describing a particular entity. Among other things, cataloging is a discursive writing skill. Does RDF require that all data be represented only once, either by a literal or by a URI? Or is it perhaps possible that data that has a URI could also have a transcribed or composed form as a property? Perhaps it will even be possible to store multiple snapshots of online works that change over time to document variant forms of a name for works, persons, and so on.

2. *Will the Internet ever be fast enough to assemble the equivalent of our current records from a collection of hundreds or even thousands of URIs?* In RDF, links are one-to-one rather than one-to-many. This leads to a great proliferation of reciprocal links. The more granularity there is in the data, the more linking is necessary to ensure that atomized data elements are linked together. Potentially, every piece of data describing a particular entity could be represented by a URI leading out to a SKOS list of data values. The number of links necessary to pull together

all of the data just to describe one manifestation could become astronomical, as could the number of one-to-one links necessary to create the appearance of a one-to-many link, such as the link between an author and all the works of an author. Is the Internet really fast enough to assemble a record from hundreds of URIs in a reasonable amount of time? Given the often slow network throughput typical of many of our current Internet connections, is it really practical to expect all of these pieces to be pulled together efficiently to create a single display for a single user? We yet may feel nostalgia for the single manifestation-based record that already has all of the relevant data in it (no assembly required).

Bruce D'Arcus points out, however, that

> I think if you're dealing with RDF, you wouldn't necessarily be gathering these data in real-time. The URIs that are the targets for those links are really just global identifiers. How you get the triples is a separate matter. So, for example, in my own personal case, I'm going to put together an RDF store that is populated with data from a variety of sources, but that data population will happen by script, and I'll still be querying a single endpoint, where the RDF is stored in a relational database.[16]

In other words, D'Arcus essentially will put them all in one place, or in one database that "looks" from a URI perspective to be "one place" where they're already gathered.

3. *Is RDF capable of dealing with works that are identified using their creators?* We need to treat *author* as both an entity in its own right and as a property of a work, and in many cases the latter is the more important function for user service. Lexical labels, or human-readable identifiers for works that are identified using both the principal author and the title, are particularly problematic in RDF given that the principal author is an entity in its own right. Is RDF capable of supporting the indexing necessary to allow a user to search using any variant of the author's name and any variant of the title of a work in combination and still retrieve all expressions and manifestations of that work, given that *author* will have a URI of its own, linked by means of a relationship link to the work URI? Is RDF capable of supporting the display of a list of one thousand works, each identified by principal author, in order first by *principal author*, then by *title*, then by *publication date*, given that the preferred heading for each *principal author* would have to be assembled from the URI for that *principal author* and the *preferred title* for each work would have to be assembled from the URI for that work? For fear that this will not, in fact, be possible, I have put a human-readable work-identifier data element into my model that consists of *principal author* and *title* when appropriate, even though that means the preferred name of the principal author may not be able to be controlled by the entity record for the principal author. Any guidance from experienced data modelers in this regard would be appreciated.

According to Bruce D'Arcus, this is purely an interface or application question that does not require a solution at the data layer.[17] Since we have never had interfaces or applications that would do this correctly, even though the data is readily available in authority records, I am skeptical about this answer!

Perhaps Bruce's suggestion under item 9 of designating a *sortName* property for each entity is the solution here as well. My human-readable work identifier consisting of the name of the principal creator and uniform title of work could be designated the *sortName* poperty for the work. It would have to be changed whenever the preferred form of the name for the principal creator changed, however.

4. *Do all possible inverse relationships need to be expressed explicitly, or can they be inferred?* My model is already quite large, and I have not yet defined the inverse of every property as I really should to have a correct RDF model. In other words, for every property there needs to be an inverse property; for example, the property *isCreatorOf* needs to have the inverse property *isCreatedBy*; thus "Twain" has the property *isCreatorOf*, while *"Adventures of Tom Sawyer"* has the property *isCreatedBy*. Perhaps users and inputters will not actually have to see the huge, complex RDF data model that would result from creating all the inverse relationships, but those who maintain the model will have to deal with a great deal of complexity. However, since I'm not a programmer, I don't know how the complexity of RDF compares to the complexity of existing ILS software.

5. *Can RDF solve the problems we are having now because of the lack of transitivity or inheritance in the data models that underlie current ILSes, or will RDF merely perpetuate these problems?* We have problems now with the data models that underlie our current ILSes because of the inability of these models to deal with hierarchical inheritance, such that whatever is true of an entity in the hierarchy is also true of every entity below that entity in the hierarchy. One example is that of cross-references to a parent corporate body that should be held to apply to all subdivisions of that corporate body but never are in existing ILS systems. There is a cross-reference from "FBI" to "United States. Federal Bureau of Investigation," but not from "FBI Counterterrorism Division" to "United States. Federal Bureau of Investigation. Counterterrorism Division." For that reason, a search in any OPAC name index for "FBI Counterterrorism Division" will fail. We need systems that recognize that data about a parent corporate body is relevant to all subdivisions of that parent body. We need systems that recognize that data about a work is relevant to all expressions and manifestations of that work. RDF allows you to link a work to an expression

and an expression to a manifestation, but I don't believe it allows you to encode the information that everything that is true of the work is true of all of its expressions and manifestations. Rob Styles seems to confirm this: "RDF doesn't have hierarchy. In computer science terms, it's a graph, not a tree, which means you can connect anything to anything else in any direction."[18]

Of course, not all links should be this kind of transitive or inheritance link. One expression of work A is linked to another expression of work A by links to work A, but whatever is true of one of those expressions is not necessarily true of the other; one may be illustrated, for example, while the other is not. Whatever is true of one work is not necessarily true of another work related to it by related work link.

It should be recognized that bibliographic data is rife with hierarchy. It is one of our major tools for expressing meaning to our users. Corporate bodies have corporate subdivisions, and many things that are true for the parent body also are true for its subdivisions. Subjects are expressed using main headings and subject subdivisions, and many things that are true for the main heading (such as variant names) also are true for the heading combined with one of its subdivisions. Geographic areas are contained within larger geographic areas, and many things that are true of the larger geographic area also are true for smaller regions, counties, cities, etc., contained within that larger geographic area. For all these reasons, I believe that, to do effective displays and indexes for our bibliographic data, it is critical that we be able to distinguish between a hierarchical relationship and a nonhierarchical relationship.

6. *To recognize the fact that the subject of a book or a film could be a work, a person, a concept, an object, an event, or a place (all classes in the model), is there any reason we cannot define subject itself as a property (a relationship) rather than a class in its own right?* In my model, all subject properties are defined as having a domain of resource, meaning there is no constraint as to the class to which these subject properties apply. I'm not sure if there will be any fall-out from that modeling decision.

7. *How do we distinguish between the corporate behavior of a jurisdiction and the subject behavior of a geographical location?* Sometimes a place is a jurisdiction and behaves like a corporate body (e.g., United States is the name of the government of the United States). Sometimes place is a physical location in which something is located (e.g., the birds discussed in a book about the birds of the United States). To distinguish between the corporate behavior of a jurisdiction and the subject behavior of a geographical location, I have defined two different classes for place: *Place as Jurisdictional Corporate Body* and *Place as Geographic Area*. Will this cause problems in the model? Will there be times when it prevents us from making elegant generalizations in the model about place *per se*? There is a similar problem with events. Some events are corporate bodies (e.g., conferences that publish papers) and some are a kind of subject (e.g., an earthquake). I have defined two different classes for event: *Conference or Other Event as Corporate Body Creator* and *Event as Subject*.

8. *What is the best way to model a bound-with or an issued-with relationship, or a part–whole relationship in which the whole must be located to obtain the part?* The bound-with relationship is actually between two items containing two different works, while the issued-with relationship is between two manifestations containing two different works (see figure 2). Is this a work-to-work relationship? Will designating it a work-to-work relationship cause problems for indicating which specific items or manifestation-items of each work are physically located in the same place? This question may also apply to those part–whole relationships in which the part is physically contained within the whole and both are located in the same place (sometimes known as analytics). One thing to bear in mind is that in all of these cases the relationship between two works does not hold between all instances of each work; it only holds for those particular instances that are contained in the particular manifestation or item that is bound with, issued with, or part of the whole. However, if the relationship is modeled as a work-1-manifestation to work-2-manifestation relationship, or a work-1-item to work-2-item relationship, care must be taken in the design of displays to pull in enough information about the two or more works so as not to confuse the user.

9. *How do we express the arrangement of elements that have a definite order?* I am having trouble imagining how to encode the ordering of data elements that make up a larger element, such as the pieces of a personal name. This is really a desire to control the display of those atomized elements so that they make sense to human beings rather than just to machines. Could one define a property such as *natural language order of forename, surname, middle name, patronymic, matronymic and/or clan name of a person* given that the ideal order of these elements might vary from one person to another? Could one define properties such as *sorting element 1*, *sorting element 2*, *sorting element 3*, etc., and assign them to the various pieces that will be assembled to make a particular heading for an entity, such as an LCSH heading for a historical period? (Depending on the answer to the question in item 11, it may or may not be possible to assign a property to a property in this fashion.) Are there standard sorting rules we need to be aware of (in Unicode, for example)? Are there other RDF techniques available to deal with sorting and arrangement?

Bruce D'Arcus suggests that, instead of coding the name parts, it would be more useful to designate *sortName* properties;[19] might it not be necessary to designate a *sortName* property for each variant name, as well,

for cases in which variants need to appear in sorted displays? And wouldn't these *sortName* properties complicate maintenance over time as preferred and variant names changed?

10. *How do we link related data elements in such a way that effective indexing and displays are possible?* Some examples: number and kind of instrument (e.g., music written for two oboes and three guitars); multiple publishers, frequencies, subtitles, editors, etc., with date spans for a serial title change (or will it be necessary to create a new manifestation for every single change in subtitle, publisher name, place of publication, etc?). The assumption seems to be that there will be no repeatable data elements. Based on my somewhat limited experience with RDF, it appears that there are record equivalents (every data element—property or relationship—pertaining to a particular entity with a URI), but there are no field or subfield equivalents that allow the sublinking of related pieces of data about an entity. Indeed, Rob Styles goes so far as to argue that ultimately there is no notion of a "record" in RDF.[20] It is possible that blank nodes might be able to fill in for fields and subfields in some cases for grouping data, but there are dangers involved in their use.[21] To a cataloger, it looks as though the plan is for RDF data to float around loose without any requirement that there be a method for pulling it together into coherent displays designed for human beings.

11. *Can a property have a property in RDF?* As an example of where it might be useful to define a property of a property, Robert Maxwell suggests that *date of publication* is really an attribute (property) of the *published by* relationship (another property).[22] Another example: In my model, a variant title for a serial is a property. Can that property itself have the property *type of variant title* to encompass things like spine title, key title, etc.? Another example appeared in item 9, in which it is suggested that it might be desirable to assign sort-element properties to the various elements of a name property.

12. *How do we document record display decisions?* There is no way to record display decisions in RDF itself; it is completely display-neutral. We could not safely commit to a particular RDF–based data model until a significant amount of sample bibliographic data had been created and open-source indexing and display software had been designed and user-tested on that data. It may be that we will need to supplement RDF with some other encoding mechanism that allows us to record display decisions along with the data. Current cataloging rules are about display as much as they are about content designation. ISBD concerns the order in which the elements should be displayed to humans. The cataloging objectives concern display to users of such entity groups as the works of an author, the editions of a work, and the works on a subject.

13. *Can all bibliographic data be reduced to either a class or a property with a finite list of values?* Another way to put this is to ask if all that catalogers do could be reduced to a set of pull-down menus. Cataloging is the art of writing discursive prose as much as it is the ability to select the correct value for a particular data element. We must deal with ambiguous data (presented by Joe Blow could mean that Joe created the entire work, produced it, distributed it, sponsored it, or merely funded it). We must sometimes record information without knowing its exact meaning. We must deal with situations that have not been anticipated in advance. It is not possible to list every possible kind of data and every possible value for each type of

Issued-with relationship

A copy of Charlie Chaplin's 1917 film *The Immigrant* can be found on a videodisc compilation called *Charlie Chaplin, The Early Years* along with two other Chaplin films. This compilation was published and collected by many different libraries and media centers. If a user wants to view this copy of *The Immigrant*, he or she will first have to locate *Charlie Chaplin, The Early Years*, then look for the desired film at the beginning of the first videodisc in the set. The issued-with relationship between *The Immigrant* and the other two films on *Charlie Chaplin, The Early Years* is currently expressed in the bibliographic record by means of a "with" note:

First on Charlie Chaplin, the early years, v. 1 (62 min.) with: The count – Easy Street.

Bound-with relationship

The University of California, Los Angeles Film & Television Archive has acquired a reel of 16 mm. film from a collector who strung five Warner Bros. cartoons together on a single reel of film. We can assume that no other archive, library, or media collection will have this particular compilation of cartoons, so the relationship between the five cartoons is purely local in nature. However, any user at the Film & Television Archive who wishes to view one of these cartoons will have to request a viewing appointment for the entire reel and then find the desired cartoon among the other four on the reel. The bound-with relationship among these cartoons is currently expressed in a holdings record by means of a "with" note:

Fourth on reel with: Daffy doodles – Tweety Pie – I love to singa – Along Flirtation Walk.

Figure 2. Examples of part–whole relationships. How might these be best expressed in RDF?

data up front before any data is gathered. It will always be necessary to provide a plain-text escape hatch. The bibliographic world is a complex, constantly changing world filled with ambiguity.

What are the next steps?

In a sense, this paper is a first crude attempt at locating unmapped territory that has not yet been explored. If we were to decide as a community that it would be valuable to move our shared cataloging activities onto the Semantic Web, we would have a lot of work ahead of us. If some of the RDF problems described above are insoluble, we may need to work with Semantic Web developers to create a more sophisticated version of RDF that can handle the transitivity and complex linking required by our data. We will also need to encourage a very complex existing community to evolve institutional structures that would enable a more efficient use of the Internet for the sharing of cataloging and other metadata creation. This is not just a technological problem, but also a political one. In the meantime, the experiment continues. Let the thinking and learning begin!

References and notes

1. "Notation3, or N3 as it is more commonly known, is a shorthand non–XML serialization of Resource Description Framework models, designed with human-readability in mind: N3 is much more compact and readable than XML RDF notation. The format is being developed by Tim Berners-Lee and others from the Semantic Web community." Wikipedia, "Notation 3," http://en.wikipedia.org/wiki/Notation_3 (accessed Feb. 19, 2009).
2. FRBR Review Group, www.ifla.org/VII/s13/wgfrbr/; FRBR Review Group, FRANAR (Working Group on Functional Requirements and Numbering of Authority Records), www.ifla.org/VII/d4/wg-franar.htm; FRBR Review Group, FRSAR (Working Group, Functional Requirements for Subject Authority Records), www.ifla.org/VII/s29/wgfrsar.htm; FRBRoo, FRBR Review Group, Working Group on FRBR/CRM Dialogue, www.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue_wg.htm.
3. Library of Congress, *Response to On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control* (Washington, D.C.: Library of Congress, 2008): 24, 39, 40, www.loc.gov/bibliographic-future/news/LCWGRptResponse_DM_053008.pdf (accessed Mar. 25, 2009).
4. Ibid., 39.
5. Ibid., 41.
6. Dublin Core Metadata Initiative, DCMI/RDA Task Group Wiki, http://www.dublincore.org/dcmirdataskgroup/ (accessed Mar. 25, 2009).
7. Mikael Nilsson, Andy Powell, Pete Johnston, and Ambjorn Naeve, *Expressing Dublin Core Metadata Using the Resource Description Framework (RDF)*, http://dublincore.org/documents/2008/01/14/dc-rdf/ (accessed Mar. 25, 2009).
8. See for example table 6.3 in FRBR, which maps to manifestation every kind of data that pertains to expression change with the exception of language change. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records* (Munich: K. G. Saur, 1998): 95, http://www.ifla.org/VII/s13/frbr/frbr.pdf (accessed Mar. 4, 2009).
9. Roy Tennant, "MARC Must Die," *Library Journal* 127, no. 17 (Oct. 15, 2002): 26.
10. W3C, *SKOS Simple Knowledge Organization System Reference, W3C Working Draft 29 August 2008*, http://www.w3.org/TR/skos-reference/ (accessed Mar. 25, 2009).
11. The extract in figure 1 is taken from my complete RDF model, which can be found at http://myee.bol.ucla.edu/ycrschemardf.txt.
12. Mary W. Elings and Gunter Waibel, "Metadata for All: Descriptive Standards and Metadata Sharing Across Libraries, Archives and Museums," *First Monday* 12, no. 3 (Mar. 5, 2007), http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1628/1543 (accessed Mar. 25, 2009).
13. OCLC, *A Holdings Primer: Principles and Standards for Local Holdings Records*, 2nd ed. (Dublin, Ohio: OCLC, 2008), 4, http://www.oclc.org/us/en/support/documentation/localholdings/primer/Holdings%20Primer%202008.pdf (accessed Mar. 25, 2009).
14. The Library of Congress Working Group, *On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control* (Washington, D.C.: Library of Congress, 2008): 30, http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf (accessed Mar. 25, 2009).
15. Talis, *Sir Tim Berners-Lee Talks with Talis about the Semantic Web: Transcript of an Interview Recorded on 7 February 2008*, http://talis-podcasts.s3.amazonaws.com/twt20080207_TimBL.html (accessed Mar. 25, 2009).
16. Bruce D'Arcus, e-mail to author, Mar. 18, 2008.
17. Ibid.
18. Rob Styles, e-mail to author, Mar. 25, 2008.
19. Bruce D'Arcus, e-mail to author, Mar. 18, 2008.
20. Rob Styles, e-mail to author, Mar. 25, 2008.
21. W3C, "Section 2.3, Structured Property Values and Blank Nodes," in *RDF Primer: W3C Recommendation 10 February 2004*, http://www.w3.org/TR/rdf-primer/#structuredproperties (accessed Mar. 25, 2009).
22. Robert Maxwell, *FRBR: A Guide for the Perplexed* (Chicago: ALA, 2008).

APPENDIX. Entity/class and attribute/property comparisons

Entities/classes in RDA, FRBR, FRAD compared to Yee Cataloging Rules (YCR)

RDA, FRBR, and FRAD	YCR
Group 1: Work	Work
Group 1: Expression	Expression Surrogate
Group 1: Manifestation	Manifestation Title-manifestation Serial title
Group 1: Item	Item
Group 2: Person	Person Fictitious character Performing animal
Group 2: Corporate body	Corporate body Corporate subdivision Place as jurisdictional corporate body Conference or other event as corporate body creator Jurisdictional corporate subdivision
Family (RDA and FRAD only)	
Group 3: Concept	Concept
Group 3: Object	Object
Group 3: Event	Event or historical period as subject
Group 3: Place	Place as geographic area
	Discipline
	Genre/form
Name	
Identifier	
Controlled access point	
Rules (FRAD only)	
Agency (FRAD only)	

Attributes/properties in FRBR compared to FRAD

	Model	
Entity	**FRBR**	**FRAD**
Work	title of the work form of work date of the work other distinguishing characteristics intended termination intended audience context for the work medium of performance (musical work) numeric designation (musical work) key (musical work) coordinates (cartographic work) equinox (cartographic work)	form of work date of the work medium of performance subject of the work numeric designation key place of origin of the work original language of the work history other distinguishing characteristic
Expression	title of the expression form of expression date of expression language of expression other distinguishing characteristics extensibility of expression revisability of expression extent of the expression summarization of content context for the expression critical response to the expression use restrictions on the expression sequencing pattern (serial) expected regularity of issue (serial) expected frequency of issue (serial) type of score (musical notation) medium of performance (musical notation or recorded sound) scale (cartographic image/object) projection (cartographic image/object) presentation technique (cartographic image/object) representation of relief (cartographic image/object) geodetic, grid, and vertical measurement (cartographic image/object) recording technique (remote sensing image) special characteristic (remote sensing image) technique (graphic or projected image)	form of expression date of expression language of expression technique other distinguishing characteristic
Surrogate		

Attributes/properties in FRBR compared to FRAD (cont.)

	Model	
Entity	**FRBR**	**FRAD**
Manifestation	title of the manifestation statement of responsibility edition/issue designation place of publication/distribution publisher/distributor date of publication/distribution fabricator/manufacturer series statement form of carrier extent of the carrier physical medium capture mode dimensions of the carrier manifestation identifier source for acquisition/access authorization terms of availability access restrictions on the manifestation typeface (printed book) type size (printed book) foliation (hand-printed book) collation (hand-printed book) publication status (serial) numbering (serial) playing speed (sound recording) groove width (sound recording) kind of cutting (sound recording) tape configuration (sound recording) kind of sound (sound recording) special reproduction characteristic (sound recording) colour (image) reduction ratio (microform) polarity (microform or visual projection) generation (microform or visual projection) presentation format (visual projection) system requirements (electronic resource) file characteristics (electronic resource) mode of access (remote access electronic resource) access address (remote access electronic resource)	edition/issue designation place of publication/distribution publisher/distributor date of publication/distribution form of carrier numbering
Title-manifestation		
Serial title		
Item	item identifier fingerprint provenance of the item marks/inscriptions exhibition history condition of the item treatment history scheduled treatment access restrictions on the item	location of item

Attributes/properties in FRBR compared to FRAD (cont.)

	Model	
Entity	**FRBR**	**FRAD**
Person	name of person dates of person title of person other designation associated with the person	dates associated with the person title of person other designation associated with the person gender place of birth place of death country place of residence affiliation address language of person field of activity profession/occupation biography/history
Fictitious character		
Performing animal		
Corporate body	name of the corporate body number associated with the corporate body place associated with the corporate body date associated with the corporate body other designation associated with the corporate body	place associated with the corporate body date associated with the corporate body other designation associated with the corporate body type of corporate body language of the corporate body address field of activity history
Corporate subdivision		
Place as jurisdictional corporate body		
Conference or other event as corporate body creator		
Jurisdictional corporate subdivision		
Family		type of family dates of family places associated with family history of family
Concept	term for the concept	type of concept
Object	term for the object	type of object date of production place of production producer/fabricator physical medium
Event	term for the event	date associated with the event place associated with the event

Attributes/properties in FRBR compared to FRAD (cont.)

	Model	
Entity	**FRBR**	**FRAD**
Place	term for the place	coordinates other geographical information
Discipline		
Genre/form		
Name		type of name scope of usage dates of usage language of name script of name transliteration scheme of name
Identifier		type of identifier identifier string suffix
Controlled access point		type of controlled access point status of controlled access point designated usage of controlled access point undifferentiated access point language of base access point script of base access point script of cataloguing transliteration scheme of base access point transliteration scheme of cataloguing source of controlled access point base access point addition
Rules		citation for rules rules identifier
Agency		name of agency agency identifier location of agency

Attributes/properties in RDA compared to YCR

	Model	
Entity	**RDA**	**YCR**
Work	title of the work	key identifier for work
	form of work	language-based identifier (preferred lexical label)
	date of work	variant language-based identifier (alternate lexical label)
	place of origin of work	language-based identifier (preferred lexical label) for work
	medium of performance	language-based identifier for work (preferred lexical label) identified by PrincipalCreator in combination with uniform title
	numeric designation	language-based identifier (preferred lexical label) for work identified by title alone (uniform title)
	key	supplied title for work
	signatory to a treaty, etc.	variant title for work
	other distinguishing characteristic of the work	original language of work
	original language of the work	responsibility for work
	history of the work	original publication statement of work
	identifier for the work	dates associated with work
	nature of the content	original publication/release/broadcast date of work
	coverage of the content	copyright date of work
	coordinates of cartographic content	creation date of work
	equinox	date of first recording of a work
	epoch	date of first performance of a work
	intended audience	finding date of naturally occurring object
	system of organization	original publisher/distributor/broadcaster of work
	dissertation or theses information	places associated with work
		original place of publication/distribution/broadcasting for work
		country of origin of work
		place of creation of work
		place of first recording of work
		place of first performance of work
		finding place of naturally occurring object
		original method of publication/distribution/broadcast of work
		serial or integrating work original numeric and/or alphabetic designations—beginning
		serial or integrating work original chronological designations—beginning
		serial or integrating work original numeric and/or alphabetic designations—ending
		serial or integrating work original chronological designations—ending
		encoding of content of work
		genre/form of content of work
		original instrumentation of musical work
		instrumentation of musical work—number of a particular instrument
		instrumentation of musical work—type of instrument
		original voice(s) of musical work
		voice(s) of musical work—number of a particular type of voice
		voice(s) of musical work—type of voice
		original key of musical work
		numeric designation of musical work
		coordinates of cartographic work
		equinox of cartographic work
		original physical characteristics of work
		original extent of work
		original dimensions of work
		mode of issuance of work

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Work (cont.)		original aspect ratio of moving image work original image format of moving image work original base of work original materials applied to base of work work summary work contents list custodial history of work creation of archival collection censorship history of work note about relationship(s) to other works
Expression	content type date of expression language of expression other distinguishing characteristic of the expression identifier for the expression summarization of the content place and date of capture language of the content form of notation accessibility content illustrative content supplementary content colour content sound content aspect ratio format of notated music medium of performance of musical content duration performer, narrator, and/or presenter artistic and/or technical credits scale projection of cartographic content other details of cartographic content awards	key identifier for expression language-based identifier (preferred lexical label) for expression variant title for expression nature of modification of expression expression title expression statement of responsibility edition statement scale of cartographic expression projection of cartographic expression publication statement of expression place of publication/distribution/release/broadcasting for expression place of recording for expression publisher/distributor/releaser/broadcaster for expression publication/distribution/release/broadcast date for expression copyright date for expression date of recording for expression numeric and/or alphabetic designations for serial expressions chronological designations for serial expressions performance date for expression place of performance for expression extent of expression content of expression language of expression text language of expression captions language of expression sound track language of sung or spoken text of expression language of expression subtitles language of expression intertitles language of summary or abstract of expression instrumentation of musical expression instrumentation of musical expression—number of a particular instrument instrumentation of musical expression—type of instrument voice(s) of musical expression voice(s) of musical expression—number of a particular type of voice voice(s) of musical expression—type of voice key of musical expression appendages to the expression expression series statement mode of issuance for expression notes about expression
Surrogate		[under development]

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Manifestation	title	key identifier for manifestation
	statement of responsibility	publication statement of manifestation
	edition statement	place of publication/distribution/release/broadcast of manifestation
	numbering of serials	manifestation publisher/distributor/releaser/broadcaster
	production statement	manifestation date of publication/distribution/release/broadcast
	publication statement	carrier edition statement
	distribution statement	carrier piece count
	manufacture statement	carrier name
	copyright date	carrier broadcast standard
	series statement	carrier recording type
	mode of issuance	carrier playing speed
	frequency	carrier configuration of playback channels
	identifier for the manifestation	process used to produce carrier
	note	carrier dimensions
	media type	carrier base materials
	carrier type	carrier generation
	base material	carrier polarity
	applied material	materials applied to carrier
	mount	carrier encoding format
	production method	intermediation tool requirements
	generation	system requirements
	layout	serial manifestation illustration statement
	book format	manifestation standard number
	font size	manifestation ISBN
	polarity	manifestation ISSN
	reduction ratio	manifestation publisher number
	sound characteristics	manifestation universal product code
	projection characteristics of motion picture film	notes about manifestation
	video characteristics	
	digital file characteristics	
	equipment and system requirements	
	terms of availability	
Title-manifestation		key identifier for title-manifestation
		variant title for title-manifestation
		title-manifestation title
		title-manifestation statement of responsibilities
		title-manifestation edition statement
		publication statement of title-manifestation
		place of publication/distribution/release/broadcasting of title-manifestation
		publisher/distributor/releaser, broadcaster of title-manifestation
		date of publication/distribution/release/broadcast of title-manifestation
		title-manifestation series
		title-manifestation mode of issuance
		notes about title-manifestation
		title-manifestation standard number

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Serial title		key identifier for serial title variant title for serial title title of serial title serial title statement of responsibility serial title edition statement publication statement of serial title place of publication/distribution/release/broadcast of serial title publisher/distributor/releaser/broadcaster of serial title date of publication/distribution/release/broadcast of serial title serial title beginning numeric and/or alphabetic designations serial title beginning chronological designations serial title ending numeric and/or alphabetic designations serial title ending chronological designations serial title frequency serial title mode of issuance serial title illustration statement notes about serial title serial title ISSN-L
Item	preferred citation custodial history immediate source of acquisition identifier for the item item-specific carrier characteristics	key identifier for item item barcode item location item call number or accession number item copy number item provenance item condition item marks and inscriptions item exhibition history item treatment history item scheduled treatment item access restrictions

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Person	name of the person preferred name for the person variant name for the person date associated with the person title of the person fuller form of name other designation associated with the person gender place of birth place of death country associated with the person place of residence address of the person affiliation language of the person field of activity of the person profession or occupation biographical information identifier for the person	key identifier for person language-based identifier (preferred lexical label) for person clan name of person forename/given name/first name of person matronymic of person middle name of person nickname of person patronymic of person surname/family name of person natural language order of forename, surname, middle name, patronymic, matronymic and/or clan name of person affiliation of person biography/history of person date of birth of person date of death of person ethnicity of person field of activity of person gender of person language of person place of birth of person place of death of person place of residence of person political affiliation of person profession/occupation of person religion of person variant name for person
Fictitious character		[under development]
Performing animal		[under development]
Corporate body	name of the corporate body preferred name for the corporate body variant name for the corporate body place associated with the corporate body date associated with the corporate body associated institution other designation associated with the corporate body language of the corporate body address of the corporate body field of activity of the corporate body corporate history identifier for the corporate body	key identifier for corporate body language-based identifier (preferred lexical label) for corporate body dates associated with corporate body field of activity of corporate body history of corporate body language of corporate body place associated with corporate body type of corporate body variant name for corporate body
Corporate subdivision		[under development]
Place as jurisdictional corporate body		[under development]

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Conference or other event as corporate body creator		[under development]
Jurisdictional corporate subdivision		[under development]
Family	name of the family preferred name for the family variant name for the family type of family date associated with the family place associated with the family prominent member of the family hereditary title family history identifier for the family	
Concept	term for the concept preferred term for the concept variant term for the concept type of concept identifier for the concept	key identifier for concept language-based identifier (preferred lexical label) for concept qualifier for concept language-based identifier variant name for concept
Object	name of the object preferred name for the object variant name for the object type of object date of production place of production producer/fabricator physical medium identifier for the object	key identifier for object language-based identifier (preferred lexical label) for object qualifier for object language-based identifier variant name for object
Event	name of the event preferred name for the event variant name for the event date associated with the event place associated with the event identifier for the event	key identifier for event or historical period as subject language-based identifier (preferred lexical label) for event or historical period as subject beginning date for event or historical period as subject ending date for event or historical period as subject variant name for event or historical period as subject
Place	name of the place preferred name for the place variant name for the place coordinates other geographical information identifier for the place	key identifier for place as geographic area language-based identifier (preferred lexical label) for place as geographic area qualifier for place as geographic area variant name for place as geographic area
Discipline		key identifier for discipline language-based identifier (preferred lexical label) (name or classification number or symbol) for discipline translation of meaning of classification number or symbol for discipline

Attributes/properties in RDA compared to YCR (cont.)

	Model	
Entity	**RDA**	**YCR**
Genre/form		key identifier for genre/form language-based identifier (preferred lexical label) for genre/form variant name for genre/form
Name	scope of usage date of usage	
Identifier		
Controlled access point		
Rules		
Agency		

Note: In RDA, the following attributes have not yet been assigned to a particular class or entity: *extent, dimensions, terms of availability,* contact information, *restrictions on access, restrictions on use, uniform resource locator, status of identification, source consulted, cataloguer's note, status of identification,* and *undifferentiated name indicator. Name* is being treated as both a class and a property. *Identifier* and *controlled access point* are treated as properties rather than classes in both RDA and YCR.