The library catalogue as a bibliographic tool: document analysis and error corrections as parts of an enhanced workflow

Stefano Bolelli Gallevi

The library catalogue should be, first of all\(^1\), a really efficient and correct bibliographic tool\(^2\); as librarians, we need to pay attention to internal consistency of the catalogue, developing the best possible cataloguing workflow to obtain not only records’ (both descriptive and authority ones) accuracy, but also time saving procedures and tools for cooperative and network cataloguing.

In this article, I will try to make a proposal for an enhancement of cataloguing workflow\(^3\) focusing on document analysis and error corrections.

---

\(^1\) See Bade (2012) for a critical discussion on IT expected influence on library catalogues design.

\(^2\) Petrucciani (2006): “In today centuries-old professional librarianship, the library catalog is a structure devoted to the systematic control of publications” From here on, translations from Italian language are mine.

\(^3\) Petrucciani (2012): the “real life cataloguing”.

JLIS.it Vol. 6, n. 2 (May 2015)

DOI: 10.4403/jlis.it-10250
Cataloguing now

Having a look to poor quality of present library catalogues⁴, actual cataloguing workflow can be described with a simple flowchart.

![Cataloguing flowchart](image)

According to the goal of increasing library catalogue accuracy, this workflow should be enhanced with at least two tasks: publication analysis and error handling.

Before going to consider error correction in next part of the paper, here it is my purpose of an enhanced cataloguing workflow, using a more complex flowchart based on the above simple one.

---

⁴ See Bade (2008) and Petrucciani (2012).
JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 144
Both publication analysis and record analysis depend on cataloguing rules, that are the instrument devoted to give the cataloguer the know-how necessary to perform his job\(^5\); I will discuss the role of publication analysis in both REICAT (2009), the Italian cataloguing

\(^5\) Petrucciani (2012): “a code of cataloging rules, that is, an effective reference and training tool for real-life cataloging in libraries, bibliographic agencies, and similar offices or institutions”.

Figure 2 Cataloguing flowchart enhanced
code, and RDA (2010), the international “guidelines to represent and discover resources”\(^6\).

According to the Italian cataloguing rules, REICAT, document analysis appears as a very important part of the cataloguing job: cataloguer is claimed to start from copy in hand (part I, chapter 1) and to (or at least try to) understand it, starting from material characteristics as copy completeness, how it was issued, continuing (chapter 2 and 3) on how to describe it and where to keep information. In addition to that, document analysis is also the basis to define the access points (e.g. uniform titles and names) and the relations between them, as shown by the sequence of the Italian rules. In other words, REICAT follows the “cataloguing tradition [that] has pointed out, until present days, which data have to be recorded and in which way they have to be displayed”\(^7\).

RDA seems to pay no attention to publication analysis, instead: the summary\(^8\) shows a lack of parts devoted to such analysis (as made by the recovery of copy/item peculiarities, followed by the reconstruction of the way it was published or issued, and the ascertaining if it’s a monograph or a serial publication), while REICAT\(^9\) discuss such task in detail in the part I (particularly 1.2, 1.4 and 1.5). RDA chapter 1 is definitely only an introduction, (mainly

---

\(^6\) Bianchini and Guerrini (2014), other title information; see also Guerrini and Bianchini (2015).

\(^7\) Trombone (2014); Petrucciani (2012) explains the reason underlying such tradition: “The bibliographic description itself, while in some parts an aggregate of simple unrelated elements, is mostly and as a whole a structured text, not a set of data elements: its parts may be related to each other and, as in any text, their meaning may depend on previous parts or on their position with respect to other elements”.

\(^8\) See Trombone (2013) for a critical presentation of the RDA summary in Italian.


JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 146
terminological) to publication analysis\(^\text{10}\): “in RDA [the] workflow for the description of the publication and the item does not exist: guidelines prescribe which data have to be described, while how they have to be displayed and in which order depend on technological choices, and institutional too, of data producers”\(^\text{11}\); moreover, “RDA is structured like a long enumerative list”; Bianchini and Guerrini 2009 clarify that “RDA, even if it declares that its main focus is data content, deals also with presentation of data” and “RDA’s choice to consider ISBD […] as a display format for cataloging data, means […] making a cataloging revolution as great as in the past, but in this case regression not a progression”\(^\text{11}\). Connection between data display and cataloguing workflow (i.e. publication analysis) is made by Petrucciani 2012: “The bibliographic description itself, while in some parts an aggregate of simple unrelated elements, is mostly and as a whole a structured text, not a set of data elements: its parts may be related to each other and, as in any text, their meaning may depend on previous parts or on their position with respect to other elements”.

Differences between REICAT and RDA referring to publication analysis may depend from the latter’s goal to produce guidelines not specific for library environment\(^\text{12}\). Petrucciani 2012 advises from risks in producing a professional tool, particularly a cataloguing code, unspecific and directed to other professional communities: “It

---

\(^{10}\) To make just only one example, RDA “1.1.3 Mode of Issuance” just refers to “single unit”, “multipart monograph”, “serial”, “integrating resource”, while REICAT “1.4. Mode of publication or issuance” refers to “publications with separate parts or pieces”, “multipart publications”, “integrating publications”, “accompanying material”, “supplements”, “publications issued or marketed together”.

\(^{11}\) See Bade (2012) for semiotic problems related to FRBR/RDA and even XML separation of form and context.

\(^{12}\) “RDA has the ambition to present itself as a unique code to register data: for resources that can be found in libraries […], in archives […], in museums […], and for resources produced and disseminated using digital technologies” Bianchini and Guerrini (2014), p. 20; Trombone (2014) underlines that “RDA […] scope is metadata creation for resources transverse to contexts and disciplines”.

JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 147
was felt that a more generalized approach, not focussed on the materials for which libraries have specific responsibilities (as stated in Italy and other countries by legal deposit and library legislation) and not targeted to specific library needs, would not be an effective tool for the library community. Conversely, other communities (e.g., those of archivists or catalogers of art objects for the national cultural heritage database) will also have their specific needs and would like to develop their own professional tools (and usually have them already). The development of professional tools by one community for another (or for many others) is usually ineffective and sounds unfair13; even in terminology, crucial for document analysis and ascertaining of entities and attributes in general, RDA seems at the opposite of the “conservative” choice of REICAT, falling in problems derived by use of “fashionable” (Petrucciani 2012) terms like “resource”14.

In conclusion, only a professional reference tool with a strong structure15 and with a deep theorectic background devoted to

---

13 See also Mitchell (2013): “although RDA is the focus of library cataloging, alternative standards such as Cataloging Cultural Objects (CCO), the archival standard DACS, and the Cultural Objects Name Authority (CONA) standard being developed by Getty are addressing vocabulary, record relationship, and encoding issues in different ways for their own communities. The adoption of these standards in their communities pose considerable challenges for interoperability and adoption for organizations that work with a diverse range of materials and patrons. While some research examines the technical methods by which these standards can interoperate, there are considerable philosophical differences behind these standards. How LAM and other cultural heritage institutions research and resolve these interoperability issues are likely to be an interesting area of research and experimentation in the coming year”.

14 See also Trombone (2014) for the large covering of the term: “Depending on the mode of publication a resource may indicate a single physical unit, a multipart monograph, a serial, a (sic) resource enhanced in an integrating mode. RDA uses the term resource instead of entity in all the definitions of the user functions”.

15 Trombone (2013) demonstrates with clear and detailed examples how in RDA the serial numbering of the summary is not revealing any systematic structured presentation of the concepts, while some grouping (i.e. systematic) titles are JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 148
libraries, as REICAT, seems able to design the enhanced cataloguing workflow presented before\textsuperscript{16}. Nevertheless, RDA, with the continuous effort to updating\textsuperscript{17}, is exemplar on how library community should care of the maintaining of cataloguing rules and job in general.

**Problems in real cataloguing\textsuperscript{18}**

In day-by-day cataloguing, every research on a library catalogue is an occasion to discover, in bibliographic and in authority entries, a huge amount of mistakes, especially in printed books and movies descriptions. And it is not rare to doubt whether the record the cataloguer is looking at has been created for the same edition of the copy-in-hand, or for a different one.

\textsuperscript{16} Here there are ten examples of concepts – useful in “real life cataloguing” and particularly relevant to publication analysis – deepened in REICAT while treated in brief – or with no treatment at all – in RDA: 1) work and expression definition, 2) the systematic attention to modifications originating a new expression, with attention to the limits occurring between works/expressions connected but distinguished, 3) the relationships (different from ownership and custody) that can occur between an author, both personal or corporate, and an item, 4) instructions for multilingual (and multiscrypt) managing of data, 5) detailed corporate author types and their authorship treatment, 6) series extensive treatment (i.e. as a type of publication), 7) definition and separate treatment of multipart publications (joint with analytic description), 8) definition and separate treatment of unpublished documents, 9) collected uniform titles, 10) the systematic structuring of description notes.

\textsuperscript{17} Shortly after REICAT publication, the RICA commission, devoted to its creation and planned to be permanent, was discontinued. See http://www.iccu.sbn.it/opencms/opencms/en/main/attivita/gruppilav_commissioni/.

\textsuperscript{18} This part of the paper was the object of a poster presented to the conference Faster, smarter and richer. Reshaping the library catalogue. International conference, Rome 27-28 February 2014 (http://www.aib.it/attivita/congressi/fsr-2014/2013/38189-fsr-programme/).
In this part of the paper, I will show some examples that I came across during the cataloguing job, in which document analysis showed problems hard, maybe impossible, to solve without some research on other copies of same and/or different editions, often requiring a direct contact with other cataloguers\textsuperscript{19}.

Example 1:\textsuperscript{20}

The only difference between the two DVDs is the series statement “I maestri” on the left one, noticeable only having both images.

---

\textsuperscript{19} In example description I do not mention all minor corrections needed by records.
\textsuperscript{20} Please take a look at article's supplementary files for higher resolution images http://leo.cineca.it/index.php/jlis/rt/suppFiles/10250/0
JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 150
Example 2:

Despite the numbering is not far from titlepage, the cataloguing for the two volumes was not coherent in SBN, where numbering was missing for both volumes, and for volume II the title “La chanson française” was incorrectly registered as subtitle: consequently, no link was made between the two volumes. To obtain a correct cataloguing of volume II, it was necessary first to check the copy of the library creating the wrong record, and then to correct it; the colleague that make such correction had to be recontacted to fix minor inaccuracies (es. Accents).

Example 3:

21 Library owning the volume I was contacted as well because of missing numbering.
Looking as different editions of two different works of the same author, these two volumes are reprints of the first edition of the two parts of the work “Féerie pour une autre fois”: published in 1952, the first part, measuring 21 cm, has only the title of the work, evidently accompanied by a number designation represented by the symbol “*”; second part, published in 1954, has a completely different physical appearance: only 19 cm, title of the work in smaller characters, numbering expressed in roman numbering “II”, a title of the part, “Normance”, in evidence. There is no evidence of any precedent publishing of the second part, nor republishing of the first according to the second. Obviously there is a chaos in SBN OPAC, and even BNF records aren’t well-formed, but even wanting to obtain a better situation, I was in doubt on how should these two books be catalogued; should they be considered a single edition with differences between the two parts or two different editions linked? Even more, they are conveying two different works linked or a single one in two parts and only one with a particular title? After having contacted the three colleagues of libraries that created the records, we decided for first option, resulting in one level cataloguing registering all differences between parts. In the correcting process there was a record to delete but wrong record is still present; I am trying to recontact the colleague to complete the correction.

22 This example has another problem, the “nrf” on title pages: is it a series statement or something else? I do not discuss here because it requires more research; even BNF have only a few records registering as a series statement. Gallimard website shows as even today books with “nrf” in title pages are published, most of them related, by the publisher, to a series called “blanche” on the site but in no one book.

Example 4:

San Antonio is the protagonist of many novels by Frédéric Dard, that signs them with the name of his character. In this Fleuve noir edition, the novel is claimed to be part of a series but without naming it. Finding the series title is not easy, especially having only a few novels: some OPACs use "San Antonio", some others "Fleuve noir", some use both and some don't register any series at all... It is quite sure that, in more than a hundred titles (often in more than one edition), presentation on books changes, but a few libraries have many of them... I decided to contact a colleague that has some of them, and after some weeks I had a feedback claiming for “Fleuve noir” being the series name. After asking for the reason of the choice, answer was that “Fleuve noir” is the publisher and “San Antonio” the series title.

These examples show not only how poor is accuracy in Italian National library catalogue SBN, but how hard and time expensive is to correct errors and mistakes in the present situation.
Starting from the assumption that the library catalogue\textsuperscript{23} is one of the main, if not the most, important instrument that librarians offer to the wide social community, that no other professionals will care of the quality of it, and that library catalogue is different from any commercial and/or generic finding tool Google-like\textsuperscript{24}, it is very important that librarians start to pay attention to cataloguing job, claiming for better instruments and building a better workflow and procedures for catalogue maintenance.

Most of the errors presented in above examples (and other more) would be avoided if the cataloguer could look at the information sources of the copy used as a basis for the catalogue record; moreover, easiness in contacting other cataloguers would permit to quickly signal, discuss and decide eventual corrections needed by the record, who will made them and so on.

Decisions and corrections should be traced and be available for other cataloguers, and the most relevant ones should be presented to the worldwide cataloguing community.

Unfortunately, in present situation, as I will try to show in next part, Italian cataloguers have insufficient tools and support to do their job at the best.

\textbf{Tools for digital cataloguing: what Italian cataloguers have now and what they should ask for}

In Italian libraries, the main digital tools for cataloguers in the perspective of maintaining accuracy in catalogs are:

\begin{itemize}
\item the software they use for cataloguing;
\end{itemize}

\textsuperscript{23} Library catalogue is intended here in a generic, but specific from the librarians community perspective, as the specific tool that all libraries have, or should have, to organize, and give access to, their documents.

\textsuperscript{24} A specific comparison between a library catalogue and google see Mann (2005); Yee (2007) discuss difference between metadata creation and cataloguing.
reference tools as the LOC Cataloger’s desktop, WebDewey Italiana, the BNCF Nuovo soggettario etc;

- collaborative initiatives (Fairclough 2013) like mailing lists (both institutional or not), blogs, forums.

All of them have specific problems that should be solved to enhance the cataloguers’ job; first of all, they lack in integration between themselves, i.e. cataloguers have not the possibility to use them in an integrated and collaborative way; moreover, each cataloguing software have different capabilities and the shortage in standardization results in usability problems.

To enhance cataloguing job, some guidelines should be produced on what any cataloguing software should do and what can be considered optional; then some documents on cataloguing workflow and processes organization should be produced, or enhanced with the publication analysis and error handling tasks. Cataloguing software should be regularly tested and evaluated. Integration between tools should be asked to vendors, as well as functions oriented to the communication between cataloguers (i.e. chat) and the use of images of the publications being catalogued should be evaluated and enhanced.

---

25 For example, Aleph requires a windows-only app, Koha works with any browser.


27 With digital cameras, smartphones and apps is really easy, fast and cheap take pictures of publication and snapshots of electronic publications and/or OPACs; I use, even for images showed in this article, my IPhone and the app Genius scan, in the free version.
Cataloguers themselves should enhance collaboration, standardize cataloguing level\(^{28}\) and claiming for better and more efficient support, especially in participated environment\(^ {29}\); attention should be paid to collaboration with “IT people”\(^ {30}\). For example, to improve error correction in SBN, the present function “proposta di correzione” (proposal of correction) could be redesigned for library codes automatically provided (now the proposer has to manually copy and past from SBN OPAC), direct providing of cataloguer’s contacts (now the proposer has to find the colleague adress through the web), possibility to manage images and, last but not least, sending advices (email, sms...) that give feedback of the sending of the proposal (now any library have to manually check if it has received any proposal).

\(^{28}\) I heard a colleague, cataloguing in SBN, saying that ISBN was not recorded for a precise choice. In SBN records made by BNI (Italian National Bibliography) we could find reprint designation in edition area, despite REICAT does not allow it; when I claimed to correct it to avoid record duplication, one of the BNI cataloguers answered that it was not possible because it’s a record with the designation of National bibliography number; I do not know if BNI, or other libraries participating to SBN, should have particular cataloguing practices in conflict with catalogue quality, but I think that if there is a good reason for it, this information and consequent cataloguing practices should be diffused to all other SBN libraries.

\(^{29}\) See the AIB document “AIB per il #nuovoSBN” (http://www.aib.it/attivita/2013/39838-nuovosbn/), that is a good start for claiming for improvement of all SBN aspects and tools, participated catalogue included; see also the answer of ICCU (http://www.aib.it/attivita/2014/41208-iccu-nuovosbn/), hardly insufficient on such topic; in particular, note that ICCU answer to point 6 of the AIB document focuses on the importance of translating RDA in Italian, despite such task, for the Italian cataloguing environment, should be considered low in priority, having precedence, just for giving one simple example, to the publication of REICAT online tool (presented in conferences like REICAT: contenuti, applicazione, elementi di confronto: Seminario REICAT. Roma, Biblioteca Nazionale Centrale di Roma, 18 febbraio 2010); in the AIB document RDA translation is no mentioned at all, claiming for just a deep analysis of the RDA attention to open linked data.

\(^{30}\) See http://blog.jweinheimer.net/2012/08/rant-on-cataloguers-and-it-people-was.html and http://librarylab.law.harvard.edu/blog/2012/05/24/quality-rules/ for informal discussion of the problem; in Bade (2012) IT design problems are discussed from a theoretical point of view.
Conclusions

If the efficiency of cataloguing workflow is a crucial factor to build a library catalogue with clear and correct descriptions of the materials owned by the library, then back-office cataloguing needs to be as fast, rich and smart as the catalogue is.

In the past, cataloguing was a well standardized practice in the offices, though mainly on a practical basis; nowadays, when attention is paid to management problems in library, back office cataloguing, and catalogue with it, have to be rethought to became more efficient and accurate.

On the theoretical way, as Petrucciani pointed out, “Better cataloging rules and the enhancement of bibliographic formats and software programs are necessary to achieve a permanent, authoritative, and effective record of the whole published output of human knowledge and creativity” (Petrucciani 2012). Now that important cataloguing rules start to be evaluated even in a comparative way, it would be important to better understand how comparison should be conducted: Forassiepi 2015, for example, points out, as a problem, how REICAT is devoted to printed material; even if it was true31, it could be seen as a point of strengthness, considering not only that printed material is far from disappear from libraries, but also how it is hard (even nowadays) to catalogue (as I hope examples above have shown); moreover, digital publications, i.e. ebooks or digital versions of periodicals, are far to be stable in their paratextual elements32, if not merely transpositions of print originals, so that

31 As Forassiepi himself cites, REICAT deals with “publications of any kind and in any media”.
32 As an example can be considered the extension in pages of ebooks and digital articles: sometimes there is one (but page numbers could not remain constant when you enlarge the text or if page numbers are device-specific), sometimes there is not (i.e. Plos one, http://www.plosone.org, articles).
attention paid to cataloguing of print material can be considered as a task useful to any cataloguer.

On the practical side, cataloguing should be based on the assumption that examples are both important to build good cataloguing rules and to apply them, i.e. all examples in cataloguing rules and in manuals\(^{33}\) should be kept from real publications, and attention has to be paid on the presentation of them. Communication between cataloguers should be considered the basis to effort catalogue accuracy, and the cataloguing workflow should be standard for all cataloguers.

Cataloguing codes and cataloguer’s workflow enhanced on the side of publication analysis and development of collaborative tools and practices, would permit to reduce differences between cataloguing and descriptive bibliography focused by Yee 2007, with the aim to enhance the bibliographic accuracy of library catalogues: if large OPACs are similar to bibliographies, cataloguers have to be(sim(e) similar to bibliographers too\(^{34}\); or, at least, they should take into consideration their point of view and their skills.

---

\(^{33}\) In Italy some books were published to help in application of past cataloguing rules RICA showing and discussing practical cataloguing examples; see Quaderno RICA (1981), Quaderno RICA – ISBD (M) (1981) and Petrucciani and Turbanti (2006).

\(^{34}\) Petrucciani (2012) particularly in chapter “Work and expression: unsolved questions and open issues” clarify the importance of textual studies in taking critical decisions for the Italian cataloguing code; see also Petrucciani (2006) for claiming for cataloguers, and librarians in general, paying attention to bibliographer G. Thomas Tanselle writings. See also Yee (2007) for discussion of Tanselle writings relating to cataloguing environment.
References


———. 2012. «From the FRBR model to the Italian cataloging code (and vice versa?)». *Cataloging & Classification Quarterly* 50 (5-7): 603–21.


Yee, Martha M. 2007. «Cataloging Compared to Descriptive Bibliography, Abstracting and Indexing Services, and Metadata». *Cataloging & Classification Quarterly* 44 (3/4): 307-327

*JLIS.it. Vol. 6, n. 2 (May 2015). Art. #10250 p. 160*
STEFANO BOLELLI GALLEVI. Alma Mater Studiorum – Università di Bologna stefano.bolelli@unibo.it.


ABSTRACT: The paper aims to discuss cataloguing workflow as a part of catalogue improvement.

As a librarian involved in day-by-day cataloguing, every research on a library catalogue is an occasion to discover, in bibliographic and in authority entries, a huge amount of mistakes, especially in printed book and film description, and it is far from rare to have the doubt if the record I am looking has been created for the same edition of my copy-in-hand, or for a different one.

Referring to the Italian (i.e. REICAT) situation but trying to having a look on the international (i.e. RDA) one, there will be discussed some practical examples of book and film cataloguing, on the assumption that examples are both important to build good cataloguing rules and to apply them.

Particular attention will be paid to record duplication and inaccuracies, trying to understand how to avoid them improving collaboration between cataloguers.

KEYWORDS: Cataloguing; Data quality; Inaccuracy; REICAT; RDA.