A Reconsideration of Mapping in a Semantic World

Gordon Dunsire	Diane Hillmann	Jon Phipps	Karen Coyle
Consultant, Scotland	Metadata Management	Metadata Management	KCoyle.net, USA
gordon@	Assoc., USA	Assoc., USA	kcoyle@kcoyle.net
gordondunsire.com	metadata.maven	jonphipps@gmail.com	
	@gmail.com		

Abstract

For much of the past decade, attempts to corral the explosion of new metadata schemas (or formats) have been notably unsuccessful. Concerns about interoperability in this diverse and rapidly changing environment continue, with strategies based on syntactic crosswalks becoming more sophisticated even as the ground beneath library data shifts further towards the Semantic Web. This paper will review the state of the art of traditional crosswalking strategies, examine lessons learned, and suggest how some changes in approach--from record-based to statement-based, and from syntax-based to semantic-based--can make a significant difference in the outcome. The paper will also describe a semantic mapping service now under development.

Keywords: semantic maps; mapping; crosswalks; crosswalking; ontologies; libraries; library metadata; dumb-down; Semantic Web; Linked Data; Linked Open Data; LOD; RDF; RDFS; OWL; XML schema; RDA; FRBR; FRAD; FRSAD; MARC; ONIX; Vocabulary Mapping Framework; VMF; Dublin Core Metadata Initiative; DCMI

Introduction

Much of the impetus for crosswalking has occurred in the library community, only now beginning to emerge from the forty-plus years of building data based on the venerable MARC data format. Although the future for library metadata is still evolving, a great deal of recent discussion has focused on the enormous amount of legacy data that is managed and stored in legacy ILS systems. Other cultural heritage communities have seen the development of many different non-MARC formats such as VRA Core, and also find themselves facing similar issues.

The library community has put a great deal of effort into the creation of crosswalks that transform metadata between complex library-specific and more general-purpose schemas. Development of these translations between metadata formats increased in the late 1990s, when Dublin Core began to make inroads into areas where traditional library practices had little uptake, primarily in digital library projects attempting to create access to resources like images and archival materials not typically included in library catalogs. Another wave of crosswalking was instigated when the ONIX format was adopted by publishers for metadata that is also of interest to libraries.

We are facing a new round of discussion about translation as libraries consider replacing the machine-readable metadata formats they have used for the last half-century with something more suitable for the linked data environment. The vast bulk of library data is currently in one or more formats primarily based on ISO 2709, and we will refer to them here, for simplicity, as MARC. Using the concepts of crosswalking and data translation that have served us in recent decades, it is difficult for many traditionally trained librarians to understand how it is possible to intermingle library bibliographic metadata and metadata from other communities in a single data environment. This difficulty is in part based on the assumption that equivalencies between different metadata schemas must be done as in the past: with a literal translation of elements from one schema to another. Instead, in the linked data environment, "mapping" takes on an entirely different meaning. Although some of the discussion in this paper is relevant to general concepts of linked data, we focus on the specific environment which uses the Resource Description Frame-

work (RDF) data model along with RDF Schema (RDFS) and Web Ontology Language (OWL) vocabularies for use in that framework, operating in an "Open World" of data on the world-wide web. We also focus on mapping between format elements rather than content expressed as controlled value vocabularies, but recognize that mappings between values may be an important consideration for specific applications.

Crosswalking and Mapping

The RDF data model allows anyone to say anything about any thing, including relationships between any two identifiable things in the broadest possible sense. In this paper, we use the noun "mapping" to refer to a semantic relationship between metadata elements, and the noun "map" to refer to a set of such mappings that relate one schema to another. The Semantic Web allows us to focus on mappings rather than maps, on relationships between one RDF class or property and another.

In many discussions of crosswalking and mapping, the terms are used almost interchangeably. Early work on crosswalking by St. Pierre and LaPlant presents crosswalking as one tool in the attempt to 'harmonize' metadata standards to provide consistency between them.

> A crosswalk is a set of transformations applied to the content of elements in a source metadata standard that result in the storage of appropriate modified content in the analogous elements of a target metadata standard. A complete or fully specified crosswalk consists of both a semantic mapping and a metadata conversion specification. The metadata conversion specification contains the transformations required to convert the metadata record content compliant to a source metadata standard into a record whose contents are compliant with a target metadata standard. (St. Pierre & LaPlant, 1998)

They further suggest a requirement for absolute consistency that has echoed through all subsequent discussions of crosswalks:

> A fully specified crosswalk requires that all implementations of the crosswalk on a specific source content result in the same target content. If two different implementations of a crosswalk operating on the same source content result in a different target content, the crosswalk is not fully specified.

Much of the crosswalking research of the last decade or so focuses less on the creation of the maps and more on the resulting transformational process. In this traditional world, maps are developed, ingested and maintained as documents (usually spreadsheets) that are not actionable. Thus a further, but separate, step beyond the intellectual process of creating a map is the creation of programs that implement the mapping and transform data based on the decisions made during the creation of the maps.

Godby, in her discussion of the services built to transform ONIX to MARC, describes the complexity of the processes.

> The crosswalk is represented as a human-readable Microsoft Excel spreadsheet that can guide an implementation of software that translates from ONIX to MARC. Organized into fifteen worksheets, the most important worksheet is labeled *ONIX*. … subordinate worksheets … typically specify how ONIX codes may trigger maps to multiple MARC fields or spell out complex conditional logic that cannot be stated succinctly. (Godby, 2010)

More than a decade's worth of published research identified strategies that worked fairly well in closed environments, primarily meeting limited goals for transforming non-MARC data into basic MARC records, but a number of seemingly intractable problems had been identified. One issue was that 'equivalence' was the only relationship supported, so that any shades of gray created problems that could not easily be remedied.

> One of the problems of crosswalking is the different degrees of equivalency: one-to-one, one-to-many, many-to-one, and one-to-none … details may extend from elements-only to

elements-plus-qualifiers/refinements or sub-elements. However, usually only the names of the elements and their definitions are taken into consideration in a crosswalk." (Chan & Zeng, 2006)

This means that when mapping individual elements, often there are no exact equivalents. Meanwhile, many elements are found to overlap in meaning and scope. For this reason, data conversion based on crosswalks could create quality problems. (Chan & Zeng, 2006)

This discussion of 'degrees of equivalency' exemplifies one of the problems of mapping: the concept of equivalence is described in binary terms but then discussed in terms of levels or degrees of exactness. It is hard to translate this into data conversion algorithms that must make a yes/no decision about whether data can be effectively moved between elements 'mapped' to each other. This inability to effectively express a degree of exactness lies at the core of the 'quality problems' noted above.

Hub-and-spoke, or switch, mapping architectures, have been developed to address the problem of schema proliferation; they are inherently more economical because pair-wise or peer to peer mapping involves a combinatorial explosion as more and more schemas are involved. Instead, mappings are established between a central hub or switch vocabulary and the source and target elements, which are then related by joining the source-hub and hub-target mappings. Current approaches based on switch architectures are top-down; that is, the switch vocabulary needs to be complete before any mappings can be made. The introduction of a new entry to the switch vocabulary should normally trigger a review of all mappings in the semantic vicinity of the term, in case it provides a better fit or equivalence. This strategy is equivalent to pre-coordinated indexing in resource discovery systems.

An example of this model is the Vocabulary Mapping Framework (VMF) Project (VMFa). This is a switch vocabulary intended for resource relators and categories used in major standards from the publisher/producer, education and bibliographic/cultural heritage communities. The VMF attempts comprehensive pre-coordination through the use of concept families:

a family of classes and relators ... defined around a single 'verb' concept". The switch vocabulary is thus "the sum of all the concepts which are mapped into it, plus a large number of other intermediate concepts which are needed to create computable relationships between the mapped terms …The VMF ontology is then a matrix of Concept Families organized within a hierarchy. (VMFb)

This structure should ensure that there is always an exact-equivalence mapping from a spoke vocabulary term to the matrix. The matrix itself is a black box, with the ontology guaranteeing an internal pathway between any pair of switch terms, and thus between any pair of terms mapped to them. In fact, there will be many pathways between any pair of terms, and algorithms for computing shortest paths and corresponding "semantic distance" or partial equivalences are being developed. (Das, Chong, Eadon, & Srinivasan, 2004)

An issue with this approach is the number of pre-coordinate terms required in the switch vocabulary. The more terms, the more likelihood there is of finding an exact equivalence mapping from a spoke vocabulary. The VMF has 30,000 triples covering bibliographic relationships between agents and resources, such as "translate", "translator", and "translation". There is currently no coverage of detailed bibliographic attributes such as "equinox" and "foliation", or of refinements such as "parallel title", which are likely to be the concepts required for mappings.

The top-down approach will not be suitable for all communities. There may be delays in reviewing, approving, and integrating additional classes and attribute and relationship properties in the hub, and processes may not be flexible enough. A bottom-up approach would be a useful complement, in which mappings are neither pair-wise nor involve a switch vocabulary with internal completeness and coherency. Rather, a mapping is made to an appropriate existing RDF graph using OWL equivalence properties, as with a switch vocabulary in the top-down scenario. However, that graph need not, in itself, contain a mapping to the target vocabulary. Instead, the graph is "mapped" or connected to another graph, and so on until a graph containing the target

concept is reached. The hub is replaced by an ad hoc set of connected graphs, as shown in Figure 1.
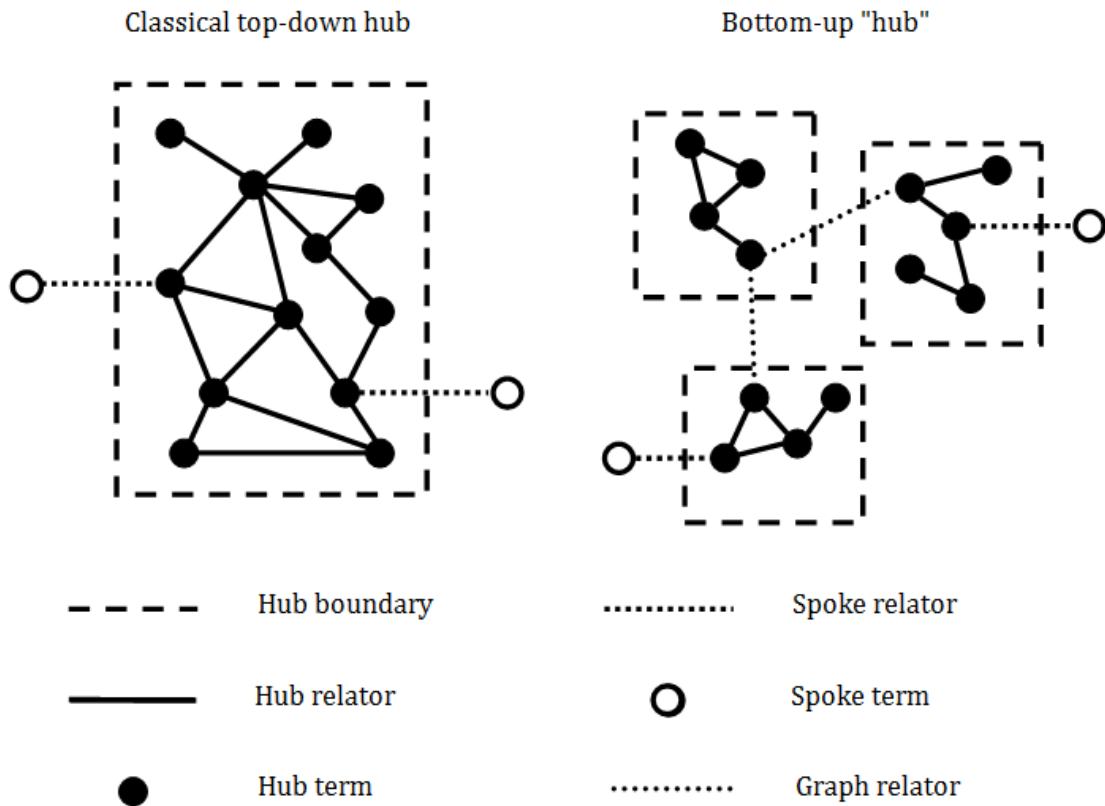

FIG. 1. Hub-and-spoke architecture diagrams

Godby and her OCLC colleagues experimented early on with strategies where crosswalks were created to and from a central switch vocabulary to provide clearer separation between syntax and semantic mapping. (Godby 2003) This work provides an important window on the extensive investment made by OCLC in providing crosswalking services, particularly between publishers and libraries. Her most recent work anticipates changes in the environment to shift her efforts as well:

> But even if the library community moves toward a more modern standard, there will be a need for robust crosswalks to ingest the hundreds of millions of legacy records created in the library community and mine the knowledge contained in them. This process will require a metadata model that extracts elements one at a time and recombines them, putting them to previously unanticipated uses, much as ONIX elements are deployed now to process transactions in the publisher supply chain. (Godby, 2010)

Whether the lessons of the past decade will hold true in a linked data environment that is more focused on statements rather than records, where data is unlikely to always be exchanged in record packages or from known sources, has not yet been addressed. This paper will focus on the intellectual process of creating relationships, not the transformative processes built on the maps themselves; we will not use the term 'crosswalking', keeping that term for the combined approaches of the past, but use 'mapping' in pursuit of a forward looking discussion.

An Environmental Shift

The meaning of "mapping" changes radically on moving from a database and record based approach to an open, multi-domain, global, shared environment based on linked data technologies --

where anybody can say anything about any topic, validity constraints are not acknowledged, a nearly infinite number of properties can be defined to describe an infinite number of entities, and authority is multi-dimensional and often ephemeral. The classic approach to such apparent chaos is to attempt increased control, increased filtering, increased restrictions, and limited access. This approach hinders appreciation of the broad diversity of perspective that comes with a world of open data.

Traditional databases are designed primarily as private data stores so metadata formats can be individualized for a particular application; the shared open data environment must accommodate community data views for all metadata created. In the domain-specific database environment, mapping to other data is related to data exchange, and requires an actual translation of elements from one metadata format to another. In the linked open data environment, mapping is inherent in the definition, redefinition, and refinement of properties and vocabulary elements. While actual migration of values from one property to another is certainly possible, it does not need to take place because any given property or element can maintain multiple simultaneous relationships to any number of other properties. In addition, relationships between properties can have their own distinct semantics. In traditional transformative mapping, elements must ultimately be treated as equivalent since the *values* of one will be assigned to the *properties* of another. In an open, non-transformative world, mapping applications can choose to treat all relationships as statements of equivalence or near-equivalence, or they can make use of more complex relationships in their applications.

This view does not obviate the need for attention to value spaces, often presented as text strings. Even when property and class relationships can be handled with the semantic fluidity described above, text strings as values dead-end the chains of linkages inherent in object properties. The semantic mapping of data properties to related object properties supports effective normalization of forms and the general transition from text strings to URIs, particularly an issue with the transition from MARC to RDA.

Despite the fact that some metadata developers do not think of the relationships as a form of mapping, the DCMI 'dumb-down' principle (Woodley) is an example of a pre-defined semantic mapping that retains great value in a mapping environment, particularly when the level of granularity between schemas is very different. One of the core principles behind the notion of dumb-down is a requirement that the more refined of two related properties must be considered to be a subset (subproperty) of the more broadly defined property. Even though the loss of specificity may mean less clarity of understanding, the data does not lose meaning when 'dumbed down' to its broader relative. Of equal and perhaps even greater value, broader properties defined by Simple DC may be 'smartened up' to provide increased specificity when possible. An example of a process where this occurs is described in the early implementation of the NSDL Metadata Repository (MR):

> The MR harvests a simple DC record from Provider A containing [DC] unqualified Title, Identifier, Creator and Type elements. The safe transforms done at the MR recognize the Identifier as a valid URI, so the URI encoding scheme is added to the Identifier element, and similarly, the Type value is a valid DCMIType, so the DCMIType encoding scheme is added to the Type element. (Hillmann, Dushay & Phipps, 2004)

From the point of view of mapping, the transformation process described above is inherently aware of the semantic relationships between Simple DC elements, and the refinements now known as DC Terms (formerly Qualified DC). These relationships, explicitly encoded in the Dublin Core Registry and on the DCMI website [DCMI metadata terms], represent and expose the intellectual effort carried out by the Dublin Core Usage Board, which maintains the DC vocabularies, and has a strong incentive to explicitly provide the capability to 'dumb-down' data to Simple DC as well as 'smarten-up' data, as the NSDL demonstrated, by using the same relationships.

The fact that the DC Usage Board chose to explicitly define the property/subproperty relationship between elements of the two vocabularies does not make that definition any less of a semantic 'mapping' in the context within which we define mapping. That relationship could have been defined by anyone at any time, although with substantially less authority, and would be no less valid in the context of the Global Web of Data.

Data Refinement as Mapping

In an RDF-based data environment, mapping may take place not as data exchange but as data refinement. For example, the RDA property labeled "Preferred title for the work" can be treated as a refinement of the DC property labeled "Title". Useful mapping opportunities depend on clarity of both the formal semantics and the human-understandable definitions of a property or vocabulary term. For example, RDA's "Title proper (Manifestation)" and ISBD's "has title proper" have near-equivalent definitions but different domains of Manifestation and Resource, respectively, whose semantic relationship is, as yet, not defined and remains unclear. This is a change for metadata developers because it requires them to consider the overall environment when developing a schema, with mapping capabilities as part of the process. It also means anticipating, where possible, future mapping needs. This effort to create data that "plays well" in the environment of the Semantic Web puts pressure on developers to think more broadly than in a closed database environment, but the end result could be metadata more clearly defined on its own and in relationship to other metadata.

Defining properties that are intended to be mapped to more (or less) expressive properties is particularly helpful as a modeling exercise. An example of building-in mapping capability is described by the DCMI/RDA Task Group in its discussion of the RDA Vocabularies:

> Where there is no inherent FRBR entity assignment in the name of a property, two properties were created for the RDA elements: first, a general property with no explicit FRBR assignment; and second, a FRBR-bounded subproperty. The general property carries no specific association with a FRBR entity and can be used by any application that determines that it is useful in its context, whether or not the application is based on FRBR. These general properties are fully compatible with the Semantic Web and not specific to library applications. In addition, the general property can be used to extend RDA by associating the property at an application level with a different FRBR entity, other than the one chosen by the RDA developers. (Hillmann, Coyle, Phipps & Dunsire, 2010)

The generalized properties are perceived as an important asset in providing mapping functionality, given that strongly FRBR-associated properties would not be useful when the entity associated with the property could not, or should not, be inferred by consuming systems to be a member of a FRBR-defined class.

Mapping through linked data semantics does not have the same result as mapping through crosswalks. Crosswalks are generally managed through applications that not only translate values stored in one metadata schema to another, but also transform the values themselves, as mentioned in the NSDL example above. In the new environment, without the necessity of considering a transformative end process, the idea of one 'best' map (or collection of maps) no longer seems relevant. Limiting mappings to equivalences -- in an RDF or OWL context being limited to 'sameAs' relationships--also seems outdated. (W3C, 2009) A more useful approach would define a mapping strategy that is open, extensible, and built using technology that goes beyond the spreadsheet.
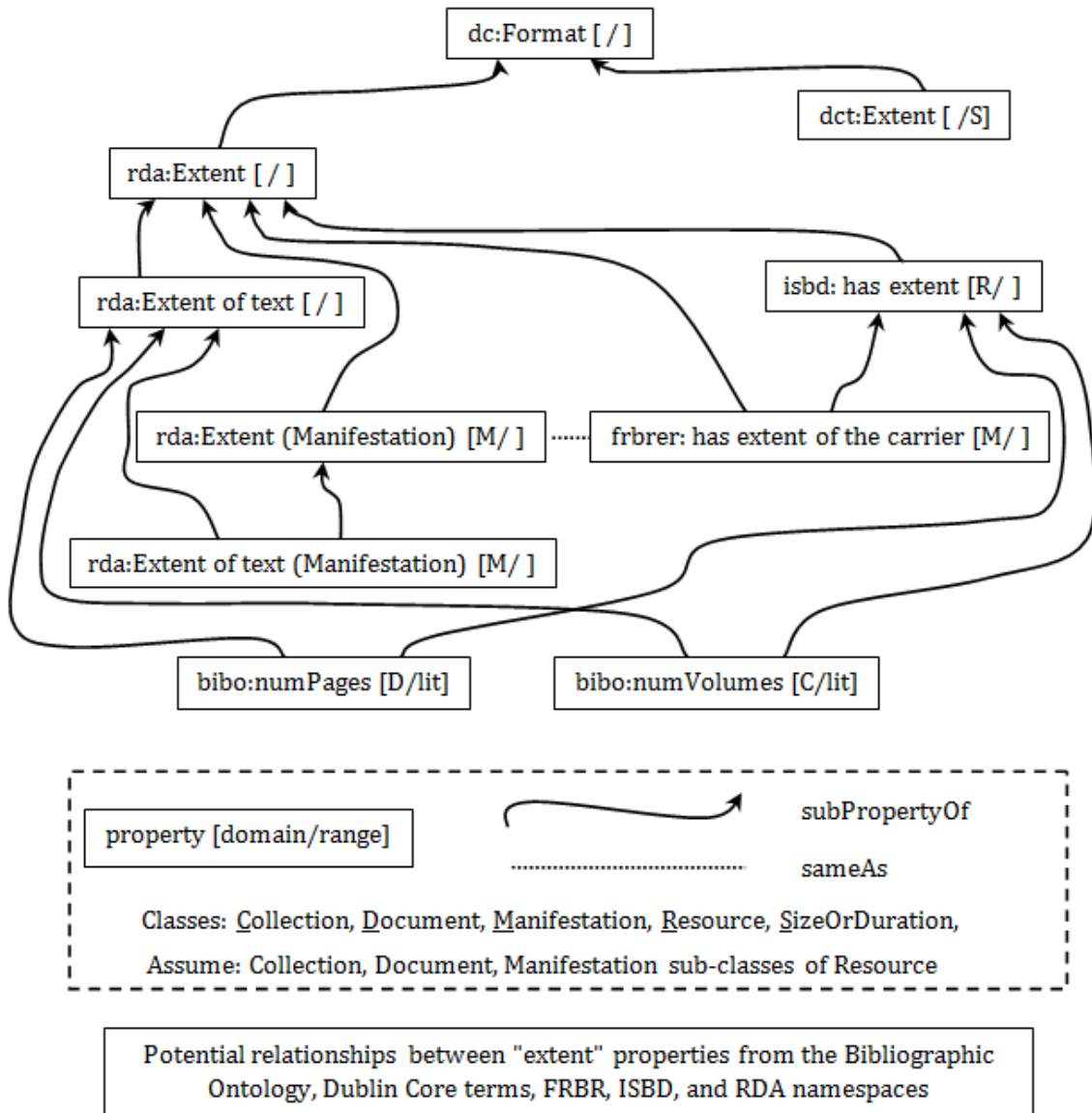
FIG. 2 Namespace relationships

Figure 2 shows some possible ways of linking properties related to the attribute "extent" from several namespaces for bibliographic description. Properties are linked by the owl:sameAs and rdfs:subPropertyOf properties. This web of mappings might be created by linking each property, one by one, to create an RDF graph, or by linking two or more smaller RDF graphs. A toolkit to support this activity needs to provide namespace element set classes and properties, the RDFS and OWL properties to link them, existing graphs, warning of inappropriate choice of linking property based on the element domain and range, and an interface to manage the mappings.

In an RDF-based data environment, transformation can still take place within an application, but as a separate operation from mapping. Transformation performed in a semantically mapped environment removes some of the anxiety surrounding potential degradation when round-tripping the contents of data elements because it is done without loss of information about the original semantics. Without the necessity of defining an 'authoritative' or 'best' mapping, a metadata el-

ement can have more than one set of semantics at the same time; this means it should be a simple matter to move from different but compatible definitions as needed within an application.

For example, there are now at least two RDF representations of the FRBR model. The first was developed by Davis and Newman in 2005 (Davis & Newman, 2005); it covers the entities and relationships between them, but not the entity attributes. It also includes relationships to external vocabularies. The second representation was published by the FRBR Review Group in 2011 (FRBRer Model), and covers attributes as well as entities and relationships, but with no reference to external vocabularies. Although there are many points of equivalence between the resulting RDF classes and properties, there are also significant differences. For example, the earlier representation includes classes for each of the FRBR "groups", such as Endeavour, while the later representation does not, on the grounds that the groups are not intended to indicate shared characteristics of the entities they contain, but are a mechanism for simplifying the model's entity-relationship diagram. This is not apparent from the source documentation, and required clarification by the FRBR Review Group.

The earlier FRBR RDF classes and properties have been used in several projects, such as Linking Petterson (Massey). The newer representation is likely to be used in work related to RDA, either directly or via equivalent RDF representations in the RDA namespace, not least because the FRBR Review Group is charged with maintaining and developing the model and its representation, and has a formal liaison with the Joint Steering Committee for Development of RDA. There is therefore a need for a map between the two representations, and possibly with the RDA namespace. Such a map can be created by anyone, at any time, and like the FRBR representations themselves, the degree of authoritativeness of the map will depend on its creator. There may be political issues if the creator is also responsible for one of the constituent namespaces: does the map somehow "authorize" all of the RDF classes and properties in the other namespace? Who decides how to resolve any semantic conflicts? Does the map constrain development of the creator's namespace? A map created by an independent third party with some recognized authoritativeness might avoid some of these issues.

Additional issues arise from the need to manage ongoing change. For example, the FRBR Review Group intends to consolidate the existing three models in the FR family: FRBR itself, and Functional Requirements for Authority Data (FRAD), and Functional Requirements for Subject Authority Data (FRSAD). FRAD and FRSAD extend and expand the model to fill gaps noted in the original FRBR analysis, but were not published until over 10 years later. Preliminary work has identified a number of FRBR elements that have been semantically modified by the later extensions, for example the class Corporate Body. Although the RDF representation of FRAD re-uses FRBR classes and properties as appropriate, and FRSAD re-uses FRAD, any semantic modifications will require mappings between corresponding elements of the models. Mappings will also be required between the separate models and the future consolidated model, and some classes and properties may need to be deprecated. Thus "semantic drift" between old and new RDF representations will occur, creating potential problems for applications using older versions and consumers of instance data based on them.

A Place for Services

The ability to create independent definitions of the semantic relationships embodied in a map, and the facility of the RDF data model to merge mapping graphs with ontological graphs does not imply a simple process of graph merger. There are issues of provenance and authorship of the map, version control and change management over time, the editorial and publishing cycle, management of group authorship and roles within the group including discussion and voting, and even evaluation of the validity of individual mapping statements based on the declared domains and ranges of mapping predicates. While some aspects of these requirements are being met by proprietary software, there is a tremendous need for open public services to provide assistance

and support for creators of semantic maps, and places where such maps can be registered and discovered.

One such toolkit, providing a set of open services to support the development and publication of group-managed maps, is in an early development phase for ultimate deployment as a service of the Open Metadata Registry (OMR). The planned toolkit would allow both individuals and communities to create a customized interface to facilitate the selection of the ontologic resources to be mapped. Properties from the same or different namespaces can be mapped using relationships such as property/sub-property or using relationships from appropriate ontologies. Mapping can also be used with instance data such as the assignment of FRBR entity relationships between Works, Expressions and Manifestations. In addition to the ontologic mapping facilities, interfaces will be developed that allow selection of resources based on the results of SPARQL queries, as well as selections from predefined sets of resources available from a chosen domain. Regardless of type, the mapping interface will detect domain and range mismatches between the mapped resources and the mapping relationship and warn the user of the potential for inappropriate inferences.

The service will support group development of maps by providing the ability to manage 'membership' of individuals and organizations and enabling assignment of distinct publishing roles to members such as map administrator, map editor, and map contributor with role-specific editorial abilities within a well-defined editorial workflow. Maps may be entirely private, with both the mapping interface and the resulting RDF map available only to group members, entirely public with all activities on display, or a combination of the two.

Discussions of mapping decisions that surround and enhance effective semantic map development, both within the group and optionally with the public, are facilitated by the ability to attach a discussion to a single mapping statement as well as aggregate those atomic discussions at multiple levels within the map. Both members and non-members of a group may subscribe to these discussions. Interested parties may subscribe to maps via email as well as subscribe to an RSS/ATOM feed to be notified of changes. This capability will be integrated with the OMR to inform ontology owners that a mapping relationship involving their ontology has been registered with the Mapping Service, providing owners with the ability to monitor and interact with the creators of maps that might ultimately affect the semantics of their ontology.

Once the Mapping Service is fully integrated with the OMR, ontology owners will be able to selectively incorporate independently-developed maps into the published versions of their ontologies and SKOS-based value vocabularies on-the-fly, taking advantage of the OMR's change-history tracking and versioning capabilities. Maps may also be visualized by map owners in conjunction with ontologies in the OMR to see the extent and nature of existing relationships.

Despite the relative sophistication of the interface, there is a commitment to avoiding any sense of data lock-in. At any point in the lifecycle of any map, public or private, the map will be available to the map owners as RDF (in several serializations), spreadsheet compatible formats, and of course (X)HTML. An API will be provided to enable programmatic interaction with other semantic registries to support such global interests as federated searching, distributed and offline development, inter-organization collaboration, and closed proprietary systems.

If all of this sounds vaguely familiar it may be because much of the design is based on previous experience with the development of the OMR. The basic goals are very similar: provide a set of open tools for a group or an individual to manage a potentially complex development process in an intellectually accessible way.

Conclusion

Some individuals and organizations in the library domain perceive Linked Open Data, and the attendant globalization of metadata creation and dissemination, as a threat -- a direct infusion of chaos into what was once perceived as an effectively ordered world. There have been many attempts to maintain that order and allow organizations to create data and metadata specific to their

domain while providing a framework for interoperability based on the translation of metadata from format to format. This has of necessity created a perception of interoperability based on the use of the common, shared syntax of public XML schema. The notion of interoperability based on shared schema has found a new home in the Semantic Web and its junior cousin Linked Data, but without a strategic reevaluation of a fundamentally different data model and related technologies. This has resulted in attempts to control the perceived chaos rather than to understand, harness, and ultimately benefit from it.

The possibilities presented by an effective transition to a new, open ecology of mapping based on the platform of the RDF data model and Semantic Web technologies developed over the last 10 years require a better understanding of the differences between traditional syntactic crosswalking and modern semantic mapping. Yet, clearly there are significant questions remaining that will require further experience and continued discussion that are beyond the scope of this paper. Among them are:

- What will be the relationship of Application Profiles, specifying how sets of data elements should be assembled into packages or "records" for particular applications, to this ecology of mapping? Will communities wish to designate mappings that reflect their metadata point of view?

- We see the value of mappings as independent ontological statements with visible authority and ownership separate from the originating ontologies. Is there a value too, to formal endorsement in this environment?

- How does the mapping of individuals in value vocabularies fit in? Can these techniques be applied to value vocabularies in a useful way? Is the value different or less?

- What is the value of metadata registries such as the Open Metadata Registry, id.loc.gov, the Dublin Core registry, and vocab.org, in this environment. Can tools based within those registries encourage the growth of this environment?

We look forward to being a part of that continuing conversation.

References

Chan, Lois Mai and Marcia Lei Zeng. (2006). "Metadata Interoperability and Standardization -- A Study of Methodology Part I: Achieving Interoperability at the Schema Level." DLib Magazine, v. 12, no. 6. Available at: http://dlib.org/dlib/june06/chan/06chan.html

Das, Souripriya, Eugene Inseok Chong, George Eadon and Jagannathan Srinivasan. (2004). "Supporting Ontology-based Semantic Matching in RDBMS". In Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004. Available at: http://www-scf.usc.edu/~csci586/paper/ontology-imp.pdf

Davis, Ian and Richard Newman. (2005). Expression of core FRBR concepts in RDF. Available at: http://vocab.org/frbr/core.html

DCMI Usage Board. DCMI metadata terms. Available at: http://dublincore.org/documents/dcmi-terms/

FRBRer Model. Available at: http://metadataregistry.org/schema/show/id/5.html

Godby, Carol, Devon Smith, and Eric Childress. (2003). "Two Paths to Interoperable Metadata" /International Conference on Dublin Core and Metadata Applications/[Online], 0 28 Sep 2003

Godby, Carol Jean. (2010). "From records to streams: Merging library and publisher metadata." *International Conference On Dublin Core And Metadata Applications*. Available at: http://dcpapers.dublincore.org/ojs/pubs/article/view/1033/989

Hillmann, Diane, Naomi Dushay, & Jon Phipps. (2004). "Improving Metadata Quality: Augmentation and Recombination." *International Conference On Dublin Core And Metadata Applications, 2004*. Available at: http://dcpapers.dublincore.org/ojs/pubs/article/view/770/766

Hillmann, Diane, Karen Coyle, Jon Phipps, and Gordon Dunsire. "RDA Vocabularies: Process, Outcome, Use." D-Lib Magazine, v. 16, no. ½, Jan./Feb. 2010. Available at: http://dlib.org/dlib/january10/hillmann/01hillmann.html

IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional requirements for bibliographic records. Available at: http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

Joint Steering Committee for Development of RDA. RDA: resource description and access. Available at:
    http://www.rda-jsc.org/rda.html

Massey, David. Linking Petterson: Visualizing FRBR data with Protovis.
    Available at: http://home.hio.no/~davm/foredrag/linking_petterson.html

Open Metadata Registry (OMR). Available at: http://metadataregistry.org

St. Pierre, Margaret, and William P. LaPlant, Jr. (1998). "Issues in Crosswalking Content Metadata Standards, October,
    15, 1998. Available at: http://www.niso.org/publications/white_papers/crosswalk/

Vocabulary Mapping Framework Project (VMFa). Available at: http://www.doi.org/VMF/

The Vocabulary Mapping Framework (VMFb): An introduction: v1.0, December 12, 2009.
    Available at:
    http://www.doi.org/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0%28091212%29.pdf

Woodley, Mary S. DCMI glossary. Available at: http://dublincore.org/documents/usageguide/glossary.shtml

World Wide Web Consortium (W3C). (2009). The Web Ontology Language (OWL). Available at:
    http://www.w3.org/TR/owl2-overview/