RDA: a content standard to ensure the quality of data

Carlo Bianchini and Mauro Guerrini

The user's interest at the center of RDA's objectives

RDA, Resource Description and Access, is the international standard for the description of, and access to, resources designed for the digital word. The first version of the standard was released in 2010.[1] RDA replaces AACR2, *Anglo-American Cataloguing Rules*, 2nd edition, which was in place from 1978 until the early 2000s, with many revisions. The new standard goes beyond the previous cataloging codes.

RDA is not presented as a set of compulsory rules, but as guidelines and instructions developed systematically, and updated continuously, on the basis of new knowledge and different

---

[1] In 2010 and 2013 two loose-leaf editions of RDA were published. For the subsequent editions (2014 and 2015), a bound edition was decided upon and plans for the diffusion of an "update packet" were abandoned, due to the large number of pages affected by changes. On the official RDA website, it states: "Cataloging practice described by RDA has not altered dramatically due to these changes, but over 70 percent of the pages in RDA Print were affected by the changes, making an RDA Print update packet impracticable." http://www.rdatoolkit.org/rdaprint. For other information on RDA, see the websites: http://www.rda-rsc.org; http://rda-rsc.org/content/links-related-resources; http://resourcedescriptionandaccess.blogspot.it; https://resource descriptionandaccess.wordpress.com; https://rdabibliography.wordpress.com.

conventions. RDA is a set of instructions for recording data, designed for libraries and all institutions responsible for recording knowledge, particularly archives and museums. The standard is, however, still affected by the bibliographic tradition of AACR, and it is still structured primarily for resources typically and traditionally found in, and accessible from, libraries. RDA, fully acknowledges the FRBR model – or FRBR MRL, as it is currently known (Riva and Žumer 2015), and the *International Cataloguing Principles* (ICP).

The adhesion of RDA to FRBR and FRAD is significant as it demonstrates that the framework is structured according to a conceptual model based on a bibliographic – rather than a technological – context, thereby ensuring the presence of an enormous number of guidelines that are useful to identify and describe the wide variety of resources in today's market, and to offer high quality authority data. Therefore, librarians and agencies that create quality, structured data for resources will retain an important role in the future, thanks to their wealth of continually updated knowledge of descriptive techniques for the bibliographic world and for the creation of authoritative data.

RDA, in step with FRBR and ICP, focuses its scope on users' interests, providing them with a series of functions to assist resource discovery. To this end, point 0.0 reads:

> RDA provides a set of guidelines and instructions on recording data to support resource discovery. The data created using RDA to describe a resource are designed to assist users in performing the following tasks:
>
> - *find* - i.e., to find resources that correspond to the user's stated search criteria

- *identify* - i.e., to confirm that the resource described corresponds to the resource sought, or to distinguish between two or more resources with similar characteristics
- *select* - i.e., to select a resource that is appropriate to the user's needs
- *obtain* - i.e., to acquire or access the resource described.

Any published or unpublished resource, in any media, is described, identified and formalized by the access points. RDA takes into account all kinds of resources produced and used to transmit recorded knowledge. The standard, in fact, places particular importance on the ever more varied bibliographic world, the principal characteristics of which are typical of most modern publishing technologies, which, for example, allow the same work to be read in differing manifestations, or, in term of BIBFRAME, instances:[2] the same text may appear as a printed book, an e-book, a PDF file, html, or a Word document.

From record management to data management

The recording of data in RDA reflects the process of atomization of the data in the digital environment. With the hand-written, printed or typed catalog card, the (static) record is read exactly as it is written; the ways in which the data is registered and presented coincide. Data entry and consultation are, in fact, two separate activities; as the elements are distinct and, as such, can be assembled and then viewed and read in different (dynamic) formalizations. The static nature of the catalog card, due to the

---

[2] For the concept and the definition of instance, see BIBFRAME https://www.loc.gov/bibframe.

way in which it is recorded (paper), despite allowing an organized representation of the content, does not fully permit the identification and self-management of individual component elements by machines. The advent of information technology has led to major changes as regards the methods of recording data (metadata). For decades, however, the structure of the record has been conditioned by the "paper card" model. Thanks to web-based technologies, from the mid-1990s the development of increasingly advanced OPACs has made it possible to overcome the limit of strict equivalence between the form of data registration and the presentation of the same data, resulting in the creation of search tools that have made the consultation of catalogs increasingly advanced and user-friendly. These transformations have highlighted the enormous benefits and increased flexibility for the presentation of data as separated from data recording.

In RDA, the forms of data registration and presentation no longer coincide. The standard focuses its attention on the concept of *data*, and no longer on the textual record, reflecting the granularity of data typical of the digital environment. A single piece of data, separated from the others, can be now re-used in different contexts to meet different needs.

Now the textual record, formerly rigid and static, turns into a *dataset* in which each piece of data is perfectly identified and described. The data is, therefore, an independent, self-contained element, which is usable (or re-usable) in different contexts and in different representations. With RDA, in fact, there is a transition from *record management to data management*, or from a concept of creating a record to the concept of identifying data and describing every important individual element, thereby opening up the prospect of linked data. RDA and linked data

represent a happy binomial: the guidelines and the technology of linked data are specially designed for new generation indexing tools and for the semantic web. Both of these contribute to building the new architectural structure of information retrieval systems that are able to manage and retrieve information with an enhanced semantic connotation. The new organizational structure of the information ensures that each identified and controlled individual element of data be connected to other data, which are, in turn, identified with appropriate and controlled attributes, in order to create a flexible configuration that is interpretable and re-usable by machines. This enables an architecture in which users, starting from an entity, can develop and satisfy their need for information and can follow the relationships between the entities that interest them.

In particular, RDA provides the possibility to build a wide network of relationships between entities (the reference to FRBR is explicit). Not only the traditional relationships between works and their creators, but a full set of relationships can be identified among different and related works, expressions of a work, associated persons, families, corporate bodies, etc. The expansion of the syndetic, relational structure, already theorized by Charles Ammi Cutter for his cataloging rules in 1876, is an innovative feature and very important – one may say this is the main feature of the new standard. The relational structure is designed to make users more aware of the publishing history of each work through links to its various editions (commented, reduced, illustrated, critical, for children), translations in other languages, the availability of the work in different physical formats and coding and its mode of use. It is a relational graph network that also includes works derived from the original work, such as parodies, remakes, reductions in prose of works in verse, sequels, etc., when they exist. The construction of the network of relationships

has a strong cultural and, primarily, philological connotation. For example, think of the importance and functionality of linking the original title of a work to the titles used in the versions in other languages, which are often so different, as in the case of American films rendered in Italian with completely different titles.

Data recording and presentation

"Identify" and "relate" are the verbs that epitomize the two fundamental objectives of RDA: to identify a resource by the selection of a set of attributes that allow it to be to distinguished from another; and to connect the resource to other associated resources by creating meaningful relationships.

Recording data using a format and support, and *presenting* data, that is to arrange data on a paper card or monitor, are operations that depend on the technologies used. RDA 0.1 is very clear: "In RDA, there is a clear line of separation between the guidelines and instructions on recording data and those on the presentation of data. This separation has been established in order to optimize flexibility in the storage and display of the data produced using RDA. Guidelines and instructions on recording data are covered in chapters 1 through 37; those on the presentation of data are covered in appendices D and E." RDA is, in fact, a *content standard*, which establishes core elements, according to FRBR and FRAD, to describe and to give access to a resource. RDA instruction 0.6.2 reads:

> The RDA core elements for describing resources were selected according to the FRBR assessment of the value of each attribute and relationship in supporting the following user tasks:

- *Identify* and select a manifestation
- *Identify* works and expressions embodied in a manifestation
- *Identify* the creator or creators of a work.

The RDA core elements for describing entities associated with resources were selected according to the FRAD assessment of the value of each attribute and relationship in supporting the following user tasks:

- *find* a person, family, or corporate body associated with a resource
- *identify* a person, family, or corporate body.

However, RDA does not impose a form for the presentation of data, the order of elements, or a conventional punctuation. In short, RDA states *what* to describe but not *how* to present data, unlike AACR2, which, in contrast, placed much importance on the formal presentation of the elements. RDA, therefore, establishes the information required to describe a resource, the content, or the mandatory and optional elements of the description, but it does not deal with methods of visualization. RDA allows freedom of decision on the content of data, from the technology used to register it (paper or electronic), to the format (MARC 21 or UNIMARC and perhaps, in the future, BIBFRAME), and all other aspects of the description that were previously closely related to the description itself. This is not only a technological or conceptual passage, but a cultural distinction that establishes a clear boundary between the management of the meaning and the role of data - the very foundation of the library profession. This is in contrast with the handling of modes of storage and display (purely technological aspects, which can, and should, be left to the expertise of computer engineers).

The visualization of data, however, is not random. The list of the core elements in RDA 0.6 (see 2.6) appears to be an implicit order of presentation for the set of elements. The standard also shows numerous examples structured according to ISBD, which remains one of the possible standards for the presentation of data (appendix D). RDA, in fact, does not prescribe any modality for the output of data. The use of a particular punctuation style for added elements – for example, *Authorized access point of the work*: Bulletin (Geological Survey (South Africa)); authorized access point for family: Austen (Family : Austen, Jane, 1775–1817) – reflects traditions inherited from AACR2 and agreements with the ISSN and ISBD(S) communities. As a content standard, the RDA instruction should, in reality, have indicated only the necessary data, not the form of visualization. On the other hand, it is necessary for the elements to be presented in some way in the examples given by the RDA guidelines. The standard often uses ISBD, which is no longer used for its descriptive system but provides an ordered way of presenting elements that have been identified and defined. ISBD, as we know, comes from a comparison of the descriptive behavior of the eight bibliographic agencies deemed the most representative in the world.[3] It confirms a long-standing practice dating back at least to Charles Ammi Cutter, if not even earlier; a cataloging practice of uniformity that has never excluded the possibility of adaptation to the needs of individual libraries. With RDA, these demands for freedom of presentation are fully recognized under the principle

---

[3] At the International Meeting on Cataloguing Experts (IMCE) Copenhagen 1969 was examined and compared: British National Bibliography, Bibliographie de la France, Deutsche Bibliographie, Svensk Bokförteckning, Bibliografija Jugoslavije, National Union Catalog, Magyar könyveszet, and Boletín bibliographic nacional, Buenos Aires (Gorman 1969, 3).

of user-centricity sanctioned by ICP. Another equally important factor is the consideration of data production by a wide variety of individuals as a value and not a limit. "The more we can share the workload of describing resources internationally, the better our users will be served", Barbara Tillett writes in the Preface to *Introduzione* a RDA (Tillett 2014, 13). While these varied data creators (libraries, publishers, distributors, book shops, aggregators, users) should guarantee significant compliance with the description of information, one cannot insist on uniformity in storage or in the viewing of elements described according to standards designed in different contexts. Systems can display the elements for the name of author or the title of the work in a primary position, formulated in uppercase or lowercase, round or italic characters as deemed most useful to local users; but the core elements of RDA description are required to be present, as shown by the following examples of an authorized access point according to various display possibilities:

- Melville, Herman, 1819-1891. Moby Dick
- Melville, Herman (1819-1891). Moby Dick
- Melville, Herman <1819-1891>. MOBY DICK
- Moby Dick [Melville, Herman. 1819-1891
- Works [Dante Alighieri]
- Works [Wolfgang Amadeus Mozart]
- Works [Federico Fellini]

In the cases shown above, the commutative property could be paraphrased in arithmetic: even if the order of the elements is modified, the functions of the authorized access point are guaranteed. The standard presentation for all disappears; to cite the great lesson of Ranganathan, a personalized description, a contextualized description remains. In this way, it can be said that the concept at the core of RDA is one that has long been a

hypothesis (as stated more or less explicitly in the introductory "Report" to the Italian Cataloguing Rules for Authors (RICA): "The uniformity of the rules leaves enough space to the needs of individual libraries" (Regole italiane di catalogazione per autori 1979, XVII). All of this is today more feasible from a technological perspective. This also reflects the fact that RDA is a content standard.

We should also ask ourselves: are the representations all valid? The response is negative. The arrangement of the data is of marginal importance for the user, who potentially discovers something he had not been aware of and perhaps had not been looking for, but that, once found and identified, proves to be both interesting and relevant. When searching through Google, the majority of users discover, or believe they discover, what they are looking for, despite not knowing how to search effectively and, in many cases, without paying attention to the path they have navigated. In informed searching and explorative searching, the displayed data should be consistent with the value of the data itself, on the one hand; and with the function and use of the data, on the other. In fact, some of the data subsets have a functional coherence and in syntax express a very close link between them, such as that between the name of the creator and the title of the work, or between the place/date of publication and the publisher. Secondly, visualization cannot be anything other than consistent with the function and use that it assumes in the different contexts in which arises: a catalog, a list, a bibliography, or a database. The effectiveness of visualization depends, therefore, on the way in which data is shown for its semantic value, syntactic articulation, and for the overall (pragmatic) manner in which it is presented to the user. These dimensions of cataloguing language must be expressed clearly. Therefore, a correct visualization of the data can be neither accidental nor arbitrary. The question of

information architecture is raised: it should ensure better perception and understanding of the elements. RDA, in fact, overcomes many formalisms that have characterized cataloging over the last forty years, but it would be wrong to think that this justifies just any kind of solution. The way in which data is presented must be effective and compliant with the principles set out by RDA, principles in large part taken from the ICP. The RDA standard also introduces a specific terminology. This is in step with ICP that dispenses with various obsolete, historical terms related to cataloging, such as "record", "header", "uniform title". Indeed, even the term *cataloging* is no longer present in the title - simply Resource Description and Access. This inaugurates a new path that would see all those who create information following the same direction, regardless of the cultural institution or information agency from which they operate: libraries, archives, museums, publishers, aggregators resources, distributors, cultural centers, etc.

RDA pays particular attention to the quality of data. The standard is used to declare, accurately:

- *type of content* (for example, text, annotated music, performed music, cartographic image, computer program, sounds, the spoken word);
- *type of media* (a higher level – the means by which the content is communicated, for example, microform, audio, projection, video);
- *type of format* (a lower level or a specific level, since it refers to the kind of format for a type of media – how the content is packaged; for example, for videos, the format can be: video cartridges, video cassettes, tape reel, video disc);

- *modes of issuance* (how the resource is issued, for example, as a single unit, a monograph in multiple parts, a serial, or an integrating resource).

The *focus* of RDA on the description and access to the resource allows in-depth analysis of the resource as a single entity, with a high level of attention to the identification of its attributes, and at the same time allows the cataloguer to contextualizes that entity in the bibliographic world through the construction of the network of relationships. The two methodological principles of the first chapters of the guidelines - "take what you see" and "accept what you get" - tend towards a faithful transcription of the attributes of the manifestation (and of the item), in the exact form in which they appear on the resource, without manipulation or standardization. For example, the title presented in uppercase on the resource may be recorded in uppercase. These methodological principles also tend towards accepting data in the form in which it is received without changes. For example, data may be derived by automatic processes from a publisher or by optical recognition systems from the typeface of the book title. In fact, RDA instruction 0.1 reads: "RDA is designed to take advantage of the efficiencies and flexibility for data capture, storage, retrieval, and display that are made possible by the new database technologies." RDA is also designed to be compatible with the older technologies still used in many resource discovery applications. The RDA principle "take what you see" is linked to the objective of providing a more accurate and detailed description of the resource than previous descriptions based on international paper card format (the structure and display taken in first generation electronic catalogues), restrained by physical size (paper 7.5 x 12.5 cm), which obliged the cataloguer to select the descriptive elements to be recorded and to impose the

standardization of data, again to save space. In the digital era, the physical size of attribute mapping has no limit, except common sense and the ICP principle of economy. The cataloger is faced instead with a decision about the choice of the elements to be displayed. In this perspective, the "rule of 3"[4] is obsolete. That rule for decades dictated the organization of catalogs and in fact prevented the realization of the catalog's aim to give access to *all* the works of an author (particularly if a person was the third or other subsequent author). In contrast, an RDA description may contain attributes for the all the important descriptive elements for a resource, including an abstract of the work, which typically would not have been included in paper card catalogs. The accuracy of the description is guaranteed by RDA's meticulousness in providing access to the work through related people, organizations, and families, as well as derivative and associated works. The creation of detailed and accurate descriptions and access should therefore trigger a virtuous cycle in favor of the quality of the data, which can be re-used by everyone in any context and domain (interoperability).

Local variation

RDA is intended for use in a global context and yet is flexible and designed to include national options, or cultural and/or linguistic variations that are important to a community, according to the principle of benefitting the user. This is consistent both with the idea of local variation developed by Ranganathan and with the new concept of universal bibliographic control (UBC) accepted by IFLA. The UBC concept recommends descriptions created

---

[4] According to the "rule of 3" in earlier cataloging codes, for works with more than three authors, only the first was transcribed and indexed and others were omitted.

for international sharing with the possibility of differentiated visualizations of the shared data to meet local needs or to express shared conventions in a defined area. Furthermore, in RDA recurrent phrases and terms (such as, *place of publication not identified* or *film, currency, globe, fortnightly, selections, pentagrammatic notation,* etc.) can be represented by URLs that in turn can be displayed automatically in the preferred language chosen by the user, if the search/display system enables this capability. Bibliographic agencies and libraries in the United States, Canada, England, Australia, Germany, and other countries also provide numerous options where RDA allows them, to reflect specific national or regional cataloging traditions, while respecting the same guidelines. RDA, in fact, does not require the descriptions produced from different agencies to be identical, other than for core data elements.

In this way, RDA enhances the great aim of allowing users to discover the resources they want and other resources or persons/families/corporate bodies related to them, making the most of the possibilities offered by the digital environment. This supports the characterization of RDA as the first standard of metadata attribution to be designed for the digital environment and for the Web. It offers the potential to take advantage of the Web to search and discover resources through bibliographic and authority data that could be viewed on smartphones and through related apps, thus encouraging further connections between the users, resources, and information. All of this demonstrates an evolution from previous conventions, the result of over twenty years of theoretical discussion in the international library community (in particular IFLA and ALA), and heralds an interesting and exciting development of the new standard on the international scene for the discovery and transmission of recorded knowledge.

References

Gorman, Michael. *Bibliographical data on national bibliographic entries. A report on descriptive cataloguing made for UNESCO and IFLA. Provisional abridged text.* Copenhagen, 1969

Regole italiane di catalogazione per autori. Roma: ICCU, 1979

Riva, Pat, and Maja Žumer. "Introducing the FRBR Library Reference Model." *IFLA WLIC 2015.* Cape Town. http://library.ifla.org/1084/1/207-riva-en.pdf.

Tillett, Barbara B. "Prefazione." In *RDA: Resource Description and Access. Linee guida per identificare e collegare entità nel web semantico*, edited by Carlo Bianchini and Mauro Guerrini, 11–13. Milano: Editrice Bibliografica, 2014

CARLO BIANCHINI, Università degli Studi di Pavia. carlo.bianchini@unipv.it.

MAURO GUERRINI, Università degli Studi di Firenze. mauro.guerrini@unifi.it.

ABSTRACT: RDA Resource Description and Access are guidelines for description and access to resources designed for digital environment and released, in its first version, in 2010. RDA is based on FRBR and its derived models, that focus on users' needs and on resources of any kind of content, medium and carrier. The paper discusses relevance of main features of RDA for the future role of libraries in the context of semantic web and metadata creation and exchange. The paper aims to highlight many consequences deriving from RDA being a content standard, and in particular the change from record management to data management, differences among the two functions realized by RDA (to identify and to relate entities) and functions realized by other standard such as MARC21 (to archive data) and ISB (to visualize data) and show how, as all these functions are necessary for the catalog, RDA needs to be integrated by other rules and standard and that these tools allow the fulfilment of the variation principle defined by S.R. Ranganathan.

KEYWORDS: RDA; Resource Description and Access; metadata; content standard; data management; principle of local variation.