Jacob Jett, UIUC, jjett2@illinois.edu David Dubin, UIUC, ddubin@illinois.edu Bobby Bothmann, MNSU, Mankato, bobby.bothmann@mnsu.edu

To Map or Not to Map: Rethinking Crosswalk Agendas

Abstract

In the two decades since their publication, the Functional Requirements of Bibliographic Records and succeeding standards such as the Library Reference Model have had a marked impact on discourse concerning descriptive theory and practice. The BIBFRAME model, which began as an effort to replace MARC as a linked data-capable modeling format, offers an alternate view of the bibliographic universe with three principal entities rather than four. Differences between BIBFRAME and LRM are based in competing intuitions on the nature of creative works, and at first the two approaches appear to compete for the same intellectual space. BIBFRAME offers us a less constrained model of bibliographic descriptions than the FRBR models, and if interoperability between BIBFRAME and WEMI-aligned standards like Resource Description and Access requires translation of RDA records both to and from BIBFRAME descriptions, then the latter's flexibility poses problems for mapping between the models. Proposed solutions to those problems reveal as much about different modeling philosophies as they do about different views of creative works and their relationships to texts and copies. Linked data protocols are intended to support resources and scenarios that are far too diverse for either a single account of creative works or for a subsumption-based taxonomy of models. But a need for descriptions flexible enough to include them all does not require us to retreat from modeling commitments to either reductionism or operationalism. BIBFRAME can be seen as reaching for or pointing toward a descriptive domain that supports a complementary role to the IFLA standards.

Introduction

Twenty-two years ago, the International Federation of Library Associations (IFLA) changed the conversation of contemporary cataloging and knowledge organization practices with the first publication of their Functional Requirements of Bibliographic Records (FRBR [IFLA 1998]). In the two decades since its publication, FRBR and its Work-Expression-Manifestation-Item (WEMI) model of Group 1 Entities have had a marked impact on the discourse concerning descriptive theory and practice. It has shaped the Resource Description and Access (RDA) cataloging standard to a great extent (JSCD-RDA 2014).

We find another response to the needs of catalogers in the Library of Congress's (LoC) BIBFRAME standard (Miller et al. 2012).¹ BIBFRAME, which began as an effort to replace the MARC metadata standard as a linked data capable data modeling format, offers an alternate view of the bibliographic universe with three principle entities rather than four. With more precise WEMI definitions emerging in the twenty-one years since the original 1998 report, and to bring all of the FR-series of standards into a single aligned document, IFLA recently released a new standard—Library Reference Model (LRM [Riva et al. 2017]). In many ways, these continuous changes to cataloging approaches and conceptual models are an extension of an ever-evolving milieu of cataloging practices that find the roots in the mid-twentieth century.

The Nature of Creative Works

The BIBFRAME model and each of those presented in the FRBR family of specifications reflect competing intuitions about the nature of creative works. Does

¹ <u>https://www.loc.gov/bibframe/</u>

translating a novel from Russian into English produce a new means of access to the Russian author's work, or should we consider the translation a new derivative work? We want to hold on to our belief that reading works in translation is possible, but at the same time acknowledge that faithful translations require (per the language of LRM) a "significant degree of independent intellectual [and] artistic effort."

To what extent does the identity of a work depend on a particular narrative or expository structure? LRM and BIBFRAME define works as conceptual objects and conceptual essences, respectively. But concepts don't stand in rhetorical, narrative, or expository relationships the way that words, sentences, paragraphs, and chapters do. Concepts and propositions may stand in logical relationships with each other or in intentional relationships to objects in time and space, but they aren't arranged into stories or explanations. So, are creative works the "constellation of concepts" of LRM's E2, or do they include the stories, plots, and procedural descriptions that FRBR₀₀'s E89 superclass (Bekiari, et al. 2015) suggests?

Is the artistic content of a musical or choreographic work conceptual? Works of art resist the kind of straightforward paraphrase that a diagram in a technical manual might admit. Musical works typically admit some freedom of arrangement, but only within limits; not every composition can be performed using only harmonicas, for example.

When are the circumstances of a work's composition essential or incidental to it being the particular work that it is? If we uncover evidence that a pseudonymous text that we believed had been written during the first century CE was actually written during the second century, then have we gained a new understanding of the same work or should we conclude that the text realizes an entirely different work? The intellectual content expressed via the text is presumably the same, but our appreciation of its meaning and significance may be radically different (as, for example, with the discovery of a forgery).

Tensions such as these invite us to consider the purposes served by identifying bibliographic entities, the philosophical commitments that inform our standards for success, and (in light of those) to review particular recommendations for resolving apparent differences.

Evolutions in 20th Century Cataloging Practices

Anglo-American cataloging during the twentieth century experienced a continuous change from the overly complicated and "legalistic" rules at the beginning of the century decried by Osborn (1941), which paved the way for a pragmatic approach to cataloging, led by Semour Lubetzky. Lubetzky was responsible for much of the philosophy codified in the 1961 International Cataloguing Principles (ICP), or Paris Principles as they are commonly known, and the development of the Anglo-American Cataloging Rules (AACR). The ICP laid the groundwork for the development of the International Standard Bibliographic Description (ISBD), which in turn provided the framework for the development of the Anglo-American Cataloguing Rules.

Concurrent with the beginning of the use of the pragmatic cataloging rules found in AACR2 in North America, Elaine Svenonius and her doctoral students Barbara Tillett and Martha Yee were developing a theoretical framework for describing bibliographic resources based on an entity relationship model. Tillett's dissertation research (1987), in particular, became the basis for the Functional Requirements for Bibliographic Records (FRBR) theoretical model, developed and published by the International Federation of

Library Associations and Institutions (IFLA) in 1998, which would later become the framework for a new descriptive cataloging code, RDA: Resource Description and Access (RDA). The principles of description set forth by Svenonius in *The Intellectual Foundation of Information Organization* (2001) provide the philosophical underpinnings for RDA, chief among them the principle of representation.

For Svenonius, bibliographic entities such as works, editions, and author sets are fundamental because they are the primary objects in bibliographic languages, which is to say they are the intentional foci of bibliographic descriptions. But to define these entities, their attributes, and relationships Svenonius rejects descriptive accounts in favor of operational (i.e., constructive) definitions for both practical and philosophical reasons. She takes the position that defining concepts operationally enables a discipline to advance (page 4) and that collective agreements to accept empirical correlates of concepts that satisfy (some) of our intuitions provide a basis for resolving identity puzzles like those discussed above (page 12). As a demonstration, Svenonius proposes definitions of works, editions, superworks, author and subject sets each as a set of physical documents that stand in a relevant similarity relationship with the others in the set (e.g., sharing the same information or having a common origin).

The pragmatic approach of descriptive cataloging under AACR and AACR2 worked well for the chiefly analog resources that dominated during the 20th century. AACR2 provided some guidance for the description of computer files, but the rapid development of electronic versions of various media (text, moving image, audio, etc.) proved to be too complex for the format-based description bias in AACR2. By the time of the 1998 revision to AACR2 catalogers were struggling with the description of electronic resources over fundamental descriptive attributes such as the publication date, content versus format, and how this related to the overall work.

The impetus for change from AACR2 to RDA came not only from the growing number of physical and electronic formats, what we now term as carriers, but also from the nature of AACR2 in its 1998 revision. AACR2's "cardinal principle," as it was known, or Rule 0.24, instructed the cataloger to consider the "class of materials to which that item belongs." The effect of this rule was to set up a debate around the concept of "content versus carrier" (Delsey, 1998 3-7; Howarth, 1998). The tension between analog and digital description in AACR2 was part of the problem that the entity relationship model in FRBR was meant to alleviate. Goals for the new code were for rules based on principles and would include attributes for all types of materials, and would be easy to interpret (Tillett 2004, 10). RDA is designed to support digital and analog, traditional and nontraditional resources using well-formed data that can be managed by current and emerging database technologies (Oliver 2010, 2).

Further development of the functional requirements conceptual models necessitated a need for the revision of the 1961 ICP. A revision was published in 2009 (IFLA 2009); its revision was made in tandem with the development of the functional requirements models for Authority Data (FRAD) and Subject Authority Data (FRSAD). Where the 1961 ICP was written for textual resources, the revised 2009 ICP broadened the scope to all types of resources and incorporated the FRBR entities, thereby setting the emphasis of the "content vs. carrier" debate squarely on the "content" side.

In terms of the concept of a work, AACR2 part II focused on access points, much of which relates to identifying a work; the instructions are based on information from the

item in hand, raising the question of "whether the apparent anomaly in those cases is simply the result of inconsistency in following through on the principle or whether it is indicative of a more fundamental difficulty in operationalizing the concept of the work as an entity independent of the physical entity or entities in which the work is reflected." (Delsey 1998, 8). AACR2 part II was complex and probably most often consulted by those catalogers trained in the creation of name authorities, and seldom consulted by the generalist cataloger. While the work as a concept was something all catalogers described in the creation of a bibliographic description, it was not generally something that was a discrete action or easily identifiable in a MARC 21 bibliographic record.

Even now, the FRBR concept of Work is operationalized through a preferred access point for a creator in tandem with a title proper, or for cases with no creator, simply by use of the title proper. Seldom is a preferred title for Work, an RDA attribute distinct from the title proper, every explicitly used except when said preferred title of Work differs from the title proper. It is most common for works that have no creator as RDA requires each preferred access point to be unique. Moving image works, which have a long-standing practice of using the title proper for the preferred access point, also have a standardized practice of using appropriate qualifiers such as "Motion picture" or "Television program" as additions to the preferred title to provide unique preferred titles of Work. Similarly, in the history of serials cataloging, the addition of a place of publication or an institution name have served to differentiate works which otherwise have the same preferred title. While other types of resources often make use of preferred titles of Work when a creator is lacking, these tend to be less common than videorecordings and serials in many public and academic libraries.

Compatibility, Mapping, and Translation

Since BIBFRAME and LRM appear to compete for the same descriptive space, it's not surprising that people have proposed crosswalks for mapping descriptions between them. Godby (2013) presents a mapping between OCLC's extended version of the Schema.org ontology and BIBFRAME 1.0 (Miller et al. 2012). Zapounidou et al. (2016) make a comparison of FRBR, FRBR_{OO} (Bekiari et al. 2015), BIBFRAME 1.0, and the Europeana Data Model (Clayphan et al. 2014). Taniguchi (2018) showed a potential mapping between BIBFRAME 2.0 and LRM. In a series of papers Sofia Zapounidou and her colleagues compare these contrasting frameworks and explore methods for mapping among them (2017a, 2017b, 2017c, 2018, 2019). Most recently, Hahn and Dousa propose mapping between LRM and BIBFRAME works via a reduction to attribute sets, an approach they describe as "ontological and semantic minimalism" (2020). But do these mapping proposals really aim to connect alternate explanations of the same domain? Is loss-free translation between WEMI-based and BIBFRAME descriptions necessary for data sharing and system interoperability?

Crosswalks and switching languages have been designed and evaluated as means for connecting indexing languages since the 1960s (Dahlberg 1981; Svenonius 1983). Twenty years after the 1971 UNISYST Study Report on the Feasibility of a World Science Information System, Linda Smith reviewed recommendation 4 for joint efforts in developing better tools for index language conversion amidst information technology trends that prefigured the World Wide Web (1992). Smith notes developments in electronic mail and information retrieval systems that offered prospects for a "distributed

digital library with database exploration software." Although the trends in technology and the practice of science were clear at the time, Smith's characterization of indexing language compatibility (like the earlier reviews) still reflects a "closed world" view of databases, with each item having only one description (or a single level of description). Only four eventful years later, Lagoze et al. proposed the Warwick metadata framework (1996), reflecting the Dublin Core community's recognition that descriptive metadata for a single resource can take a variety of specialized and general forms, from disparate sources, serving the needs of diverse communities. The Warwick Framework proposal specifically addressed the relationship between DC and MARC descriptions, but with a broader view of descriptive compatibility that has become more familiar over the past quarter century.

Dublin Core and MARC descriptions clearly provide different degrees of specificity in the service of different needs and use cases, but can the same really be said for BIBFRAME and LRM? To even begin considering this question one should embrace Allen Renear's (2012) recommendation to take modeling seriously, rather than retreating to exclusively constructive, stipulative definitions or indulging a desire for semantic minimalism. It is modeling and data models after all that created the need for crosswalks and switching languages in the first place. And, much as Renear points out, the seminal minds in database theory in the 1960s and 70s were concerned that the databanks at the time, reflect the complexity of the world that their data described (Bachman 1969, Codd 1969, Senko et al. 1973, Chen 1976). Building formal accounts through conceptual analysis (Furner 2004) to arrive at models that describe the entities, events, and even roles (Bachman 1977) being played by both provides the cornerstone for information system design and development.

However, Svenonius, quoting Arthur Eddington, suggests that in facing up to the world's complexity, conceptual analysis threatens to plunge us into a "quagmire of indescribables," (page 33) as if the role of a conceptual definition were to resolve all competing intuitions, banish semantic infelicities, and solve the puzzles once and for all. Each of Renear's examples (inheritance across Group 1 Entities, modifiability of files, Europeana Data Model proxies, and roles vs. types in FRBR) highlights a narrow issue of logical consistency with practical implications for how systems act on records. The point of analysis is to expose one's model to critical scrutiny, not to try and represent an application domain without simplifications. As we wrote in 2013, data models for information interchange have both representational and cohortative agendas. They offer both explanatory accounts and plans of action, but as simplifications of reality they never capture domain richness with full fidelity. Analysis is carried out against the danger of having failed to recognize that two or more decisions are irreconcilable (see, for example, Jett and Dubin 2019).

If BIBFRAME and LRM may present models of different domains, rather than alternate accounts of the same domain, then where would we find evidence of that difference? We should examine what kinds of things the primary entity classes and their properties seem to be. Although the FRBR family of specifications differ to some extent on the nature of bibliographic entities, the focus of the models is (as the 1998 report authors write) "the key objects of interest to users of bibliographic data." The class of works, for example, includes Gray's *Anatomy of the Human Body*. Expressions are things like the original German text of Ellwanger's *Tennis—bis zum Turnierspieler*. Items are

physical copies of works, such as a particular autographed copy of Ronald Hayman's *Playback*.

Although some BIBFRAME vocabulary definitions suggest domain objects of direct user interest, many seem to reach for the domain of the bibliographic record, rather than entities, attributes, and relationships in a bibliographic domain. Consider these examples from the BIBFRAME 2.0 specification (emphasis added):

- Work: a *resource reflecting* a conceptual essence of a cataloging resource.
- Instance: a *resource reflecting* an individual, material embodiment of a Work.
- Agent: Entity associated with a resource or element of description, *such as the name of the entity* responsible for the content or of the publication, printing, distribution, issue, release or production of a resource.
- Arrangement: *information about* the organization and arrangement of a collection of resources.
- Subject: subject *term(s)* describing a resource.

A thoughtful reviewer of an earlier draft of this paper correctly points out that seemingly subtle differences between, for example, conceptual content and a *resource reflecting* a conceptual essence may only represent poorly worded definitions, rather than differences in intent. We note, however, that where LRM and BIBFRAME definitions are closely aligned (lrm:E5 Item and bf:Item) we have a physical resource that is directly available for a cataloger's scrutiny, while across the other definitional comparisons BIBFRAME tends to highlight levels to which a cataloger may attend more directly than the entity's LRM counterpart: a name rather than a person or organization, information about a physical pattern rather than the pattern itself, or a term rather than the topic it denotes. The BIBFRAME definitions seem to reach for a domain of bibliographic description rather than an ontology of abstract bibliographic entities.

We also observe that in the documentation for the BIBFRAME model, properties are "used with" classes and have "expected values," rather than defined domains and ranges. Many of those properties are "used with" both the Work and Instance classes, and some of those are also "used with" items. BIBFRAME apparently places fewer constraints on the classes of entities that participate in its relational properties.

It may be more productive to understand BIBFRAME works and instances connecting to LRM works, expressions, and manifestations via an intentional relationship (something like **foaf:focus** or **madsrdf:identifiesRWO**) rather than via an identity relation. Such an approach could offer more useful interpretations of the relative ease in deriving a BIBFRAME description *about* the objects of an LRM description (specific to general), as compared to linking in the other direction. Although the difference in descriptive generality would not be as wide as between MARC and Dublin Core, BIBFRAME's flexibility might be used to advantage in a broader range of use cases than those suggested by LRM user tasks such as finding, identification, and selection of materials. BIBFRAME could be a big enough umbrella for applications ranging from multiple rendition complexities of e-books to the cooperative analyses involved in creating scholarly digital editions.

One of our thoughtful reviewers correctly observes that BIBFRAME in its current form may even be too specific or *overdetermined* for this kind of interpretive adaptation. Such questions are for the BIBFRAME community of users, developers, and other stakeholders to resolve collectively. Our suggestions are intended to remind the reader that there are options for description language compatibility via relationships other than identity-based mapping, and that no one should feel reluctant to engage seriously with the descriptive agenda for information modeling.

We are extremely grateful to our anonymous reviewers for providing far more constructive and helpfully critical feedback on this paper than we're used to receiving, and to the Conceptual Foundations Research Group at the University of Illinois School of Information Sciences for their suggestions and encouragement at an earlier stage of this project.

References

- Bachman, C.W. (1969). Data structure diagrams. ACM SIGMIS Database: The DATABASE for Advances in Information Systems 1(2), pp 4-10.
- Bachman, C.W. (1977). The role concept in data models. In VLDB '77: Proceedings of the Third International Conference on Very Large Databases 3, pp 464-76.
- Bekiari, C., Doerr, M., Le Bœuf, P. & Riva, P. (eds.) (2015). Definition of FRBROO: A conceptual model for bibliographic information in object-oriented formalism (version 2.4) [revised 2016]. The Hauge: IFLA.
- Chen, P. (1976). The entity-relationship model: Toward a unified view of data. ACM Transactions on Database Systems 1(1), 9-36.
- Clayphan, R., Charles, V., & Isaac, A. (eds.) (2014). Definition of the Europeana Data Model v5.2.5. The Hauge: Europeana
- Codd, E.F. (1969). Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks, Research Report, IBM.
- Dahlberg, I. (1981). Toward establishment of compatibility between indexing languages. *Knowledge Organization* 8(2), 86–91. DOI: https://doi.org/10.5771/0943-7444-1981-2-86
- Delsey, T. (1998). Modeling the Logic of AACR. In J. Weihs, (Ed.) *The Principles and Future of AACR*. Ottawa: Canadian Library Association.

Dubin, D., Senseney, M., & Jett, J. (2013). What it is vs. how we shall: Complementary agendas for data models and architectures. Paper presented at *Balisage: The Markup Conference 2013*, Montréal, Canada, 6-9 August 2013. DOI: <u>https://doi.org/10.4242/BalisageVol10.Dubin01</u>.

- Furner, J. (2004). Conceptual analysis: A method for understanding information as evidence, and evidence as information. *Archival Science* 4(3-4), pp 233-65.
- Godby, J. (2013). The Relationship between BIBFRAME and OCLC's Linked-Data Model of Bibliographic Description: A Working Paper. Dublin, OH: OCLC Research.
- Hahn, J. & Dousa, T. M. (2020). Mapping bf:Work to lrm:Work and lrm:Expression: Towards a set-theoretical approach. Poster presented at the 83rd ASIS&T Annual Meeting, Pittsburg, PA, 25-29 October 2020 [virtual]. DOI: https://doi.org/10.1002/pra2.408
- Howarth, L. C. (1997). Content versus carrier. Paper presented at *International Conference on the Principles and Future Development of AACR* (Toronto, Canada, 23-25 October 1997).
- IFLA Study Group on FRBR (IFLA). (1998). Functional requirements for bibliographic records: Final report [revised 2009]. München: K.G. Saur Verlag.

- International Conference on Cataloging Principles (ICCP). (1961). Statement of Principles. Paris: ICCP.
- Jett, J. & Dubin, D. (2019). When conceptual models collide: Aggregates in IFLA's Library Reference Model. Paper presented at the 18th ACM/IEEE Joint Conference on Digital Libraries (JCDL'19) (Champaign, Illinois, 2-6 June 2019. Retrieved from: http://hdl.handle.net/2142/107003
- Joint Steering Committee for the Development of RDA (JSCD-RDA). (2014). *RDA: Resource Description and Access.*
- Miller, E., Ogbuji, U., Mueller, V., & MacDougall, K. (2012). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*. Washington, D.C.: Library of Congress.
- Oliver, C. (2010). *Introducing RDA: A Guide to the Basics*. Chicago: American Library Association.
- Osborn, A.D. (1941). Crisis in cataloging. Library Quarterly 11(4), 393-411.
- Renear, A. (2012). Taking modeling seriously. Presentation given at *Knowledge Organization and Data Modeling in the Humanities: An Ongoing Conversation* [workshop], Providence, Rhode Island, 12-14 March 2012. Retrieved from: https://datasymposium.wordpress.com/renear/
- Riva, P., Le Bœuf, P., & Žumer, M. (2017). *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. The Hague, Netherlands: IFLA.
- Senko, M.E., Altman, E.B., Astrahan, M.M., & Fehder, P.L. (1973). Data structures and accessing in data-base systems. *IBM Systems Journal 12*(1), pp 30-93.
- Smith, L.C. (1991.) UNISIST revisited: Compatibility in the context of collaboration. In Classification Research for Knowledge Representation and Organization: Proceedings of the 5th International Study Conference on Classification Research (Toronto, Canada, 24-28 June 1991).
- Svenonious, E. (1983). Compatibility of retrieval languages: Introduction to a forum. *Knowledge Organization 10*(1), pp 2-4.
- Svenonius, E. (2000). Bibliographic objectives. In *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press, pp 15-30.
- Svenonius, E. (2000). Bibliographic entities. In *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press, pp 31-51.
- Taniguchi, S. (2018). Mapping and merging of IFLA Library Reference Model and BIBFRAME 2.0. Cataloging & Classification Quarterly, 56(5/6), 427-54.
- Tillett, B.B. (1987). *Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging*. PhD diss., Los Angeles, CA: University of California.
- Tillett, B.B. (2004). "ALCTS Now That We've Revised Chapter 12, Where Do We Go From Here?" Program presentation for 2004 ALA Annual Conference Orlando. Retrieved from:

http://www.ala.org/alcts/sites/ala.org.alcts/files/content/events/pastala/annual/04/tillettc h12.pdf

- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2016). Representing and integrating bibliographic information into the Semantic Web: A comparison of four conceptual models. *Journal of Information Science* 43(4), 523-53.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017a). The representation of

bibliographic families in library data models and their preservation in mappings: The case of the mapping from FRBR to BIBFRAME. *Europeana Tech Insight 2017* 8 [unpaginated].

- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017b). Preserving bibliographic relationships in mappings from FRBR to BIBFRAME 2.0. In J. Kaamps, G. Tsakonas, Y. Manolopoulos, L. Iliadis, & I. Karydis (eds.) Research and Advanced Technology for Digital Libraries: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, 18-21 September 2017, 15-26.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2017c). Studying conceptual models for publishing library data to the Semantic Web. In J. Kaamps, G. Tsakonas, Y. Manolopoulos, L. Iliadis, & I. Karydis (eds.) Research and Advanced Technology for Digital Libraries: 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, 18-21 September 2017, 625-55.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2018). Assessing the preservation of derivative relationships in mappings from FRBR to BIBFRAME. In E. Garoufallou, F. Sartori, R. Siatri, & M. Zervas (eds.) *Metadata and Semantic Research: 12th International Conference, MSTR 2018*, Limassol, Cyprus, 23-26 October 2018, 230-41.
- Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2019). Mapping derivative relationships from RDA to BIBFRAME 2.0. *Cataloging & Classification* 57(8), 278-308.