What Is RDA, and Why Should E-Book Managers Care?

By Steve Kelley

In recent years, e-book managers and other librarians have posed many questions about RDA, the new cataloging code. Over the course of this chapter, we hope to answer the two-part question: what is RDA, and why should e-book managers care? RDA provides a necessary foundation for building better data structures that will change the capabilities of future cataloging systems. It is a code that was created to be compatible with digital catalogs and digital resources, as well as traditional, non-digital materials. RDA may also provide a necessary step toward integrating library resources with the much-heralded Semantic Web. In short, the future of library catalogs lies with RDA. And that is why e-book managers should care about it.

What Is RDA?

RDA is an abbreviation for *Resource Description and Access*, a new cataloging code designed to replace the *Anglo-American Cataloging Rules, 2nd ed.*, commonly known as AACR2. This new code should be of interest to e-book managers because it was developed in large measure to address AACR2's shortcomings in adequately describing internet-based resources. That RDA is intended to replace AACR2 is clearly illustrated by the fact that the JSC, the organization responsible for overseeing the cataloging rules used in the English-speaking world, has changed its full name from the Joint Steering Committee for the Revision of AACR to the Joint Steering Committee for the Development of RDA. The JSC has stopped any further revisions to AACR2, and will be devoting all of its efforts to developing RDA. AACR2 is, effectively, a dead code, while RDA is a growing, adaptive set of rules that can be changed over time.

RDA is a content code, which describes how information should be recorded in bibliographic records. RDA provides rules for how to record data such as title information, author names, organizational names, publisher names, and the physical description of bibliographic entities. Although RDA is designed to replace AACR2, the newer code is based on the older one. RDA is intended to be compatible with AACR2, which means that databases will be able to contain both RDA and AACR2 records, without a loss of functionality. Also, RDA is intended to be used in any record format, such as MARC (Machine Readable Cataloging), Dublin Core, EAD (Encoded Archival Description), MODS (Metadata Object Description Schema), etc. RDA does not specify any changes to these formats (although they may need to be revised to accommodate the data encoding required by RDA).

Just to be clear, RDA will affect most of the content of bibliographic records, but has nothing to do with subject headings and call numbers. We will continue to use LC (Library of Congress), MESH (Medical Subject Headings) and other subject heading systems like we always have. Similarly, LC (Library of Congress), Dewey and SUDOC (Superintendent of Documents) call numbers will not change due to the new code.

Why Does It Seem to Be Taking So Long to Move to RDA?

The transition away from AACR2 toward full-implementation of RDA has been, and will continue to be, a long process because there are a number of moving parts and a number of players involved. Also, quite simply, there is a lot of inertia that is slowing the process. AACR2 was adopted in 1981, in a time that not only pre-dated the Web, but a time when most libraries still relied on card catalogs. The explosion of information technology has been a boon to the functionality of modern library catalogs, but it has also created new layers of complexity. Not only must RDA be adopted broadly by libraries throughout the world, the record formats (such

as MARC) must be revised to properly capture the detailed data specified by RDA, and cataloging systems must be adapted to properly display the revised record formats. Libraries are currently moving toward adopting RDA, and a few changes have been made to the MARC formats, but we are still some years away from a really full implementation of RDA, in the sense of catalogs being able to properly display all of the functionality the new code promises to make possible.

The development path of RDA has already been a long one. The roots of the new code date back to the International Conference on the Principles and Future Development of AACR, held in Toronto, Canada, in 1997. Conference participants discussed AACR2's weaknesses in properly describing internet-based resources, among other flaws. Prompted by this conference, the JSC began discussion for developing a third edition of AACR, but this was eventually abandoned for a more thoroughgoing code revision, which became RDA (Ehlert) . Early drafts of RDA were closely scrutinized by the cataloging community and received a fair amount of criticism, which prompted the JSC to re-work the code several times. In 2010, the U.S. National Libraries (the Library of Congress, the National Library of Medicine, and the National Agricultural Library) began a test of RDA, in conjunction with twenty-three other libraries throughout the U.S. (including academic, public, and special libraries), during which they created catalog records using RDA. The testing phase closed on December 31, 2010, and, after a period of analysis, the three U.S. National Libraries announced at the ALA Annual Conference in New Orleans that they were delaying implementation of RDA until January 2013 at the earliest. However, this decision should not be interpreted to mean that the National Libraries will not eventually adopt RDA. As was mentioned earlier, the JSC has stopped further work on AACR2 and will devote all of its future efforts to the development of RDA, making it the only

option for a growing, changing cataloging code. The National Libraries determined that the current draft of RDA did not meet several of the stated goals of the code's developers and have requested that the JSC revise RDA to meet these goals, , but there is virtually no chance that they will not ultimately adopt the new code. Several libraries, including the University of Chicago, have already adopted RDA in their cataloging departments, but most libraries in the United States will likely wait to make their own decision until after the U.S. National Libraries adopt RDA.

As libraries gradually adopt RDA, they will have to encode their RDA-compliant records in bibliographic formats, which will have to be revised to accommodate the new data requirements of RDA. For most libraries, this will mean the MARC formats, although Dublin Core, MODS, EAD and other record formats will also be affected. The MARC formats are managed by the Network Development & MARC Standards Office of the Library of Congress. The MARC Standards Office has already approved several changes to the MARC formats to accommodate RDA and more are under consideration. However, the Library of Congress has begun work on an even more fundamental change to the bibliographic record formats by beginning the process of developing a format to replace MARC. This promises to be an even larger undertaking than adopting RDA, and will certainly take years to implement.

As libraries adopt RDA and record formats change to accommodate RDA, online catalogs will also have to evolve to properly display the new data encoded in the revised formats. Several catalog systems are trying to move toward RDA-compliance, but both commercially available and open-source systems have much work to do to fully implement RDA-style records, while maintaining full functionality for older records created using AACR2 (even after RDA is widely adopted, catalogs will have a mix of RDA and AACR2 records for many years to come,

due to the fact that there are millions of AACR2 records and the retrospective conversion of these records to RDA will be an expensive, time-consuming process, if undertaken). So, the transition to RDA is best viewed as an ongoing process that will continue for several more years, as libraries, record formats and catalog systems work toward fully implementing this new code. But why should we bother? What advantages does RDA confer on the world of online catalogs? This brings us to our next set of questions.

Why Do We Need RDA? What Benefits Does RDA Offer Over AACR2?

So, why *do* we need a new cataloging code? The most fundamental problem is that AACR2 is simply out of date. Although it has been revised over the years, it was initially adopted in 1981. The code was originally designed for a world of card catalogs, when computers were not in broad general use, and that is to say nothing of the very small Internet of the time, or the not-yet-invented World Wide Web. A variety of changes and additions have been made to AACR2 over the past thirty years to bring it up to date with current information technologies, but the basic code was written for a pre-digital age. RDA addresses the problem of describing digital works for a digital environment by adopting the principles of FRBR, which brings us to our first sub-question:

What Is FRBR?

FRBR stands for *Functional Requirements for Bibliographic Records*, a conceptual model for bibliographic entities developed by the International Federation of Library Association and Institutions (or IFLA). First published in 1997, FRBR is based on entity-relationship modeling. It is not a cataloging code, but rather a method for thinking about cataloging and the relationships between various bibliographic entities.

FRBR identifies the four basic tasks performed by catalog users. They use data to **find**, **identify**, **select**, and **obtain** bibliographic materials (books, e-books, journals, DVDs, etc.) (FRBR 2.2). That is, all catalog users are trying to accomplish one or more of these tasks, and need the proper data to accomplish the task or tasks. Every search of a catalog is an attempt to find a work that is needed, identify a useful work, select a work from among various options, and/or obtain the work by finding its unique location (including, in some cases, directly accessing an online work through a URL).

With these four tasks recognized as essential to all uses of the catalog, FRBR then identifies four types of related bibliographic entities that are of interest to catalog users. These entities have varying levels of abstraction and are useful to different sets of user needs. The entities are, from most abstract to most particular, **works**, **expressions**, **manifestations**, and **items** (FRBR 3.1). We will attempt to explain the differences between the four types of entities by discussing the various ways we use the word "book."

In the sentence, "Leo Tolstoy is the author of the book 'War and Peace,'" we are using the word "book" to refer to a **work**. That is, we are talking about 'War and Peace' as an intellectual creation in its most abstract sense. We have even abstracted it from its original language, because Tolstoy wrote it in Russian, with the title "Voina i Mir." In the example sentence we are referring to a novel by Tolstoy about a set of characters in 19th century Russia, regardless of the language or format in which it is captured. No matter how "War and Peace" is rendered, Tolstoy is the author of the **work**.

In the sentence, "Millions of copies of the German version of the last Harry Potter book were sold in all formats," we are using the word "book" to refer to an **expression**. Now obviously we are talking about a **work** when we refer to the last Harry Potter book, but we are

looking at just the German translation of this work, which is one of several possible **expressions** of the same **work**. An **expression** is the realization of a **work** in text, sound, image or other representational forms (FRBR 3.2.2). When talking about a book as an **expression**, we are referring, in an abstract sense, to a textual arrangement of certain words in a certain order. The German text of the last Harry Potter book consists of some thousands of German words arranged in a particular order. Whether these words are arranged inside a hardback book, a paperback, or on an e-reader, they constitute a single **expression** of the **work**.

In the sentence, "The latest John Grisham book is the number one hardback bestseller in the U.S.," we are using the word "book" to refer to a **manifestation**. The Grisham book is, of course, a **work**, and the English-language version is a particular **expression** of the book, but we are also talking about the hardback edition, which is a particular **manifestation** of the book. A **manifestation** is the physical embodiment of an **expression** of a **work** (FRBR 3.2.3). In terms of books, a **manifestation** generally corresponds with an edition. The **manifestation** is the first of the four entities described by FRBR to be tied to a physical presence. Both **work** and **expression** are inherently abstract concepts. The different **manifestations** of a book may be the hardback edition, the paperback edition, the Kindle version, the Nook version, etc.

Finally, in the sentence, "Please place the James Herriot book on the shelf," we are using the word "book" to refer to an **item**, a specific physical copy of a particular book. An **item** is a single exemplar of a **manifestation** (FRBR 3.2.4), and it is also the most concrete and conceptually easy to grasp of the four FRBR entities. An **item** is a physical object that you can put your hands or eyes on. The **item** is a single copy of a **manifestation**, which embodies an **expression** of a **work**.

Figure 9-1 illustrates how the four FRBR bibliographic entities relate to a specific book. At the **work** level, we have "Voina i Mir," which is "War and Peace" as conceived by Leo Tolstoy. It could be in Russian or translated into any other language, but it will always be the intellectual creation of Tolstoy. Next, at the **expression** level, we have an English translation of "War and Peace," which captures the ideas of Tolstoy's work into a specific sequence of English words. These words can be printed in a paperback book, or published in an e-book. The physical carrier does not matter, what matters is the intellectual content of the expression. Next, at the **manifestation** level, we have the Oxford University Press edition published in 1983. This refers to a group of thousands of books that all look the same, and all have the same expression of the work on their pages. And finally, at the **item** level, we have the copy of this edition held at Z. Smith Reynolds Library.

INSERT FIGURE 9-1 APPROXIMATELY HERE.

Now, you may be wondering, what does all this mean? Well, RDA is based on the ideas of FRBR and the four types of bibliographic entities. This means that, by using RDA, catalogers can create records that have varying levels of specificity to meet different user needs. Although AACR2 does not use FRBR terminology, the rules of AACR2 require catalogers to work at a **manifestation** level (in FRBR terms), referring to the particular physical forms of bibliographic works. But by using RDA, catalogers can create more abstract **work** and **expression** level records. This will be particularly useful for e-books. With e-books, the user is not tied to any one delivery device or physical format. Someone could search a catalog looking for "*The Adventures of Huckleberry Finn*," and with the more abstract **expression** level record, find all of versions of the English text of the book. The user may not care whether the text is available as a loan to her Nook, or by borrowing a Kindle from a library, or by downloading a file to her

laptop. Or, she may be willing to go to the library to retrieve a print copy of the book. The point is, by allowing for more abstract searching for the intellectual content of a work, RDA frees us from AACR2's limitations requiring us to describe specific physical objects.

At the same time, RDA will make faceted searching by material format more robust, because it allows for more detailed description of formats than did the old system of general material designations (GMDs). For example, the GMD "videorecording" was applied to both DVDs and videocassettes, putting the formats together in one category, while RDA allows for clearly encoding whether an item is a DVD or a videocassette. The combination of abstracting the intellectual content of a work from its physical carrier, while simultaneously providing a more detailed description of the carrier, will allow users to search the catalog to more easily find, identify, select and obtain the intellectual content they seek, across formats and physical carriers.

What Are the Other Benefits of RDA?

Because RDA was built from the ground up, rather than simply revising AACR2, the JSC took the opportunity to try to remove the English-language bias inherent in AACR2, which, after all, stands for "*Anglo-American Cataloging Rules, 2nd edition*." The bias toward English is apparent in the very name. The JSC recognized that translated versions of AACR2 are used in various countries, with varying degrees of difficulty arising from the English-centered nature of the rules, therefore, they drafted RDA with an eye toward making the code linguistically neutral.

Another structural problem of AACR2 that RDA addresses is the fact that AACR2 was initially written over thirty years ago to be used with card catalogs. The RDA rules were designed to be used with electronic online catalogs, and provide for detailed, rich description of electronic resources. AACR2 provides rules for describing electronic resources, but these rules were adapted from rules for print books, and can sometimes be difficult to use (the phrase

"trying to fit a square peg in a round hole" springs to mind). By beginning with a clean slate, RDA provides for better description of "born digital" materials, as well as allowing for better and easier description of the electronic versions of print books.

Because RDA is a product of the digital age, it also explicitly addresses the role that catalog records can play in the direct access of library materials. With non-digital resources, the catalog provides bibliographic information that allows the user to locate materials, but with digital resources, the URL in the catalog record is often the means of accessing the materials. The point at which a work is found in the catalog can also be the point where the work is accessed. RDA acknowledges this important development (which is apparent in the name of the code, "*Resource Description and Access*").

Another important benefit of RDA is that it allows for richer description of the relationships between multiple creators, between multiple works, and between creators and works. RDA does a better job of AACR2 of describing relationships between multiple creators by providing for more accurate, detailed description of the roles of various creators. This description can be so precise as to use terms like "recording engineer" or "lithographer." This level of detail in describing the precise role played by a creator can be of use to the patron.

RDA also provides for better description of the relationships between various works, such as derivative works, which are works based on other works. For example, the novel "Gone With the Wind" has a number of derivative works. There is a sequel novel called "Scarlett," a parody novel called "The Wind Done Gone," and a film version, also called "Gone With the Wind." A book about the film called "The Filming of *Gone With the Wind*" is a derivative work of the film, so we would want to record that relationship. However, we would not want to have

the film study directly related to the novel (see Figure 9-2). RDA allows for the clear recording of these types of relationships between works, helping patrons find the works they need.

INSERT FIGURE 9-2 APPROXIMATELY HERE.

Regarding relationships between creators and works, AACR2 limited catalog records to contain access points for the names of no more than three creators of a given work (a legacy of the card catalog), but RDA does away with this "rule of three." Using RDA rules, a cataloger can decide to provide an access point for every person involved with creating a work. This means, for example, that if a user tries to search for all the works of a given author, under RDA rules, the retrieval set will include a work where the author is the fourth contributor listed. This would not be possible under the AACR2 rule of three (Oliver, 2010).

In addition to allowing for more detailed relationships between entities in bibliographic records, RDA also simplifies the rules for recording data in records. AACR2 required the use of many abbreviations for common terms such as "pages," "volumes," "department," etc. By getting rid of these required abbreviations (another legacy of the card catalog), RDA moves toward recording data as it appears in bibliographic materials (which is also apparent in the demise of the rule of three). RDA further simplifies records by dispensing with the Latinisms that peppered records under AACR2, such as et al., sic, i.e., and others. These changes not only make it easier for catalogers to catalog materials quickly, because they do not have to remember particular abbreviations or Latin terms and when to use them, they also make the records easier for users to read and understand.

The "take it as you find it" philosophy of RDA is expected to make the cataloging rules simpler in the long run. This should result in quicker and cheaper cataloging, while also providing for richer description (in terms of relationships and electronic resources), which should

result in better-quality cataloging. Of course, all of these benefits are long-term. In the short-term, there will be some growing pains as the cataloging world transitions to RDA.

Conclusion: What Does RDA Mean For Libraries and E-Book Managers?

To be honest, RDA does not mean much for e-book managers in the short-term. Yes, a few records may pop up that have new MARC fields, and, at libraries that have in-house cataloging, the catalog librarians may be quite occupied over the next year or two learning about RDA and planning for its adoption locally. But, in truth, much of the work of implementing RDA still needs to be done. Libraries need to adopt the code, record formats need to be further changed to accommodate the code, and online catalogs need to be updated to handle the new data requirements brought about by RDA. Also, many thousands of RDA records need to be created before catalog users routinely encounter RDA records while searching.

However, RDA will help change library catalogs considerably over time. It is a foundation upon which will be built the bibliographic data structures of the future. AACR2 is simply not adequate for the digital age. RDA may not be perfect, and it definitely has its critics, but it is a code that is designed to be compatible with digital catalogs and digital resources, while AACR2 was designed for an age of card catalogs and print materials. As our bibliographic resources change, our cataloging code must adapt to these changes. RDA will allow for more precise searching for specific intellectual content, without the physical format barriers imposed by AACR2 records. RDA will also allow for richer use of faceted searches and will provide better details regarding the relationships between works. This type of information will be of enormous importance when searching for e-books especially, because they can be more difficult to browse than physical materials. We will want our catalog records to be as detailed as possible to ensure a successful search, saving the patron from downloading an e-book that is not useful.

Perhaps most excitingly, RDA may allow for bibliographic data structures to be integrated into the emerging Semantic Web. The Semantic Web will enable computers to process the meaning of information on the World Wide Web, and may facilitate such developments as natural language searching. Data on the Semantic Web will have to be finely granulated and rigidly organized to allow computers to categorize and classify information in meaningful ways to answer the queries of human users. The more finely detailed, or granulated, data required by RDA may make our bibliographic records a functioning part of the Semantic Web. This could open the door to catalogs that respond to natural language queries, which would be quite a change from our current system of searching on precise terms or the somewhat scattershot approach of using keywords.

So, in summary, RDA offers the possibility of revolutionizing the data structures that underlie our catalog systems, and, by extension, the entire process of resource discovery. As was stated at the beginning of this chapter, the future of library catalogs lies with RDA. And that is why e-book managers should care about it.

REFERENCES

Ehlert, Mark K., "RDA: Why New Cataloging Rules?" *Technicalities* 30 (May/June 2010): 19-20.

Functional Requirements for Bibliographic Records, 2007.

http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm

Oliver, Chris, *Introducing RDA: A Guide to the Basics* (Chicago: American Library Association, 2010).