

The Catalog Resisting the Web: An Historical Perspective

Philippe Bourdenet

SCD de l'Université du Maine, Av. Olivier Messiaen,
72085 LE MANS Cedex 9, France, <Philippe.Bourdenet@univ-lemans.fr>



Philippe Bourdenet has been ILS administrator and Systems Librarian for academic libraries since 1999 and has followed the successive evolutions of the descriptive bibliographic formats and reference tools. He has devoted himself to research in information science and engineering since 2008, with a particular interest in interoperability protocols in documentary platforms. He belongs to the DICEN group (Paris, Cnam), the CfU (French UNIMARC Committee), and the GE6-AFNOR group which is considering the implementation of RDA in France. These experiences come within the framework of a Ph.D. about the evolution of library services and their organization around electronic resources.

Bourdenet, Philippe. *The Catalog Resisting the Web: An Historical Perspective*. *Knowledge Organization*. 39(4), 276-282. 18 references.

ABSTRACT: Libraries are currently seeking to restructure their services and develop new cataloguing standards to position themselves on the web, which has become the main source of information and documents. The current upheaval within the profession is accompanied by the belief that libraries have a major role to play in identifying and supplying content due to their extensive high quality databases, which remain untapped despite efforts to increase catalog performance. They continue to rely on a strategy that has been proven successful since the mid-nineteenth century while seeking other models for their data. Today, they aim to exploit changes brought about by the web to improve content identification. The current intense debate on RDA implementation mirrors this desire for change. The debate is rooted in past efforts and yet tries to incite radical changes as it provides for interoperability from the creation of records through an object modeling in line with web standards and innovations. These innovations are presented through an historical perspective inspired by writings by librarians who are entrusted with helping in the development of bibliographic description standards.

Received 7 March 2012; Accepted 7 March 2012

1.0 Introduction

Standardization activity in libraries has probably never been so intense nor raised as much concern (Danskin 2006) as it currently does with the debate surrounding the implementation of RDA in France. In order to better understand the challenges that standardization presents, one must consider historical context and the underlying problems not necessarily linked to RDA itself but to the model upon which it is based, FRBR, whose emergence demonstrated an extensive review on how to improve information cataloguing and identification tools.

For libraries, the problem is twofold, because RDA clearly intends to meet the users' needs as interpreted by RDA developers, who may have a distorted under-

standing of some necessities or priorities. It supposes that users' needs will be met by an overhaul of cataloguing rules. This very hypothesis may be biased. It is indeed an approach that has been proven effective in the past. Over the past 150 years, libraries' cataloguing solutions were based on this hypothesis and were successful until the end of the 20th century. But it is unclear whether applying this principle today would bring similar success. Until the 1990s, little progress had been made in the container/contents relationship, but problems of management and identification have dramatically increased in direct proportion with the explosion of digital material in its native form as well as user accessibility and tools. FRBR is driven by the necessity to create content metadata whereas cataloguing rules, in their most evolved form,

are still producing data based on containers (Moore 2006).

2.0 Innovations in libraries

2.1 Historical Heritage

FRBR evolved out of a series of developments in the history of cataloguing (Taylor 2007). The famous 91 rules of cataloguing established by Sir Anthony Panizzi (1841), which aimed at improving the quality of the descriptions of the British Library printed books, were already highly innovative. Another development worth mentioning is Charles Cutter's cataloguing code, known for the Cutter number which integrates an abbreviation of the author's name in the classification code, as well as the third edition of its manual *Rules for a Dictionary Catalogue* (1904), which elaborates on principles first presented in 1876 and gradually enhanced. His definition of a catalog addresses three objectives: to "enable a person to find a book of which either the author (A), the title (B), or the subject (C) is known, to show what a library has by a given author (D), on a given subject (E) or in a given kind of literature (F), and to assist in the choice of a book as to its edition (G) (bibliographically) or as to its character (H) (literary or topical)." These briefly summarized basic rules require the creation of an "author-entry (for A and D), a title-entry (for B), a subject-entry, cross-references and classed subject-table (for C and E), a form-entry and a language-entry (for F), giving edition and imprint, with notes when necessary (for G) and notes (for H)," in short, nearly all the information contained in a record, even today. The system does not propose a solution for identification of multiple authors of collective publications, pseudonyms, or rules for author classification, collective responsibility or indexing resources with non-Latin alphabets. Despite these shortcomings, which must be addressed especially since the volume of collections is growing, FRBR remains the distant successor to the rules proposed by Cutter. The connection is obvious between the statements "find a book of which the author is known" and "find a particular manifestation when the name(s) of the person(s) and/or corporate body(ies) responsible for the work(s) embodied in the manifestation is(are) known" (Taylor 2007) (FRBR 1998).

As early as 1900, the United States began to harmonize cataloguing rules from the ALA, the Library of Congress, and the Dewey Decimal system with cooperation back and forth across the Atlantic, which

took into account new British cataloguing rules and "Prussian Instructions" used in Germany and some Scandinavian countries. The international cooperation, in its early stages, was supported by the work of the mathematician and librarian Ranganathan who, in 1931, wrote the Five Laws of Library Science:

Books are for use
Every reader his (or her) book
Every book its reader
Save the time of the reader
The library is a growing organism.

The word "growing" in this context does not only refer to libraries' collections in terms of linear meters of occupied shelves, but also to expansions in technology, knowledge, and international flow of knowledge, ease of global transport and migration of populations especially those in charge of their culture, conflicts, etc. of which the library is a reflection. In addition, these evolutions introduce the concept of use, a pre-occupation addressed by FRBR (Taylor 2007)—inherited from AACR—and the overhaul of cataloguing and indexing rules carried out by IFLA for more than 40 years now and initiated by the Paris principles (1961), which put user needs at the centre of the debate. Changes in the uses and mentalities which allow us to perceive the library as an organization have taken a major step forward with digital technology because information is more easily manipulated by the user, allowing an intangible appropriation of the information. This appropriation is only possible if some of the mechanisms of data processing are made transparent. This need "puts the library community in the position to move forward quickly into the broader world of information exchange and reuse outside the library silo created over the past 40 years" (Hillman et al. 2010).

2.2 The catalog as a factor of opacity

Before considering data recovery and manipulation, which are relatively new needs, we must focus on the catalog's main functions. Its "first objective is to enable the user of the catalog to determine readily whether or not the library has the book he wants" (Taylor 2007). This going back to basics may seem extreme according to Lubetzky (Taylor 2007), who, in 1953, considered certainties which, over the years, may have been overlooked by bibliographic search innovators. An example of one such certainty is that libraries consider their catalog a tool that renders

transparent their primary activity of building up collections while the user (a layperson) is unable to maximize use of the catalogue without specialized knowledge and/or training. The second main objective, subtler but linked with the first one is “to reveal to the user of the catalog, under one form of the author’s name, what works the library has by a given author and what editions or translations of a given work” (Taylor 2007). Even if it appears as a revisited rule, the concept of “work” is finally introduced here, for the sake of *identification*, and shows indirectly one inadequacy of Cutter’s vision whose searching principles were focused only on books.

Since 1961, initiatives to develop description models and cataloguing tools multiplied; in 1967, (Manning 1998) AACR’s first conference already suggested new codes, in 1990 (FRBR 1998) in Stockholm, economic realities were first discussed including the pressure on institutions to reduce cataloguing costs while improving user services and providing tools of description for different media.

2.3 The FRBR model

FRBR (Functional Requirements for Bibliographic Records) was first issued in 1998, after ten years of work. It is a conceptual model, which, if implemented, provides the user with a catalog allowing him *to find, to identify, to select and to obtain* a resource [IFL 01]. The FRBR can be briefly presented here bearing in mind that its specifications are based on a model of 10 entity-relationship, sorted into three groups which cover respectively (1) the level of bibliographic description (work, expression, manifestation, item), (2) the level of responsibility (person, corporate body) and (3) the level of subjects of works (concept, object, event, place).

In 1997, faced with inevitable dissatisfaction of both professionals and users due to inadequate cataloguing rules for the new media while the AACR were being reviewed, the Joint Steering Committee (JSC) held a conference focused on future uses of this cataloguing code and invited international participants “in the hope of freeing cataloguing rules establishment from the Anglo-American context only and developing a code which could be used worldwide” (Taylor 2007).

RDA is a cataloguing code which is based on the traditional principles of Panizzi, Cutter, Lubetzky¹ and which also validates the principles of the FRBR conceptual model and its extensions in the field of authority control (FRAD, FRSAR). The seven chapters on description summarize FRBR intentions in addi-

tion to the first one which is a general rule: (2) Identification of the resource (see “identify”); (3) Type of media (see “select”); (4) Content (see “select”); (5) Access (see “obtain”); (6) Persons, families, and corporate bodies; and (7) Linked resources (see “find”).

RDA therefore proposes a set of instructions allowing the creation of metadata to describe a content on a standardized form for the Web, able to take account of either an electronic resource or a paper document. Some of the leading bibliographic agencies are currently examining implementation options for their bibliographic models to decide whether or not to adopt RDA. Though this can have immediate benefits for bibliographies (Pisanski, Zumer, and Aalberg 2009), the objective is to improve catalogs, with their bibliographic records (B) and their authority records (A).

3.0 Implementing FRBR in RDA

3.1 Implementation scenarios worldwide and in France

Economic stakes are very high. The AACR had already intended to provide solutions to reduce cataloguing costs, by facilitating data exchange, harmonizing practices in the Anglo-Saxon world and supporting derivative cataloguing. The FRBR model and its implementation in RDA require so many changes both in the presentation of data and in their intellectual orientation through introduction of the object model in order to progress towards the Semantic Web vision (Dunsire 2009) that original MARC records are unusable. The Library of Congress currently estimates the number of MARC records at just over one billion (McCallum 2004) worldwide, as a result of collaborative work over the past four decades. Tom Delsey, RDA Editor and member of JSC, defines three implementation scenarios for the databases currently managing MARC formats (Delsey 2009):

-First scenario: “RDA data are stored in a relational or object-oriented database structure that mirrors the FRBR and FRAD conceptual models” (Delsey 2009). These structures are not those of current MARC formats.

-Second scenario: “RDA data is stored in database structures conventionally used in library applications [MARC]. In those structures, data is stored in bibliographic records and in authority records, and in some implementations in holdings records as well” (Delsey 2009). Data would be modified in order to establish bibliographic descriptions or authorized access points

representing FRBR entities, and bibliographic records will be linked to authority records for persons or corporate bodies, expression, manifestation, or work.

-Third scenario: RDA data is stored in bibliographic and authority records, based on the entities model, with no link between them.

In France, cataloguing standards (CG 46 "Information and Documentation," 28 March 2000) are prepared and updated by the AFNOR Group CN 357, which appointed the GE6 (Expert panel number 6) to do a study on RDA implementation, the feasibility of the above scenarios, the associated development costs for document applications, and to provide a solution to ensure data recovery by the system.²

There is no easy way to "FRBRise" a MARC record. The first scenario, ideal because it complies with FRBR requirements by creating records which can be exploited for web application, would be extremely expensive to implement. It would be contradictory for the AACR to adopt this solution since their objective is to streamline expenditures. It wouldn't be easy to take this decision—even if it provides "an interesting theory about a four-level model—especially that the FRBR have not been tested in actual practice" (Coyle 2004). Librarians therefore don't have the necessary hindsight or enough practical experience to refute or support it (Pisanski, Zumer, and Aalberg 2009) (Maxwell 2008), despite the new initiatives for its application with varying degrees of rigor in the United States, Australia, Sweden and even in France.

The second scenario would be more reasonable cost-wise, but it would require creation of FRBR entities from existing MARC records. This tedious work on evolutions based on strict mappings would obviously take time, but would thereby allow a gradual test of the improvements to the information systems, provided that user response would be able to be gauged along the way through accessible functionalities (Wells 2007). This expertise largely lies with bibliographic agencies and providers of MARC records in France, mainly the BnF³ and the ABES,⁴ and the decision is up to the strategic committee of CG46.

3.2 Tradition and innovation

The profession will most likely face some major upheaval, but experience has shown that, with regard to libraries, change cannot not be rapid, because even very dynamic evolutions must be rooted in a principle of sustainability due to the stability of the profession,

its uses, as well as its trade practices (Calhoun 2007), but proposed JSC deadlines are too short. Some accomplishments show that libraries are willing to change their practices and adopt other non-traditional formats, as was the case at the University of Arizona, in cooperation with the NAL,⁵ with the creation of a digital library in RDF (Han 2006). The initiative of creating web standards for library use with the W3C *Library Linked Data Incubator Group* also provides evidence of what librarians are ready to do to position themselves on the web, which has become the main source of information. The adoption of object-oriented models able to make bibliographic information at the level of the Semantic Web is inevitable and the theoretical model developed by FRBR is, from this point of view, of enormous value, but will RDA allow this? Probably disappointed by JSC decision to convert AACR2 from RDA—a courageous but risky choice—librarians have shown strong resistance to the adoption of cataloguing code in the United States. The decision raised reactions as well as criticism from librarians over RDA, because "it neither sticks with the standards we've already got, nor offers anything [the] present OPACs can make use of in any kind of a helpful way"⁶. This code is still seen as a "prediction," which "could theoretically work in the future," but which has a long way to go to prove its value. Librarians' distrust of RDA is joined by a more scientific criticism that questions whether libraries' evolution should include RDA implementation. It is perhaps a mistake to start with cataloguing rules development to change the services (Coyle and Hillmann 2007): "Prior to elaborating detailed cataloging rules for libraries, we need to decide whether the user will view a general bibliographic tool that connects users and information resources no matter their origin, or continue to view a library inventory, that requires users to look elsewhere for other information they might need." This is an uncomfortable issue for the profession, as it highlights the risk in introducing profound changes that may cause years of inaction before providing theoretically better service. There's currently no proof of RDA's effectiveness, which is all the more risky given the urgent need to rethink the services, to give them meaning and to make them compatible with users' practices. This move could turn out to be a waste of time and an inappropriate investment—to make a fundamental change in services—when libraries urgently need to reach new users, build stable foundations when, at least in the near future, they must continue to offer user tools universally regarded as obsolete, even within the profession, with the un-

certain prospect of creating more appropriate services. There is the risk that readers may permanently turn their backs on inaccessible document platforms, including the traditional OPAC.

3.3 Fears and suspicions

Reservations expressed in North America may be explained by the fact that, unlike UNIMARC, MARC21 does not manage the links between the records with the identifier indexing, which dramatically increases the adjustments to move towards RDA, but this is only one factor. An overview of these reservations reveals suspicions about the real objectives, sometimes by the strongest supporters of the FRBR model (Coyle 2004; Coyle and Hillmann 2007). Yet, there has been extensive communication efforts on RDA objectives at the very core of RDA toolkit specifications. "The first objective of RDA is to be sensitive to the user needs" (Oliver 2010). Each chapter of the RDA toolkit indicates the functional objectives and principles (see for example "Record Attributes of... / Section 1 : Manifestation & Item / ... / 1.2 Functional Objectives and Principles").

4.0 Challenges and perspectives

4.1 A great desire for change

The objectives as presented in the RDA toolkit are highly persuasive since they are based on practical considerations. If we are to still speak of "records" as aggregate data forming an intellectual unit to describe a resource, the resource is an entry to a set of services which are more or less extended according to the user authentication and then more or less personalized. More generally, a search in a FRBRised catalog should allow the disambiguation of a result, mainly in view of improving navigation and display of information. The proposed granularity is finer than it is for MARC or ISBD records. Catalogers are no longer allowed to note information that serve to identify a resource through a long character string which concatenates disparate properties (for example, the UNIMARC General Note block 3XX). Instead, all information is qualified. A record aggregates metadata on information that is an integral part of the resource by describing its properties through a pattern, the projection speed of moving images, the device or software required to read a document on a medium, or any other associated material, rather than the "Other material characteristics," which is a very vague and ambiguous

indication found in the full display records generated by traditional OPACs. In addition, RDA offers removal of abbreviations, indexing of the whole statements of responsibility which so far were limited to three, etc.

4.2 User benefits

The Results presentation allows a classification of resources according to their nature (with metadata focusing on content). A search for "le Barbier de Séville" (*The Barber of Seville*) using a catalog can return as diverse results as theater plays, prints, critical texts, music scores, video clips, DVD-ROM references, etc. (many thousands with the BnF's general catalogue)—with specific icons for each medium. Instead, RDA would propose a quick answer with potential analysis of series of results listed by content type and this would be the first visible effect of a modeling work. This achievement would also allow us to widen search for particular adaptations, parodies, or other works which are intellectually linked to the original search and identify the resource main language from the short list of results. The manifestation of these aggregates requires complex data management to create links between the data. Of course, this would be possible with traditional tools, but engineering efforts would be of such a magnitude that barriers would seem insurmountable. The RDA model creates functionalities by using information existing in the data and no longer trapped in relational databases. The traditional catalog is seen as a silo whereas RDA introduces a paradigm from the creation of records, which do not need to be stored in complex and invisible reservoirs. And this, on its own, facilitates the readability of web holdings and communication with other communities producing data, because it is not designed for a particular format. The obvious advantage is that users can combine searches from more than one source, (by) bringing together users of ONIX⁷ (Kiorgaard 2006), which is a standard designed for the industry, and academics who use bibliographic tools complying with the standards applied in libraries. Web standards allow this convergence and RDA, based on these technical prerequisites, offers the hope of common semantics and data reuse in other contexts.

4.3 Institutions benefits

From an institutional perspective, to seize this opportunity is to capitalize on technological innovations

from other countries for common implementation the boundaries of which are defined by the web technologies. A recovered record inherits all the links it contains and introduces the concept of “continuity.” Acquiring a record is recovering more than one link in the chain. Since it can be reused by another institution or a user, this link will allow the extension of another chain and connect two worlds (RDA Toolkit, Key Features, 0.1). Though FRBR is intellectually a successful model, it is sterile, with no mechanism or tool to create data. RDA is this tool and, in addition, proposes a set of benchmarks to help make the right decisions regarding hierarchical arrangement of information and strategic orientation (in terms of services strategy) to reach a category of users. RDA supporters as well as the undecided are on the lookout for arguments the likes of which they usually obtain with technical case studies, but information and documentation professionals have to agree on the need to pursue this objective: “bibliographic data were created to be read and understood by librarians and users” (Coyle 2007) at a time when OPACs have been criticized for creating misunderstanding between librarians and users.

5.0 Conclusion

These considerations make a critical viewpoint of RDA more difficult, since libraries seem determined to evolve towards the Semantic Web. There aren't enough alternative initiatives for this evolution to categorically reject the RDA model and production tool. This work requires consideration of RDF schema: “labels, areas and attributes need to be expressed as classes and properties” (Dunsire 2009), which is the basis of the object orientation, and vocabularies will have to evolve into SKOS and semantic relations into OWL. RDA proposes to develop straightaway interoperability of data from its creation, whereas the work done in the development of document platforms over the last decade implemented interoperability protocols either through connectors to build bridges toward external reservoirs or through local integration of data created elsewhere, which now seems to be inadequate to exploit the data from the web. This task was possible thanks to the high quality control on data, and this kind of exchange requires stability and *continuity*. Integrating the need for interoperability from the very outset stems both from tradition and from a very challenging dynamism—tradition because strictly respecting data formats is not new and ensures readabil-

ity by other institutions which will now be able to understand and use them, and dynamism, because the method uses processes of relations with entities, creating dynamic interoperability with semantics which allow them to be used on the web and not only in libraries. Overall, this is not fundamentally different from the changes proposed by Panizzi, Cutter, Ranganathan, or Lubetzky, who saw a need to work on records' substance in order to improve services and take into account global technology advances to meet user needs.

Notes

1. RDA Toolkit, Introduction, 0.2 : Relationship to Other Standards for Resource Description and Access
2. Working Group on RDA adoption in France : <http://rda-en-france.enssib.fr/>
3. Bibliothèque Nationale de France (= French national Library)
4. Agence Bibliographique de l'Enseignement Supérieur (= French Higher Education Bibliographic Agency)
5. National Agriculture Library
6. <http://pointsmean.blogspot.com/2011/01/rda-and-opacs.html>
7. <http://www.editeur.org/8/ONIX/>

References

- Calhoun, Karen. 2007. Being a librarian: metadata and metadata specialists in the twenty-first century. *Library hi tech* 25: 174-87.
- Coyle, Karen. 2004. Future considerations: the functional library systems record. *Library hi tech* 22: 166-74.
- Coyle, Karen, and Hillmann, Diane. 2007. Resource Description and Access (RDA): cataloguing rules for the 20th century. *D-lib magazine* 13n1/2, January/February.
- Danskin, Alan. 2006. On ne sait jamais: demain, la fin du catalogage? *World Library and Information Congress: 72nd IFLA General Conference and Council. Libraries: Dynamic Engines for the Knowledge and Information Society*, 20-24 August 2006, Seoul, Korea. Available <http://archive.ifla.org/IV/ifla72/Programme2006.htm>.
- Delsey, Tom. 2009. RDA database implementation scenarios. Available <http://www.rda-jsc.org/docs/5editor2rev.pdf>.

- Dunsire, Gordon. 2009. UNIMARC, RDA et le web sémantique. *World library and information congress: 75th IFLA general conference and assembly. Libraries create futures: building on cultural heritage*, 23-27 August 2009, Milan, Italy. Available <http://conference.ifla.org/past/ifla75/programme2009-en.php>.
- IFLA Study Group on the Functional Requirements for Bibliographic Records, and International Federation of Library Associations and Institutions. Section on Cataloguing. Standing Committee. 1998. *Functional Requirements for Bibliographic Records: final report*. München: K.G. Saur.
- Han, Yan. 2006. A RDF-based digital library system. *Library hi tech* 24: 234-40.
- Hillmann, Diane, Coyle, Karen, Phipps, Jon, and Dunsire, Gordon. 2010. RDA vocabularies: process, outcome, use. *D-lib magazine* 16n1/2, January/February.
- Kiorgaard, Deirdre. 2006. RDA/ONIX framework for resource categorization. Available <http://www.loc.gov/marc/marbi/2007/5chair10.pdf>.
- Manning, Ralph. 1998. The Anglo-American Cataloguing Rules and their future. *64th IFLA general conference*, 16-21 August 1998, Amsterdam. Available <http://archive.ifla.org/IV/ifla64/64cp.htm>.
- McCallum, Sally. 2004. An introduction to the Metadata Object Description Schema (MODS). *Library hi tech* 22: 82-88.
- Moore, Julie Renee. 2006. RDA: new cataloging rules, coming soon to a library near you! *Library hi tech news* 23n9: 12-17.
- Oliver, Christine. 2010. *Introducing RDA: a guide to the basics*. London: Facet.
- Pisanski, Jan, Zumer, Maja, and Aalberg, Trond. 2009. Frbérisation: vers un nouvel avenir brillant pour les bibliographies nationales. *World library and information congress: 75th IFLA general conference and assembly. Libraries create futures: building on cultural heritage*, 23-27 August 2009, Milan, Italy. Available <http://conference.ifla.org/past/ifla75/programme2009-en.php>.
- Taylor, Arlene G. 2007. *Understanding FRBR: what it is and how it will affect our retrieval tools*. Westport, Conn.: Libraries Unlimited.
- Wells, David. 2007. What is a library OPAC? *The electronic library* 25: 386-94.
- Maxwell, Robert L. 2008. *FRBR: a guide for the perplexed*. Chicago: ALA.