WHY RDA? ORGANIZING BIBLIOGRAPHIC INFORMATION IN THE 21ST CENTURY

Christine Oliver
Chair, RDA Board Head, Metadata and Processing,
University of Ottawa Library
Ontario, Canada
Christine.Oliver@uottawa.ca

ABSTRACT

Resource Description and Access (RDA) is an international metadata standard designed to enable the discovery of library and cultural heritage resources in both traditional and linked data environments. RDA presents a new way of thinking about bibliographic information. RDA takes the focus of creating metadata to fulfill a user's needs during the resource discovery process. It also breaks down strings of bibliographic information into entities and elements so that this information can be reliably processed as data by computers. RDA is based on a theoretical framework aligned with the internationally accepted bibliographic conceptual model, IFLA LRM; it is designed as a standard for the digital environment; it is developed as a global standard appropriate for use in many contexts.

The scope of RDA has been broadening in response to international interest in the standard. The content of the standard is designed to be flexible, to offer choices and accommodate diverse practices. It has been deliberately designed to support use by an international audience. Using the same international standard also increases data interoperability around the globe. The governance structure has also been revised to support increased international participation in RDA development.

This paper gives an overview of key features of RDA and outlines its impact on the usability and visibility of library bibliographic data. While many still operate in traditional cataloguing environments, RDA prepares us for the future environments in which libraries will function.

I would like us to look at the topic of organizing bibliographic information in the 21st century. What would an appropriate metadata standard look like? I will take RDA as our example, which is very different from cataloging manuals of the past. We will investigate some of its features and see how it is appropriately designed as a metadata standard for recording data and bibliographic information in the 21st century.

Our work in organizing bibliographic information has changed radically over the centuries. In the 21st century, the context for our work is quite different in terms of the expectations we have for our bibliographic information. We know that our users do not go out there and look for bibliographic information only at the library, but rather, they are out on the web. They are interacting with data in a very broad global context, and they expect bibliographic data to be there as well. For this reason, bibliographic data is still managed within traditional catalogs. But it is becoming increasingly also managed outside of those traditional catalogs. We live in a highly networked environment, and no library is an island. No library stands alone. But rather,

we are interconnected, and we use each other's bibliographic information. Our metadata should support linking to the metadata of other libraries around the globe. However, increasingly, we are also blurring the lines between cultural heritage institutions. Because of this, we expect our metadata to be able to link to and work with the data from other communities, such as archives, museums, and publishers. Our metadata needs to be robust enough that it can be harvested not only in ways which we expect it to be used, but also in ways in which we did not initially think about. The idea is that our metadata will be reused in many new and different ways. And it should lead to the creation of new knowledge and new kinds of applications.

Resource Description and Access (RDA) does not present itself as a cataloging manual. Rather, its definition is a package of data elements, with guidelines and instructions on how to record those data elements. It is a standard that comes out of the library community, but it is not intended exclusively for library resources. Rather, it explicitly says that it will enable one to create metadata for library resources, but also for resources from other cultural heritage domains. RDA supports the creation of metadata that can be used in traditional catalogs, but at the same time, it also supports the creation of metadata that can be used in linked data applications.

Consider two particular aspects of RDA: RDA as a standard for current and emerging technological environments, a standard for the 21st century, and RDA as an international standard, a standard that has been intentionally internationalized.

RDA as a Standard for Current and Emerging Technological Environments

In the early days of library catalogs, in the days of book catalogs and card catalogs, we recorded strings of characters. The sole purpose for the strings of characters was that a human being would be able to read and interpret the information. Towards the end of the 20th century, as we started to put these strings of characters online, we witnessed tremendous potential, but also pitfalls. Computers did not necessarily understand the context and the structuring for these strings of characters. The information was not structured as data that was suitable for machine processing. And one of the most Important things that RDA has done was to make us look at our bibliographic information and think of it as data. It is important that it be understood by a human being, but also that it can be understood and processed reliably by a computer. We will also look at how a shared element set provides the underpinnings for data interoperability. While an important aspect of RDA is that it is optimized for link data environments, there is also recognition that there are many different kinds of libraries working in many different kinds of technological environments. And therefore, a standard has to have flexibility that matches what the technological environment is for each library.

The definition of RDA is that it is a package of data elements. And this is a deliberate choice of words: data elements are intended to be understood by humans, but also to be suitable for processing by computers. We all rely increasingly on the reuse of bibliographic data; we rely on managing our bibliographic data through software, by programming, in various applications. I believe we all recognize our increasing reliance on machine processing. But we also all recognize that machines need good data in order to produce results that are reliable and useful.

The data elements in RDA are precise and granular. Each data element has a tightly defined scope, so that there is only one kind of data in each data element. Thus, you have unambiguous data. Also, each data element is precisely labeled. RDA has over 3,000 elements at this point for recording data about 13 different entities.

The data elements are precisely defined, and that makes them suitable for machine processing. These data elements also provide the basis for interoperability. RDA users all share the same element set. They may not all use the same elements, but the elements that they do use come from the same element set. And so, they share the same data structure. There is an inherent consistency and compatibility. In addition, there are also controlled vocabularies that can increase interoperability, and these vocabularies are openly published in the RDA Registry. By having definition and structure for our data, it also makes it much easier to map to non-RDA data. The structure for the RDA elements is based on the IFLA Library Reference Model (LRM). This in turn also increases the basis for interoperability because IFLA LRM is an internationally accepted bibliographic conceptual model. RDA is aligned with IFLA LRM. It is possible that there is another metadata standard that is also aligned with IFLA LRM. With both being aligned to the same bibliographic conceptual model, that also provides a degree of interoperability.

RDA is designed to be flexible. There are four different recording methods when recording data in a data element. Now, not all four methods are applicable to every single data element. But what is important is the principle of the "four recording methods," because this reflects the fact that libraries operate in different technological environments. Thus, each library chooses the recording methods that are most suitable for their own environment.

For example, I can use an unstructured description. I would be using an unstructured description when I am transcribing information from a title page. Alternately, if I were using an automated data capture method to scrape the information from [an online representation of] the title page, my result will be a string of characters.Therefore it is good for the use of a human being because we can decipher and understand what is there. However, we know that a computer does not process it very effectively. In fact, it is only good enough for keyword searching.

At the other end of the spectrum, there is another recording method, which is to use an Internationalized Resource Identifier (IRI). This is an Internet standard that is particularly well suited for the link data environment. It is a globally unique identifier, and one can use it, for example, to record the name of a person. The result is a machine readable string. Now, a human being cannot necessarily decipher and understand it. However, a computer can process it very effectively, and can pull back information that a human can decipher and understand. The IRI can also lead to a large amount of information, and link to other information that is also connected to the same IRI and thus, it expands the potential for exploration.

RDA allows one to create metadata that can be used in traditional catalogs, and also for the web and for link data environments. By using the RDA element set, one can create data that is well suited for machine processing or automated processing. In addition, it is also data that supports interoperability.

RDA as an International Standard

RDA has many evident features that mark it as an international standard. For example, it is aligned with the IFLA international standards such as the International Cataloging Principles and the IFLA Library Reference Model. It is in use in countries around the globe. It is likewise in the process of implementing a governance structure to reflect international participation. Translations and the translation process are an integral part of the standard, and translations are very visible in the RDA Toolkit. An interesting aspect of RDA as an international standard is that it has allowed for the possibility of accommodating different local traditions and the needs of different local cataloging communities.

The content of RDA has been significantly influenced by IFLA's bibliographic conceptual model, IFLA LRM, but it has also been shaped by IFLA's International Cataloging Principles. Within the text of RDA, there are also statements about its relationship to other international standards. And one example would be that metadata created according to RDA can be stored in different recognized encoding schemes such as MARC 21, or RDF.

RDA's alignment with IFLA LRM is a key defining feature of the standard. IFLA LRM was developed through international cooperation and consensus building, as well as going through a worldwide review. Thus, it comes to represent an internationally accepted understanding of the structure of bibliographic data. RDA is based on that internationally accepted understanding. IFLA LRM provides the basic structure for the content of RDA; it provides logical consistency and coherence. RDA also inherits the built-in focus on the needs of the end-user through IFLA LRM's definition of the relationships between bibliographic data and the user tasks.

One of the things that intrigues me about RDA is that it is an international standard that does not expect everybody to do the same identical thing. It is not a case of "one size fits all," but rather the focus is on the common ground that we share. The focus is on data interoperability. These are the words of Gordon Dunsire, that "RDA is designed so that metadata can be created to meet the interests of local communities within a global framework for interoperability." The basis for that interoperability is the shared element set. An additional basis for that interoperability is the fact that the data shares the same structure as the internationally accepted model, the IFLA LRM.

We can see this focus on interoperability right in the text of RDA. There are a series of objectives and principles that govern the design of RDA and one of them is internationalization. The internationalization objective focuses on data harmonization. In the words of the objective, "the metadata provided by different agents will not necessarily be identical." The emphasis is on data harmonization rather than strict compatibility.

RDA intentionally accommodates variations. IFLA LRM provides the basic structure for the bibliographic data. RDA then provides a very large element set and the option for a metadata creator to choose the precise data elements that they need. There are choices about how to record the data in the four recording methods. There are options as to which instructions to use. One can also choose which controlled vocabularies are most appropriate for one's community. And there are also options on how to punctuate and in what order

the elements should be in. Thus, a cataloging community makes the choices that suit their needs, within this framework for interoperability.

A quick example about the options when using controlled vocabularies is as follows: if I am recording data, and I am using the structured description method, there are often instructions on using one of the RDA vocabulary encoding schemes (VES), and a vocabulary encoding scheme is like a controlled vocabulary. Whenever there is an instruction on using one of the RDA vocabulary encoding schemes, it is followed by another option, and that is the option to use another suitable vocabulary encoding scheme.

Directly from the text of RDA, one can see that a cataloging community is offered the choice of either following the option to use the RDA broadcast standard vocabulary encoding scheme, or another suitable encoding scheme, and it will be up to them to choose which of the two options is appropriate for their community.

Since RDA offers so many choices and options, it is important for a cataloging community to record its decisions and to communicate those decisions to its members. There are places right within the RDA Toolkit for a community to record those decisions. However, documentation can also exist outside the Toolkit.

As an international standard that is in use around the globe, RDA has been translated into several languages. In the original Toolkit, there were eight full translations, the translations are being brought over into the new RDA Toolkit. There are new translation projects underway as well. Fourteen language versions are also integrated into the RDA Registry. And that [language versions in the RDA Registry] means these are partial translations. This denotes that the names of the elements and the controlled vocabularies are translated, but not the instructions. English is one of the language versions. The standard places an equal value on all its language versions. English is slightly different because it is the starting point for translations, but it is deliberately called a version.

What I find very interesting is the seamless integration of all the language versions into either the Toolkit or the RDA Registry. And this results in high visibility for all the different language communities that use RDA. One can see all the different languages that are available. And you become conscious of the fact that we are all using the same element set, the same vocabularies as we create our metadata around the globe in our own linguistic communities.

RDA has been designed so that it can be easily used by metadata creators around the globe. However, it is also a standard that should be maintained and developed jointly with representatives from all regions of the globe. And for that reason, RDA governance intentionally incorporates representation from all regions, in order that everyone around the globe is involved in decision-making and management.

RDA meets expectations as an international standard for the 21st century. It is interesting to note that, as an international standard, it does not demand that we produce identical data, nor is there any expectation that we will make the same choices. Rather, there is a recognition that we operate in different contexts, wherever we find ourselves around the globe. They may be different technological environments; we may be working in different linguistic communities; we may be meeting the needs of different national

and cultural communities; they may be communities within the library world or in other domains. But our common interests have always been to provide data that meets the needs of our end-users. And we rely increasingly on data that can be processed accurately by computers. But what is most important is the compatibility of bibliographic data across communities. Data that can be shared, that can be reused, and that can be integrated with data being produced by other communities around the globe. RDA enables us to achieve this compatibility by providing us a framework for interoperability.