TEN

BIBLIOGRAPHIC DESCRIPTION AND THE SEMANTIC WEB

Computer technology evolves rapidly, with change happening faster than most humans can follow. The technology of data management evolves less quickly than some other computer technologies because it often requires a rather costly conversion of data and interfaces to take advantage of new capabilities. For this reason, few can afford to jump on a new data technology bandwagon when it first appears. Instead, most wait until such a technology has matured before adoption.

The technology of the Semantic Web, although new to many, has been around for almost as long as the FRBR Final Report. The first version of RDF, the basis for the Semantic Web, was published by the World Wide Web Consortium in 1999. Adoption of RDF has been slow, but as of 2015 there is a strong movement toward implementation of this technology and integration of data with the basic functioning of the web. Most relevant to those of us in libraries, it is being used or is in the planning stages for a significant number of library and archives applications.

There is a logical progression from entity-relation modeling, such as is used in FRBR, and Semantic Web technology that models data as things and relationships. That there is a logical progression does not mean that these technologies are the same. In fact, as described in chapter 3 of this book, there are significant differences between the meaning behind entity-relation models and the Semantic Web approach that absolutely must be understood when making the transition from one to the other. But in general, an entity-relation model is not a bad predecessor to a Semantic Web data design, as long as suitable adjustments are made.

ENTITY-RELATION MODELING AND SEMANTIC WEB MODELS

> This approach differs from some other approaches in that it begins with an abstract of conceptual schema of the domain or universe in question. The universe is characterized in terms of the entities in it and the relationships that hold among them. As such, the conceptual schema is not restricted by the capabilities of any particular database system and is independent of any particular record definition. (Tillett 1994)

To understand the Semantic Web model, we need to review how data was managed before the Semantic Web, and how this relates to efforts to define bibliographic data for the Semantic Web that uses FRBR as its conceptual basis. As we've seen, FRBR was designed around an entity-relation analysis. The entity-relation method was developed to support relational database management technology, which was revolutionizing data management in the 1980s and 1990s. The problems addressed by entity-relation design are primarily related to the efficiency of storage and retrieval of data held in database management systems. An entity-relation analysis normalizes data gathered for business functions with a goal of storing each data point only once, and allowing the combination of atomized data elements to support a wide variety of business functions. At the conceptual level, entity-relation analysis identifies the primary entities that the enterprise must manage as data. Subsequent steps define data elements supporting or describing each entity, and which are needed to enable the tasks and workflow of the enterprise. These analyses may support static or dynamic processes, depending on the needs of the organization.

The Semantic Web has an entirely different approach from that used by relational databases. These differences arise from primary assumptions of the closed world of databases versus the open world of the web.

Databases reside generally within a system that is contained within the local network of an organization. Although that local network most likely connects in some way to the Internet, the database itself is not on the web. In fact, in some cases, like banks, hospitals, and government agencies, the security of the database is one of the key requirements of the data management function. For example, the banking industry has its own secure network for the exchange of information between banks; this information is not available over the open web and cannot be directly accessed by other than local users with particular privileges.

Even for those for whom the privacy and security of the data are not paramount, databases do not communicate directly with the web; the interaction with the web is managed through applications that have controlled and secure access to the database contents. This is true for library systems and their online catalogs. Although one can access the online catalog via a web browser, the data resides in a database on a private server.

In a Relational Database Management System (RDBMS), the data elements allowed are controlled by the database structure definition. This structure is made up of tables that consist of rows of data. Each data element in a table is defined as a specific data type, such as text, integer, or structured data. The data stored in the database is not in the same form as it appears in input or output actions. Instead, data in the tables is combined to present views of the data for input and output. For example, one common type of data in a database is that of names and addresses of customers, users, or employees. The data that makes up a complete mailing address may be stored across several tables such that city and state names are not repeated in the database but can be linked to the particular address when requested. The database design does not represent a record, but is a store of data elements that are input and output in different combinations, each of which may be considered a record by a particular user. A database may serve a wide variety of organizational functions, including some that do not share any data elements among them. In a large organization, management of the database is under the control of a technical department, although input and output may take place throughout the organization by different types of employees. These employees may be entirely unaware of the nature of the whole database, and only see the portion that is relevant to them through a user interface.

A common aspect of relational databases is that they have features built in for data quality control. In fact, the database design itself forces data to conform to rules. For example, certain data elements can be required in the database design. If the database has only one storage point for a data element, more than one element simply cannot be added. In particular, no new element types can be added by database users until changes are made to the database design. Because these changes often require a significant amount of planning and testing, data stored in databases has a tendency to change only with a certain effort. Anything not included in the RDBMS design is not available to the applications that will use the data.

It is common to hear the web of linked data referred to as a way to use the web as your database. Although this analogy holds, the difference between the web as database and the closed, controlled world of the enterprise database management system is enormous. The open world of the Semantic Web has no predefined boundaries. The maxim used on the Semantic Web is that "anyone can say anything about anything." The Semantic Web has no concept of quality control analogous to an RDBMS. This has advantages as well as some challenges. The advantage is that new data types can be added as needed. Rather than the slow and often painful process of database modification that traditional closed databases require, the web of linked data can allow new concepts to be added by anyone at any time. Any data on the web can be linked to your data, just as today anyone on the web can add a document that links to a web page of yours. There is also no way to require that certain data always be present in the Semantic Web model. Because the open world of the web is considered to be ever-changing (as is the real world of information), it is not possible to predict what data will be available at any given moment in time. Data that is missing at one moment may be present a moment later. Therefore the Open World Assumption, one of the fundamental tenets of the Semantic Web, does not treat missing data as an error but merely as something momentarily unknown.

In the pre-Semantic Web data environment, we generally consider ourselves to have data encapsulated in records outside of the database. The database itself has tables and does not conform to the form of the records that are input and output. However, for those who create and use the data, the record is the unit that defines an appropriate grouping of data elements. Like a database, the Semantic Web does not have records. But the input and output, when using the Semantic Web standards based on RDF, also are not packaged as records. Regardless of the view that one has at any given time, the Semantic Web is always composed of three-part statements called triples, which are autonomous atomic statements that can be combined in any desired configuration. A search against the Semantic Web returns some number of triples that match the search query.

FRBR, and subsequently RDA, were modeled as one would model data for an RDBMS. As part of the development of RDA, the Joint Steering Committee for RDA produced a document that presented three high-level database models for bibliographic data (Delsey 2009). The most advanced of these models, and presumably the preferred one, was called "Scenario 1," and made use of concepts from relational database design. The documents reflect some assumptions about systems efficiency and user service capabilities that were held by members of the committee:

> The data structures used to store the data and to reflect relationships, however, will have a bearing both on the efficiency of data creation and maintenance, and on the ease and effectiveness with which users are able to access the data and navigate the database. (Delsey 2009)

There are two ways in which efficiency of a database is commonly measured: the time needed for operations to the database to complete, and the space needs to data storage. In practice these two considerations are often in conflict. It is assumed that the storage requirements for FRBR work data (and to a lesser extent FRBR expression, because it contains fewer fields) would be reduced because there would be less duplication of these elements in the database. None of the statistics that we have on hand, however, measure data storage, so it remains an open question whether there are significant savings.

In terms of time necessary to complete operations, the efficiency of a database often depends on how many joins and reads are required. Joins and reads are what bring together related data that is stored in the separate database tables. Some efficiency that is gained by reducing redundancy is lost when multiple tables must be included in a single database activity. Database models are adjusted to meet these efficiency requirements by testing. This type of testing would be needed before any definite statements could be made about a FRBR-based database model.

Suffice it to say that at this time there are no studies available that support any claims of efficiency for RDA as a FRBR-based bibliographic model. How efficient the sharing of works and expressions for catalog maintenance depends on the extent of redundancy of entities in the local catalog, which is usually a function of a library's mission and size, as well as the extent to which bibliographic data is held in a shared data pool rather than copied to the local library system.

The Semantic Web model is also seen as holding promise in terms of data sharing. The Semantic Web is by its nature a shared "cloud," although the same technologies are used in closed enterprise systems. The primary data structure of the Semantic Web, RDF, is already being used in large corporations that need to share and link data between operations and over large geographical spaces. There is tangible tension between the Semantic Web design for the open web and the use of these technologies in closely held data environments. For many enterprises, it's not a matter of entirely open or entirely closed but, as with traditional databases, using the networking capabilities of the Internet to share select information with a wider public.

FRBR IN RDF

The conceptual model of FRBR and variants on that model have been expressed in recent RDF vocabularies developed for library data. FRBR has also been used in bibliographic models designed outside of the library. There are also Semantic Web implementations of bibliographic data, such as some develop for academic citations, that do not make use of the FRBR conceptual entities. This section focuses on some of the projects that have transformed the entity-relation model of FRBR to a Semantic Web vocabulary.

FRBRer

In 2011, the IFLA FRBR Review Group, the group that now maintains the FRBR standards (including functional requirements for authority and subject data) issued its official version of FRBR in RDF. Called FRBRer, with the "er" standing for entity-relation, this version is based on the 2009 edition of the FRBR Study Group's final report.

FRBRer is an encoding of FRBR as defined in the document. FRBRer uses the World Wide Web Consortium's standard for the definition of vocabularies, the Web Ontology Language (OWL), which is described in chapter 3 of this volume. As the literal translation of an entity-relation model, FRBRer does not follow some practices that are common in Semantic Web vocabularies. In particular, FRBRer does not make use of super- or sub-classes to define logical types of entities. Without class relationships, which are heavily used in Semantic Web vocabularies, FRBRer cannot define a single relationship that is valid between a work and any members of FRBR Group 2. Instead, FRBRer must develop a specific relationship for each individual Group 2 entity: for example, "is created by person," or "is created by corporate body." This must be done for each of the relationships involving Group 2 entities such as "realized by," "produced by," and "owned by." The later addition of "family" to Group 2 meant that new relationships had to be added for it as well. This has an effect on the ease of extensibility of FRBRer: with class relationships it would only be necessary to define the entity "family" as a sub-class of the class "responsible entity," and family would be immediately usable with all relationships defined between Group 2 and Group 1. As it is, the addition of family also requires the family-specific definition of each of the relevant relationships to the other entities.

It is clear that FRBRer is designed with a closed-world point of view rather than the open view of the Semantic Web. This view in inherited from the entity-relation origins of FRBR, because entity-relation modeling is a design method for database technologies and not for the Semantic Web. As I noted in the chapter on technology, there is a common misinterpretation of the Semantic Web vocabulary definition language OWL that reads the OWL rules as quality-control constraints on the data rather than as axioms for making inferences from the data. FRBRer is based on this misinterpretation.

One way that FRBRer affirms its closed-world view is by declaring all entities and all of the FRBR attributes as disjoint from each other. This means that it isn't possible for anyone using FRBRer to create data that varies in the assignment of attributes to entities. The result of this is that the data coded in this vocabulary is very fragile, with any deviation from the defined terms causing OWL-aware software to fail to function as desired. This fragility is not easy to mitigate because OWL does not have any functions that enforce quality control on data; instead, in the open world where "anyone can say anything about anything," vocabularies need to be as forgiving as possible. The strictness implied by disjoint classes may be unrealistic in an imperfect world, with a few exceptions for undeniable truths, such as "up" being disjoint "down." Nearly any category has exceptions that need to be handled, even the fundamental states "solid, liquid, or gas," which include intermediate transitions that can be more than one state at the same time. In the open world of the web, the strictness implied by disjoint classes generally prevents this data from interoperating with any data that is based on a different model, even if they have linkable elements in common. In the case of FRBRer, this creates an incompatibility with any bibliographic data that does not precisely separate its bibliographic description into identically defined entities. Not only does this mean that variants of FRBR as desired by some library specialist and non-book communities cannot be used alongside conformant FRBRer, but that FRBRer on the open web may not be able to link to similar data using models like FRBRcore, BIBFRAME, or even RDA in RDF.

In this way the primary advantage of the Semantic Web, discovery across heterogeneous data contributed to the web of data by different communities, is negated by the definition of vocabularies using OWL with an inappropriate closed-world assumption. FRBRer, as defined, is a vocabulary that will necessarily be usable only in its own silo.

RDA in RDF

Resource Description and Access (RDA), the cataloging rules adopted in 2013 by many North American libraries, also has a defined RDF-based vocabulary. The development of RDA was closely tied to the bibliographic concepts presented in the FRBR Final Report, and the structure of the RDA rules follows that of the FRBR groups—primarily Group 1, which is the main focus of descriptive cataloging. A list of data elements, each with their assigned FRBR entity, was published as a supplement to the final text of RDA. After a 2007 meeting between volunteers active in the Dublin Core Metadata Community and members of the RDA development group, the RDA elements were defined in RDF on an experimental basis. This activity preceded the development of FRBRer, and RDA in RDF was based on a very loose vocabulary of FRBR entities as RDF classes. Although FRBRer is now available, and there is an official version of RDA in RDF that is managed by the Joint Steering Committee (JSC) that maintains the cataloging rules, the JSC has so far chosen not to follow the vocabulary as defined by FRBRer. Instead, RDA in RDF has created its own FRBR classes to define the domain of each of the data elements, but does not attempt to impose E-R-like rules on the resulting bibliographic description. Each RDA element is also sub-classed to a parent element that is not associated with any FRBR entity, creating a FRBR-neutral version of the vocabulary that may be more acceptable to nonlibrary communities for whom the Group 1 entities of FRBR are unfamiliar, and perhaps even not useful. RDA in RDF creates a super-class for person, corporate body, and family that is called "agent," but no super-class for the Group 1 entities. Recall that the FRBR Review Group has made clear that in their analysis the groups are not to be represented as super-classes; instead, they exist as organizing elements in the documentation only.

With an awareness of the open world and the fact that there are many sources of bibliographic data, both within the library community and outside of it, RDA in RDF proposes linking RDA elements to commonly used terms from other vocabularies using sub-class relationships, with a more general vocabulary like Dublin Core Metadata Terms as the common language. This can bridge the gap between the over 900 elements in RDA and bibliographic data as approached by nonlibrarians. There probably is no other community that has so many different types of titles (key title, parallel title, series title, parallel series title, etc.), yet most communities creating bibliographic data will have an element that is compatible with Dublin Core "title," or that can be sub-classed to it. Because the primary goal of the Semantic Web is to allow linking between comparable data across the web, it is a good idea for everyone to design links into their community-specific vocabulary to well-known vocabularies used on the web. In this way one avoids being stuck in a silo where library data can only connect to other library data on the web.

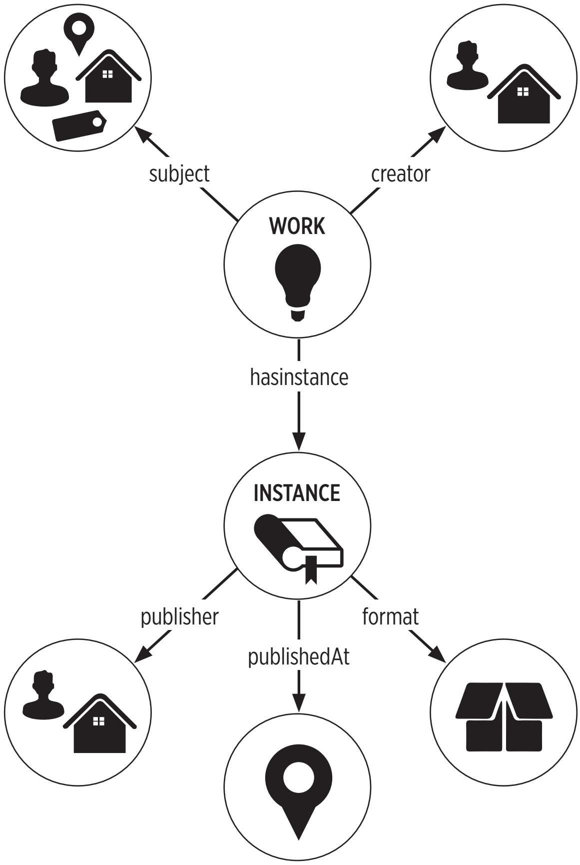
That said, RDA in RDF itself has few relationships between terms, and few uses of classes and sub-classes. This is possibly a reflection of its origins as a list of terms derived from the thirty-odd chapters of the RDA cataloging rules. You could say that RDA has not yet been subjected to the data design phase that would look at the desired functionality that could be delivered with complex bibliographic data expressed in RDF. This type of design often begins with a statement of use cases: what is it that we want to do, and that we could do, with this data? Recall that in chapte 2 on modeling we had simple use cases like "find a book of which the author is known." Given the kinds of capabilities that we have with current data management technology, including the possibility of using the entire web as the context for our information services, the number and kinds of use cases would surely grow. Do we want to promote linking from online sites for readers, like GoodReads or LibraryThing, to library materials? From the music database MusicBrainz to library collections of recorded sound? From the Internet Movie Database to film descriptions in library catalogs? Do we want to incorporate more information about authors alongside the library's bibliographic holdings? Do we want to continue to store most catalog data locally, or would "cloud-based" sharing of data be more efficient? The list would end up being quite long, but without an exploration of goals, we cannot make rational decisions to guide our development.

BIBFRAME

The RDF vocabulary of BIBFRAME, which is being developed simultaneously by Library of Congress, Zepheira, and a number of library projects, is not an implementation of FRBR, but is clearly influenced by the FRBR model. It has a two-entity model of bibliographic description, with the entities called work and instance. The BIBFRAME work represents the content portion of the bibliographic description, and the instance describes the carrier. Generally speaking, the BIBFRAME work encompasses attributes that are associated with both the FRBR work and expression; the BIBFRAME instance is analogous to the FRBR manifestation. Item-level information is not treated as one of the primary bibliographic entities in BIBFRAME. BIBFRAME also resembles the pre-FRBR data model with a central bibliographic description plus authority entities that are similar to name and subject authorities (figure 10.1).

FIGURE 10.1

BIBFRAME model



BIBFRAME and FRBR come out of different communities within the library environment. Although the FRBR Final Report provides background on the bibliographic theory that led to the creation of FRBR, the link between standard cataloging practice and BIBFRAME is less clear. As the "named successor" to the current standard bibliographic record, MARC 21, BIBFRAME is primarily a data standard. To accommodate MARC 21, BIBFRAME will necessarily have more detail than the FRBR conceptual model, which was developed ostensibly as a minimum set of bibliographic data.

SEPARATE WORK AND INSTANCE WITH LINKS	SINGLE RESOURCE WITH WORK AND INSTANCE DATA ELEMENTS
ex:ResourceA	ex:ResourceA
bf:workTitle ex:AdventuresOfTomSawyer ;	bf:creator lcna: n79021164 ;
bf:hasInstance ex:ResourceB ;	bf:workTitle ex:AdventuresOfTomSawyer> ;
bf:creator lcna: n79021164 ;	bf:language is0639–2:eng ;
bf:language is0639–2:eng .	bf:instanceTitle ex:TheAdventuresOfTomSawyer> ;
ex:ResourceB	bf:providerDate “1996” .
bf:providerDate “1996” ;	
bf:instanceOf ex:ResourceA ;	
bf:instanceTitle ex:TheAdventuresOfTomSawyer .	

BIBFRAME’s vocabulary is more compatible with the open web and with potential variations in bibliographic concepts than is FRBRer. BIBFRAME makes use of classes and sub-classes in ways that are convenient for designers of systems, which will probably facilitate searching and other system functionality. It does not define any classes as disjoint. Because of this, both of the above examples are valid instances of the BIBFRAME vocabulary, with no loss of information:

The above examples illustrate that the storage of descriptive elements in separate work and instance units is not required in BIBFRAME. Whether you keep work and instance separate or not can depend on your needs. In fact, at times data can be stored without separate works and instances, yet can be transmitted as separate BF entities when desired because the entities can always be created from a properly defined RDF vocabulary. This is an important lesson about RDF, especially as compared to the data models that we are most familiar with: in RDF, the meaning of an element is in the defined vocabulary, not in a record structure. You can define data that adheres to the concepts of work and instance, or even of work, expression, manifestation, and item, without that predetermining the structure of your data. Whatever view of your data you work with at any given moment depends only on what works best for you for that function.

ALTERNATIVE MODELS USING FRBR CONCEPTS

In previous chapters we looked at how the reality of FRBR is different from its stated goals: both the goals that led to the creation of the FRBR Study Group and the goal presented in the FRBR Final Report itself. Yet there is no question that FRBR, and in particular the Group 1 entities, resonates with many people both within librarianship and outside of it. We see a general recognition that the bibliographic resource is a complex thing that can be approached from a number of different points of view. This, at least, seems to garner wide agreement. What is not agreed upon, however, is a single interpretation of that complexity. It seems that many have a need to express the complexity of what they describe bibliographically, but those needs have a great number of potential expressions.

Those who have made use, in their own ways, of FRBR concepts have employed them to reinterpret the bibliographic entity as a cascade of abstractions, from the most abstract work through some number of intermediary levels until the actual physical item is described. The number of variants of this path, however, seem to be without limit, nor do the multi-entity models even agree on the nature of the levels of abstraction. FRBR itself is a shallow model, with a set of entities but no sub-entities. In some cases, those who borrow FRBR concepts extend the model significantly beyond its flat nature. In other cases, the model is reduced in the number of entities while at the time it is given greater depth through the creation of a hierarchy of entities.

Because it is a conceptual model, there is nothing in FRBR that has not been treated as open to reinterpretation, not even the seemingly uncontroversial concept of physical item. Where one draws the line between the physical and bibliographical is not as clear as you might think. FRBR, the library, and archive-based models often combine physical description (e.g., dimensions of the package) with a description of what is printed on the package itself. This is a mixture of information that harks back to the catalog record that presents a single line or field that includes extent, illustrative matter, and size ("xiii, 368 p. : ill. ; 24 cm.") Arguably a more logical separation would focus solely on physical properties. However, defining what is relevantly physical regarding digital resources, is not easy, and is even more difficult in a mixture of hard copy and digital resources,.

Each of the models presented below is a response to FRBR with some alterations. Given that FRBR presents itself as a conceptual model (at least in the text of the document), each of these uses the conceptual model and proposes a direction or implementation of it under different assumptions or using different technologies. Included here is just a selection of models that have riffed on FRBR's melody, or have come to a similar conclusion independently. Each of these confirms some aspects of the multi-entity bibliographic model, while drawing into question some others.

FRBRcore

The first development of the FRBR conceptual model in RDF was done in 2005 by Ian Davis and Richard Newman, in an effort that was not supported by the

IFLA group. This version of FRBR was called "FRBRcore" because it included the entities and relationships of FRBR but not the attributes. This may seem odd, as the attributes seem to have the greatest importance because they carry the descriptive information about bibliographic resources. This approach, however, makes sense if the role of classes in RDF is understood. RDF classes are similar to entities in a conception entity-relation model, and therefore entities are often designed as classes when models are interpreted in RDF. The classes in FRBRcore provide a conceptual framework that can be extended as needed. As we'll see when we discuss alternate models based on FRBR, a single conceptual framework can be used as the basis of some very different solutions. FRBRcore makes those solutions possible.

Davis and Newman were familiar with library models but they also were involved in RDF implementation. Although they were the first to develop an RDF version of FRBR, their vocabulary was never endorsed by the IFLA Working Group on the Functional Requirements for Bibliographic Records. Indeed, FRBRcore varied some from the E-R model in the FRBR Final Report. Davis and Newman added a super-class that encompasses the entities of Group 1, which they called "endeavor." They also created super-classes for Groups 2 and 3, "responsible entity" and "subject," respectively. In addition, they assigned a super-class, "spatial thing," to both object and place, because both of these could require geo-location attributes. The class structure in RDF is used in applications to address data at different levels of specificity. For example, the class "responsible entity" becomes shorthand for "person or corporate body" in program functions. The total number of entities in FRBR are relatively small, but in many RDF vocabularies the classes outnumber the descriptive elements, and are vital for efficient processing of the data.

FRBRcore could be considered experimental in nature, and is definitely not intended to be complete. The authors appear to have added a few elements intended to illustrate how FRBRcore could be extended. As sub-classes of work, they defined ClassicalWork, LegalWork, LiteraryWork, and ScholarlyWork. Each of these is a type of the more general concept of work. This extension of the work in FRBRcore may be surprising to those accustomed to library bibliographic data, because nowhere in FRBR or in any of the cataloging rules is there a discussion of types of works similar to those devised by Davis and Newman. The nearest thing is the FRBR attribute form of work, which includes as examples "play, poem, and novel." FRBR form of work isn't the same as the FRBRcore subtypes of work; however, either or both could be the basis for extending the work concept. In fact, a concept like work could be extended in a number of different directions, because in RDF classes are not exclusive. There is also no limitation in RDF on the creation of a super-class that is a generalization of one or more classes. That FRBRcore has the super-class “endeavour” does not mean that the FRBR entities of Group 1 are no longer valid or that their meaning has changed.

FRBRcore is highly visible today in the linked data cloud. The Linked Open Vocabularies project reports that FRBRcore is used in twelve vocabularies, and appears in nearly 30 million instances, although most uses are from the union catalogs of Bavaria, Berlin, and Brandenberg, and the union catalogs of Hessen and parts of the Rhineland. These catalogs account for about 24 million instances of FRBRcore in RDF triples.

FRBRoo’s Object-Oriented Model

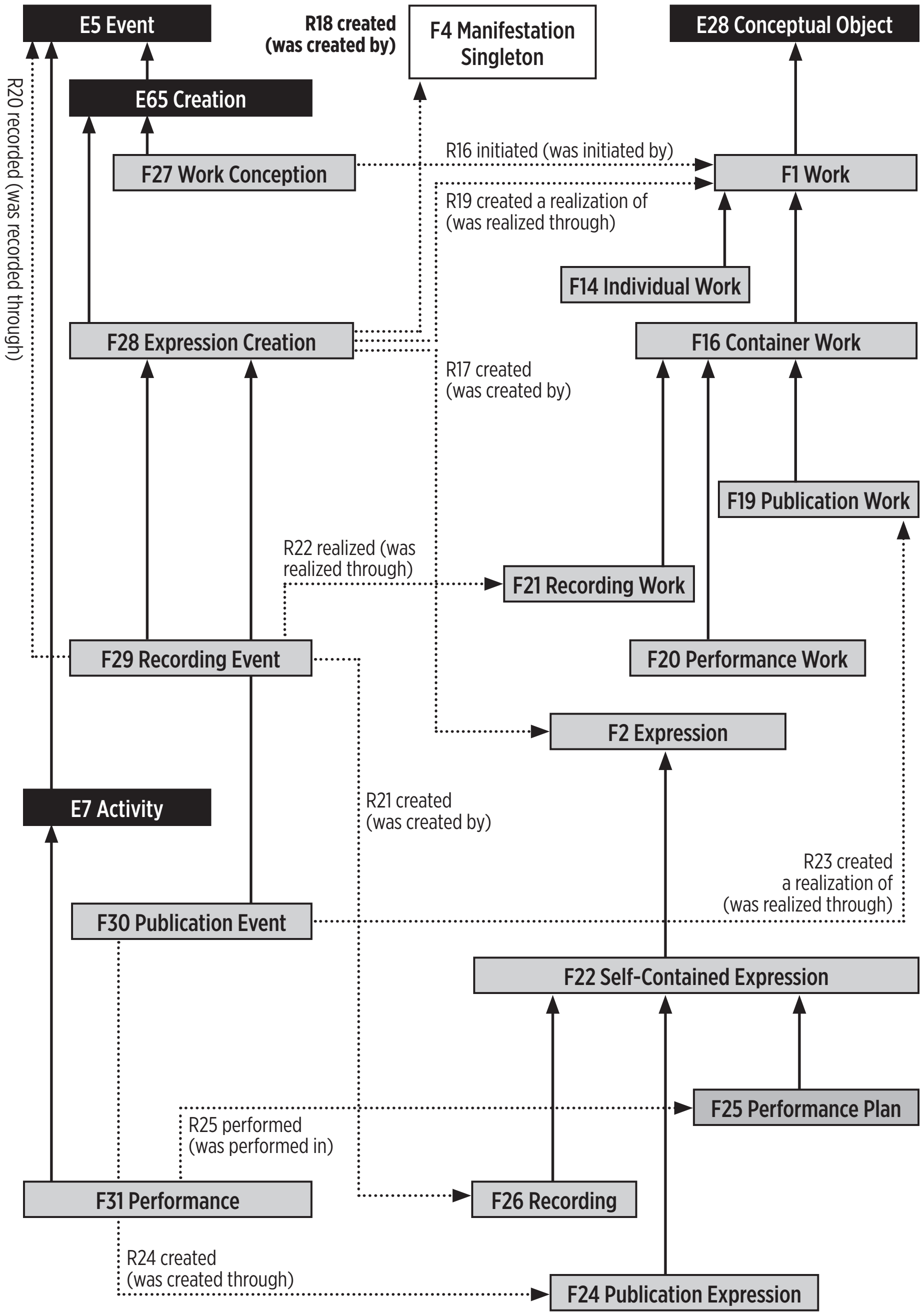
FRBRoo is a harmonization of the FRBR entity-relation model and the Conceptual Reference Model of the International Council of Museums (CIDOC CRM). It is, as its name indicates, an object-oriented data model, but the vocabulary has also been defined in RDF. FRBRoo is not intended as a replacement for FRBRer, but as an interpretation for an object-oriented environment and a harmonization with museum practice. The FRBRoo document explains that its project also serves as a proof of concept of the FRBR model:

> Expressing the FRBR model in a different formalism than the one in which it was originally developed provides a means to evaluate the model in terms of its internal consistency. It is also a good opportunity to correct some semantic inconsistencies or inaccuracies in the formulation of FRBR that may be regarded as negligible when FRBRER is only used in a library catalogue context, but that prove to be quite crucial from the moment one strives to design an overall model for the integration of cultural heritage related information. (Doerr 2006)

Object-oriented models define both things and processes, while the E-R model used in FRBR is a static definition of entities and relations. However, some of the relations in FRBR (e.g., “manifests” or “expresses”) imply some action, and FRBRoo has taken an action or event-oriented view. FRBRoo includes an event for each of the WEMI concepts, such as Work Conception, Expression Creation, Carrier Production Event, and Publication Event. It recognizes the importance of these events and the role played by the various actors that are involved in the transformation from a creative concept to something shared with others.

FIGURE 10.2

A segment of the FRBRoo model—From Work to Expression, dynamic view



> FRBRER envisions bibliographic entities as static, ever-existing things that come from nowhere, and overlooks the complicated path from the initial idea for a new work in a creator's mind to the physical item in a user's hands through the dramatically important decision-making on behalf of publishers. As a matter of fact, bibliographic records do contain implicit information about that complicated path and the relationships it implies between and among bibliographic objects; FRBRoo digs that implicit information out of bibliographic structures, e.g. the precise meaning of "date of publication." (Doerr 2006)

FRBRoo also extends the FRBR entities using super- and sub-classes, as can be seen in figure 10.2. For example, there is an abstract class super to work and expression called "conceptual object," and sub-classes under work itself for types of works, such as "individual work," "complex work," and "publication work."

Significantly, FRBRoo recognizes publication as an action, and treats Publication Work as an entity sub-classed to Container Work. This at least partially responds to the difficulty that the FRBR Review Group had with aggregates, and smooths the transition from the expression to a publicly available package that has physical and intellectual characteristics that are added by a publisher but that are not included in the expression itself. In FRBRoo, every publication is a kind of aggregate, because it always contains some aspect of creation added by the publisher. This makes a publication, by definition, a package of multiple creation activities. Works that are not published, such as works of art, can be manifested without the intervention of a publisher, and therefore do directly manifest what has been expressed. In the 2006 version of the FRBRoo documentation, edited by Martin Doerr and Patrick Le Boeuf, published works were likened to car models in their relationship between the original creator and what comes out of a manufacturing process.

> "Manifestation" can be two completely different things: Either it is an industrial product, i.e., a Type, like a particular car model, or it is a Physical Man-Made Thing that was produced as a unique carrier of an Expression. Industrially printed books belong to the first category, and are indirectly related to the main author's original creations. (Doerr 2006)

This echoes some of my own concerns about FRBR's treatment of the manifestation as being in a direct line from the expression of the work, without recognition of the many aspects of the published resource (book, sound recording, or film) that are contributed by the publisher.

> In the course of discussion however it was recognized, that virtually any book is composed of multiple, distinct works: the text, the illustrations, the editors work on lay-out, type phase etc. The latter was widely ignored in FRBR. . . . This situation demanded for a general model explicating both the individual contribution and the unity of the integrated product. (Doerr 2007)

FRBRoo appears to have a consistent grounding in object-oriented technology, and to be shepherded by a group that understands that technology. That said, FRBRoo is extremely complex, and understood by few. If it is to be adopted it will need a user-facing interface that is comprehended by a wide range of metadata producers.

<indecs> Event-Oriented Model

The <indecs> metadata model was developed in the late 1990s, and the current revised version is from 2000. <indecs> models metadata for intellectual resources around an e-commerce viewpoint, but intends its design to interoperate with the full product flow from the original creator, through the publishing and manufacturing steps, to outlets like stores and libraries, and then to the end-user.

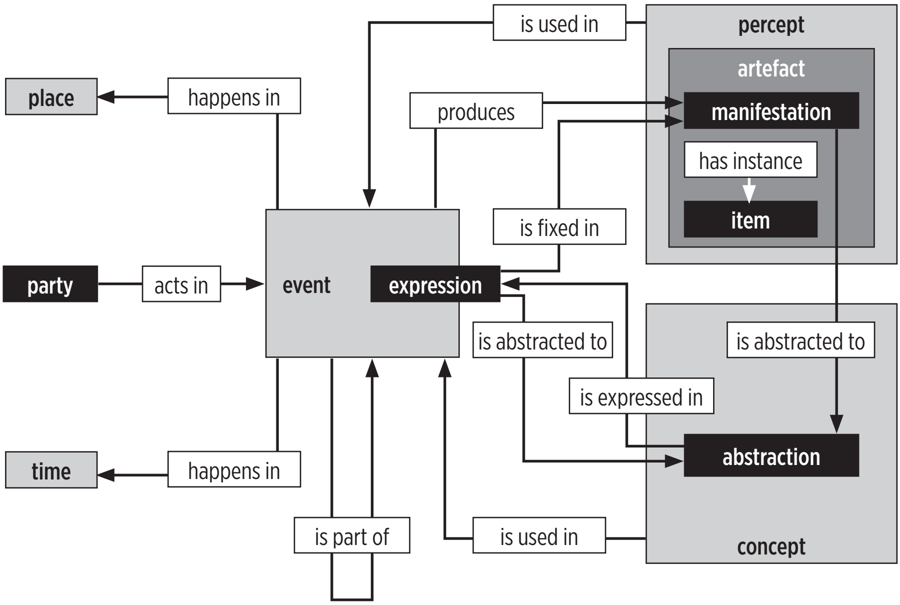
<indecs> takes an event-based view of the metadata model, and defines metadata as a "relationship that someone claims to exist between two entities." All relationships are events, either static or dynamic, and events have inputs, agents, and outputs. The agents are key in the <indecs> model because these inform the question of intellectual property rights, which are one of the main elements of commerce.

The treatment of each entity as being the result of an event with the involvement of an agent resolves some of the questions about the nature of the FRBR Group 1 entities: where FRBR has works, expressions, and manifestations as primary entities with flat relationships and no intervening human activity, <indecs> includes the persons or agents that act to create the entities (figure 10.3). This answers the question: how does a work get expressed, and by whom?

<indecs> also recognizes that different media and different products can have different events. In particular, the performance event for music produces types of expressions that are not common for textual works. <indecs> places no limits on the types of relations that can be included in the model, although it lays out a general framework of metadata properties.

FIGURE 10.3

The <indecs> model of bibliographic metadata



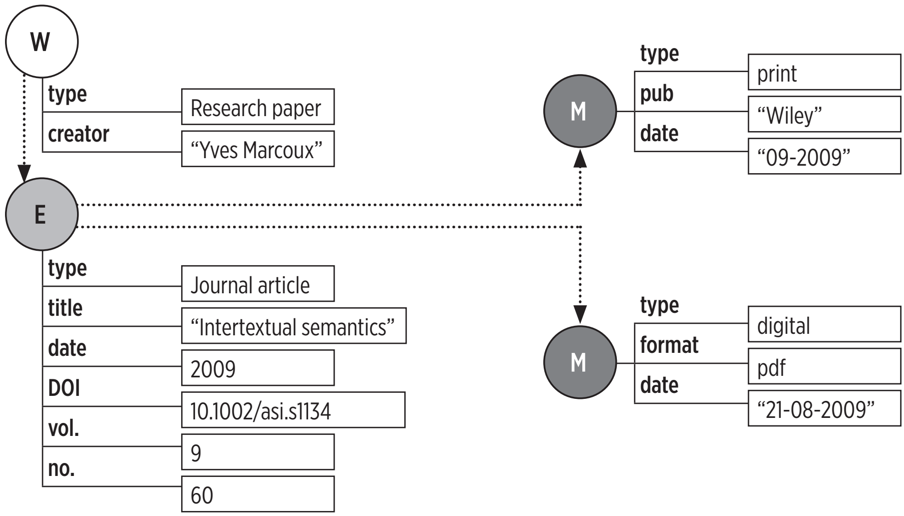
FaBiO FRBR-Aligned Bibliographic Ontology

FaBiO is one of the vocabularies defined in the Semantic Publishing and Referencing (SPAR) suite. This vocabulary development project has scholarly publishing as its main focus, with coverage primarily aimed at works that are published, textual, and/or referred to by bibliographic references. It defines its own set of entities that are sub-classed to the original FRBR entities (work, expression, etc.) using the FRBRcore RDF vocabulary. FaBiO's main approach to the FRBR entities is to use them as classes for types of scholarly publications. FaBiO sub-classes the FRBRcore work class with almost thirty different subtypes, including biography, reference work, dataset, and sound record. The FRBRcore expression class yields over fifty sub-classes, among which are chapter, editorial, presentation, spreadsheet, and Gantt chart.

In keeping with its approach of creating sub-classes of the FRBR entities defined as classes in FRBRcore, a manifestation in FaBiO is defined as having three sub-classes: analog manifestation, digital manifestation, and manifestation collection. The model is illustrated in figure 10.4.

FIGURE 10.4

FaBiO model showing expression and manifestation examples



Along with the classes derived from FRBR entities, FaBiO has dozens of properties for bibliographic description, few of which would be considered exact equivalents of descriptive elements in library data. There are some common bibliographic properties like title, publisher, data of publication, pages, identifier, and language, but FaBiO also has an extensive list of dates relating to the workflow of academic publishing, such as "has submission date," "has embargo date," and "has date received." FaBiO does not, however, align use of bibliographic properties to specific sub-classes of FRBR's WEMI. Instead, the descriptive properties are associated with the super-class that represents the bibliographic resource as a whole.

In addition to the extensions of FRBR provided by FRBRcore, FaBiO creates direct links between works and manifestation, works and items, and expressions and items.

Where FaBiO most notably diverges from a library cataloging interpretation of the entities, though, is in its emphasis on the expression. Like the expression-dominant model of Shoichi Taniguchi, FaBiO interprets the manifestation as a physical carrier, with all of the content description properties being associated with the expression. This model defines different digital formats of the same publication or manuscript in the item entity, which is a possible resolution to the situation that libraries face with electronic books and other electronic resources where only the digital format varies.

Summary

FRBRer, FRBRcore, FaBiO, <indecs>, BIBFRAME, and RDA in RDF are variations on a theme. They have much in common, but each has its own approach. webFRBRer defines a strict closed-world interpretation of the FRBR model. RDA in RDF implements the elements of the FRBR-based cataloging rules. FRBRoo is highly sophisticated in its design but is so far unconnected to mainstream library cataloging. FRBRcore and BIBFRAME show promise as vocabularies for the open web, and FRBRcore has been the basis for the development of other bibliographic vocabularies. BIBFRAME, however, is taking on the unenviable task of carrying forward into RDF the centuries-old practices of traditional library cataloging and in particular that tradition as a coded machine-readable record, MARC 21.

None of these can be considered mature at this time, and the future of bibliographic data in RDF is still at an experimental stage. Although there is some speculation in the library community about which of these models will prevail, we have to consider the possibility that there will be more than one model in use. Already there are efforts to assure that bibliographic models under development promote interoperability, not only among the library, archives, and museum bibliographic models, but also with the worlds of publishing, academe, and the reading public.