

Should catalogues wade in open water?*

Paul Gabriele Weston^(a)

a) Università degli studi di Pavia, <http://orcid.org/0000-0001-9134-2839>

Contact: Paul Gabriele Weston, paul.weston@unipv.it

Received: 18 April 2021; **Accepted:** 11 May 2021; **First Published:** 15 January 2022

ABSTRACT

In recent years, libraries, either on their own or in consortia, have carried out digitisation projects which resulted in establishing criteria to make digital items accessible through the catalogue. Pushing the boundaries of the latter, cataloguers have considered the possibility of providing access to the digital version of a work whenever available in the public domain. Librarians have now started to question whether the catalogue, moving past the idea of being just a citational tool, should open itself to the web as the place where users, thanks to quality data, can gain easy access to freely available digital bibliographic material. This should include digital publishing, as well as DH projects, all of which are based on editions published in printed format.

This scenario urges to find quick policy answers: a. how should features which could act as search keys or filters be adequately described; b. how should flexibility and changeability of digital objects be dealt with; c. how traditional cataloguing procedures should change as a consequence of the number and the peculiarities of these items; d. which criteria should be adopted in marking the new border lines of the library / catalogue mission.

KEYWORDS

Digital resources description; Metadata management; Digital preservation strategy; Professional education; Catalogue mission; Digital resources retrievability.

* To the everlasting memory of Ottavia Calini, who should have discussed her master thesis on these topics at Ca' Foscari University of Venice.

Setting the scene

The difficult relationship between catalogue and digital production has recently been the subject of reflections initiated within committees and study groups, reflections that were then shared with the library community through conferences and seminars, and professional literature. The issues are far from simple to solve, as they see an overlap of technological factors, cataloguing rules, standards and formats, and procedural choices. A thorough impact assessment of the above-mentioned factors on either the structure or the function of the library catalogue would go far beyond the scope of this paper. Nor is it possible here to ascertain whether and to what extent the changes taking place in the cataloguing rules, data coding structures and information retrieval systems comply with the task of representing the elements of the bibliographic information, which underlies the principles of cataloguing.

A passage from Diego Maltese's introduction to Trombone (2018, 11) can be taken as the starting point of these reflections:

“There's a difference between the library catalogue and a data archive. Equipping the semantic Web with a specific and even sophisticated search engine for resources of all kinds is certainly important, but it is not and should not be, in my opinion, among the tasks of the library.¹”

Maltese's observation is part of his broader discourse on the concepts of what is inside or outside the boundaries of libraries and catalogues and, consequently, the activities of librarians. Is it the task of libraries to provide data for the semantic web? If so, how important should this activity be considered among those carried out by libraries?

Or wouldn't it be better or wiser to direct intellectual, planning and creative efforts towards improving and refining the search tools of the library tradition and to entrust to the web a more or less wide part of indexing and also the retrieval of descriptions of resources or of the resources themselves, if these are electronic resources? All the more so because, as Sardo (2017, 9) puts it: “new players not previously present on the scene of document management burst forcefully and outclass libraries.”²

Inconsistencies in digital resources cataloguing

The taboo of describing electronic and digital resources in catalogues has been almost absolute for a long time for a wealth of reasons. In the first place, following a scheme that has occurred whenever a new type of material has shown up, doubts arose on whether the catalogue should include this kind of material. Subsequently, however, the cataloguers experienced some uncertainties as to which criteria to adopt to identify the type of record and the type of material, uncertainties also due to the rapid technological changes and the need to distinguish between the new emerging categories of electronic resources.

¹ “C'è differenza tra il catalogo di biblioteca e un archivio di dati. Attrezzare il Web semantico di uno specifico e persino sofisticato motore di ricerca di risorse di ogni genere è certamente importante, ma non è e non deve essere, a mio avviso, competenza della biblioteca.”

² “Altri attori prima non presenti sulla scena della gestione documentale irrompono prepotentemente e surclassano le biblioteche”.

Today there seem to be two categories of resources that run the risk of being underrepresented or, worse, represented unevenly in the catalogues: these are digital reproductions of printed books and digital editions of textual works available for free on the web.

As far as digitisation is concerned, it would appear to be an optimal solution to indicate its existence by adding a note accompanied by a link in the description of the physical item from which it was taken. For the cataloguer, this process takes a few seconds, since it is sufficient to insert the *uri* of the digital equivalent in a note or a specific field. Once the description of the physical resource has been identified, the user is made aware of the existence of a digital reproduction.³

The image shows a library record interface on the left and a scan of a book's title page on the right. The record includes fields for bibliographic level, material type, author, title, publication, physical description, language, country, imprint, notes, and a table of holdings. A link to a digital reproduction is provided in the 'Note' field.

Biblioteca	Collocazione	Inventario	Note all'inventario	Fruizione
Biblioteca Nazionale Braidense	RACC.DRAM.T.045	5 050000350	Legatura in pelle. - Note ms.	consultazione e fotoproduzione Richiesta

The title page scan shows the text: 'IL NEGROMANTE. COMEDIA DI MESSER LODOVICO ARIOSTO.' with a portrait of Ariosto and the date 'M D XXXV.' at the bottom.

Fig. 1. Link to the digital reproduction of a copy from the record of the paper edition (Source: Opac of the Biblioteca nazionale Braidense, Milano, Italy)

The fact that, as a result of its digital acquisition, the reproduction is formally identical to its original source when displayed on the screen, leads us to think that this type of resource can be considered equivalent to a set of photocopies, a microfilm or a microfiche and thus treated in the same way.

Is that true? Should a digital reproduction be considered the equivalent of the printed item from which it has been scanned?

To answer this question, a number of issues should be addressed:

³ In the December 2020 revision of MARC 21 Bibliographic, the use of field 856 (Electronic location and access) is defined as: "Information needed to locate and access an electronic resource. The field may be used in a bibliographic record for a resource when that resource or a subset of it is available electronically. In addition, it may be used to locate and access an electronic version of a non-electronic resource described in the bibliographic record or a related electronic resource. Field 856 is repeated when the location data elements vary (the URL in subfield \$u or subfields \$a and \$d, when used). It is also repeated when more than one access method is used, different portions of the item are available electronically, mirror sites are recorded, different formats/resolutions with different URLs are indicated, and related items are recorded." (Library of Congress. Network development and MARC standards office 2020)

From the point of view of the “physical” characteristics of the resource, the answer is negative. The analogue resource, like, for example, the paper book, has its own physical characteristics – the number of pages, the size or the weight – and they are not replicated in the digital object. The dimensions in cm, the rendering of the colours or the weight (these last data not included in the catalogue record) in the digital resource are simulated and suggested, respectively adding a ruler to the video images, a colorchecker, or showing the consistency of the book cut to make the idea of its thickness. When, instead, a viewer allows the reader to directly reach a specific page, the operation is the result of the correspondence created between the specific numbered page and the corresponding digital image. The equivalence between the analogue object and the digital object is then artificially reconstructed for the benefit of those who consult it from the screen.

Even from the point of view of descriptive elements, the scanning of a paper object (but it could be a parchment or a clay tablet) produces a new resource with its own characteristics, starting from the name. Only in early days was it thought that naming the digital object after the name of the analogue resource (for example its title) could be an appropriate solution. For years now file naming has been following criteria unrelated to resource identifiers. For the digital object as a resource in itself, not as a substitute for the analogue resource, in addition to a name that is its own, it might be possible to identify a creator, namely the institution responsible for the digitization project, as well as the entity responsible for its material realisation (for example, the firm that carried out the scanning).

Another crucial element for a complete and correct description is the date of realization of the resource. In the case of digitisation, it is highly unlikely that it coincides with that of the analogue resource, copyright issues being among factors which tend to favour scanning of older resources, not to mention cases where digitisation campaigns are part of special preservation projects of very old originals. The gap between the date of creation of the physical object and that of the digital reproduction is therefore substantial.



Fig. 2. (on the left) *De Arte Venandi cum avibus*. Ms. Pal. Lat. 1071, Biblioteca Apostolica Vaticana. Graz: Akademische Druck- u. Verlagsanstalt, 1969; (on the right) digital reproduction of folio 49 recto (Biblioteca Apostolica Vaticana, scan date: 23.11.2009).

What users should be entitled to know

These considerations lead us to believe that it is not appropriate to subordinate the existence and the retrievability of the so-called digital reproductions to their analogue counterparts. But this is what happens regularly. Very few catalogues describe the derived digital objects for what they are, that is, sets of images with specific technical characteristics. Yet, nothing would prevent connecting the analogue object to the derived digital object and providing readers with clear instructions. As it is a right of the latter to be able to identify the existence of one or more digitizations starting from the description of the analogue resource, so it must be equally possible to search and filter the digital resources for the characteristics that are their own, such as the date of creation, an element that could affect in a decisive way the quality of the images and the available exploitation devices, or as the technical characteristics of the images (master and derived) that make up the scanned item. Users may be interested, today and even more in the future, to search for objects created in a given period, as part of a specific project or with specific technical characteristics, not as surrogates, but as objects with meanings other than those of the analogue object. The implications of the application of new IT techniques to the processing of data, both for the purpose of managing digital repositories, and in the process of providing navigation clues to the users, are yet to be fully assessed.

The screenshot displays a digital library record for a book. The top left shows a zoomable image of the book's title page, which reads 'DIE AUSSTELLUNG VON MEISTERWERKEN MUHAMMEDANISCHER KUNST IN MÜNCHEN 1910'. Below the image are navigation controls (Zoom, Rotate, Print) and a title 'Die Ausstellung von Meisterwerken... Bd. 1, [Title page]'. The main content area is divided into several sections: 'ITEM DATA' and 'CITE THIS ITEM' at the top right; a detailed metadata section with fields for 'TYPE OF RESOURCE', 'GENRE', 'DATE ISSUED', 'DIVISION', and 'EDITOR'; a 'LIBRARY DIVISION & COLLECTION WITH THIS ITEM' section; a 'VIEW THIS ITEM ELSEWHERE' section with links to the Digital Public Library of America and NYPL Catalog; a 'RIGHTS STATEMENT' section; and an 'ITEM TIMELINE OF EVENTS' at the bottom showing a horizontal timeline with markers for 1812 (Author born), 1912 (Issued), 1912 (Created Item), 2013 (Digitized), and 2021 (Retrieved by you).

Fig. 3. The Miriam and Ira D. Wallach Division of Art, Prints and Photographs: Art & Architecture Collection, The New York Public Library. “Die Ausstellung von Meisterwerken...” New York Public Library Digital Collections. Accessed April 11, 2021. <https://digitalcollections.nypl.org/items/510d47e3-84c9-a3d9-e040-e00a18064a99>. The richness of the information and the effectiveness of their graphic layout and, at the bottom of the screen, the time-line that highlights to the reader the significant dates in the creation of the work (author’s birth and death, paper edition, digital reproduction, consultation)

The characteristics, moreover, do not always have connotations of invariability: with the passing of time, whereas it is unlikely that the same institution decides to carry out a new scan of the same item, it is entirely possible that some characteristics of the images, in particular those made available on the web, are modified, such as the format of the file, or that its quality is increased, and therefore the weight, in view of higher performance of computers and connections available to users. There are other elements that have a great relevance in terms of accessibility and that are often linked to the display context: just think of the presence or absence of a menu that provides the document structure, or features such as zoom, OCR, image editing tools (contrast, brightness, etc.), possibility of contextual display of different pages, possibility of downloading high-resolution images and plenty more.

The presence or absence of a navigable summary or, better still, the structure of the document with the indication of pages or illustrations, can make it easier or harder to consult the reproduction, especially in the case of books consisting of hundreds of pages. The same applies to those texts that, having been submitted to the OCR, allow to identify the occurrence of a term or part of it within a volume. This functionality, for example, is not reflected in the paper equivalent and is configured as one of the characteristics of digital objects with the greatest impact on the public. In all cases in which more digital resources are available from scanning the same analogue equivalent, it would therefore be very useful to also provide a description of the services available in the different viewers or on the platforms that host these objects. Considering the question from a diachronic point of view, the description of these services is as crucial as it is subject to obsolescence: interfaces, functionalities and software change in accordance to the available technology and accounting for these developments is definitely complex, especially if the updating work is carried out with conventional procedures.



Fig. 4. Busch, Frank. August Graf von Platen-Thomas Mann: Zeichen u. Gefühle. München: Fink, 1987. The digital reproduction of the volume, carried out within the Digi20 (“Digitalisierung der DFG-Sondersammelgebiete”) Project can be accessed at the URL: < https://digi20.digitale-sammlungen.de/de/fs1/object/display/bsb00042052_00001.html?leftTab=PER ent>. The digital processing has allowed the provision of separate access points for different types of data (names of people, places, references to relevant documents), as well as full-text search.

Going back to the description of the digital resource, it is clear that it should be disclosed to readers not only which exact item was digitised, but also whether an identical digital reproduction is available in multiple versions with different features on different platforms. Describing a digital object through its own characteristics, therefore not as a simple substitute for the analog object, would give the opportunity to generate appropriate filters, but also to create more meaningful links between paper and digital resources. However, these are choices that favour the paper resource and that relegate the digital one to a condition of subordination. To make a comparison, it would be like informing of the existence of an anastatic reprint in a note of the description of the ancient book that it reproduces. Both the anastatic reprint and the digitised reproduction “represent” an existing resource supported by a different medium: coated paper instead of parchment and pixels instead of paper.

Livello bibliografico	Monografia
Tipo documento	Testo
Autore principale	Accademia della Crusca
Titolo	Vocabolario degli Accademici della Crusca, con tre indici delle voci, locuzioni, e prouerbi latini, e greci, posti per entro l'opera. Con priuilegio del sommo pontefice, del re cattolico, della serenissima Repubblica di Venezia, e degli altri principi, e potentati d'Italia, e fuor d'Italia, della maestà cesarea, del re cristianissimo, e del sereniss. arciduca Alberto
Pubblicazione	In Venezia : appresso Giouanni Alberti, 1612 (In Venezia : appresso Giouanni Alberti, 1612)
Descrizione fisica	[28], 960, [104] p. ; 2°
Note generali	- Altro colophon a carta 4L4v: In Venezia : appresso Giouanni Alberti, 1611 - Segnatura: a ⁸ b ⁸ A-4L ⁸ a-h ⁸ f ⁸ , frontespizio con calcografia raffigurante l'impresa dell'Accademia della Crusca; testo disposto in colonne.
Impronta	- a,u- ilne dio- cali (3) 1612 (R)
Nomi	- [Autore] Accademia della Crusca - [Editore] Alberti, Giovanni
Luogo normalizzato	IT Venezia
Lingua di pubblicazione	ITALIANO
Paese di pubblicazione	ITALIA
Codice identificativo	IT\ICCU\PUVE\002958
<input type="checkbox"/> FI0098	CFICF Biblioteca nazionale centrale - Firenze - FI - [consistenza] 2 esemplari - [tipo di digitalizzazione] parziale - copia digitalizzata
<input type="checkbox"/> RM0267	BVECR Biblioteca nazionale centrale - Roma - RM - [consistenza] 1 esemplare - [tipo di digitalizzazione] integrale - copia digitalizzata
RM0521	IEITR Biblioteca dell'Istituto della enciclopedia italiana Giovanni Treccani - Roma - RM - Disponibilità temporaneamente limitata; informazioni sul sito della biblioteca - [consistenza] 1 esemplare - [tipo di digitalizzazione] integrale - copia digitalizzata

Livello bibliografico	Monografia
Tipo documento	Testo
Autore principale	Accademia della Crusca
Titolo	Vocabolario degli Accademici della Crusca : riproduzione anastatica della prima edizione Venezia 1612 / promossa dall'Accademia della Crusca in collaborazione con Era Edizioni
Edizione	Rist. anast
Pubblicazione	Firenze : [Accademia della Crusca] ; Varese : Era, 2008
Descrizione fisica	[30], 960, [104] p. ; 35 cm + 1 volume + 1 Cd-Rom
Note generali	- Riprod. facs. dell'ed.: In Venezia : appresso Giouanni Alberti, 1612 - In custodia - Ed. speciale f.c. per Ente Cassa di risparmio di Firenze - Edizione numerata.
Comprende	- Una lingua, una civiltà, il Vocabolario
Nomi	- Accademia della Crusca
Classificazione Dewey	- 453 (21.) LINGUA ITALIANA. DIZIONARI
Lingua di pubblicazione	ITALIANO
Paese di pubblicazione	ITALIA
Codice identificativo	IT\ICCU\LUA\0531190

Fig. 5. a) Record of the 1612 edition of the “Vocabolario della Crusca”, taken from the Opac of SBN. The holding records of three scanned copies are shown below. These reproductions, carried out within distinct scanning campaigns, show different features (the digital reproduction of the National Library in Florence includes only four pages; the copy from the National Library in Rome was entirely scanned within Google Books project and can be downloaded in both PDF and ePUB formats; the digital reproduction of the copy belonging to the Istituto dell’Enciclopedia Treccani is made of just 16 pages taken from various parts of the volume, despite the reproduction is declared “complete”); b) Record of the anastatic reproduction of the same edition, also taken from the Opac of SBN

If, in the future, the number of digitisations increases and if this is to be adequately and independently represented in the catalogue, the data contained in the hosting digital libraries, whenever available, should be used to create autonomous descriptive data, which then should be properly connected to the original resource.

In fact, in most cases, it can be assumed that the user is not interested in examining the reproduction of a specific copy, but rather the edition, so he or she would be happy to consult the digital equivalent of any specimen. In many other cases, however, his/her interest is directed to the text, the content, regardless of the specific edition. To satisfy this large percentage of research, it would be sufficient to point out, that is, to describe, within the catalogue, the existence of a text or a translation of it in one of the major projects offering works that are now outside the copyright, like Project Gutenberg or LiberLiber. And this brings us to the other category of resources that tends to be underrepresented in catalogues, digital editions.

The image displays two screenshots of digital editions of Dante Alighieri's *Commedia*. The top screenshot is from the 'Biblioteca italiana' website, showing a metadata page for the 'Commedia' edition. The page includes fields for 'Autore' (Alighieri, Dante), 'Genere' (Poesia), 'Pubblicazione' (Roma: Biblioteca Italiana, 2003), and 'Periodo' (300). It also provides a 'Descrizione fonte cartacea' and 'Descrizione versione digitale' (Dimensione: 1073657 bytes). The bottom screenshot shows a digital edition interface with a row of file format options: ePub, HTML, HTML+ZIP, PDF, RTF+ZIP, TXT+ZIP, and audiobook, all marked as 'GRATIS'. Below this, there is a detailed metadata section including the title 'La Divina Commedia', author 'Dante Alighieri', and publication information: 'Edizione Petrocchi', 'Divina Commedia (La)', 'Edizione Nazionale a cura della Società Danteasca Italiana. A cura di Giorgio Petrocchi.', 'opera di riferimento: "Le opere di Dante Alighieri"; Edizione Nazionale a cura della Società Danteasca Italiana. Comprende: "La Commedia secondo l'antica vulgata" di Dante Alighieri, a cura di Giorgio Petrocchi, 3 volumi. A. Mondadori Editore, Milano, 1966-67.', and 'licenza: Creative Commons "Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale"'. Additional metadata includes 'data pubblicazione: 20/06/2005', 'opera elenco: D', 'ISBN: 9788890359729', 'soggetto BISAC: FICTION / Classici', 'affidabilità standard', 'impaginazione: Leonardo Costa Coscarelli, lcosta@planetarium.com.br; Vittorio Dell'Alto; William I. Johnston, wj@world.std.com', 'impaginazione ePub: Marco Calvo, http://www.marcovalvo.it/', 'pubblicazione: Marco Calvo, http://www.marcovalvo.it/', and 'revisione: Marco Calvo, http://www.marcovalvo.it/; Marina De Stasio, marina_de_stasio@rcm.inet.it; Claudio Paganelli, paganelli@rcm.inet.it; Catia Righi, catia.righi@rsores.it'.

Fig. 6. Two digital editions of Dante Alighieri's *Commedia*. The first (top) is taken from Biblioteca Italiana, a project aimed at the publication of texts for study purposes; the second (bottom) is taken from LiberLiber, a project aimed at the creation of a public library, which fact explains the variety of formats

This certainly meritorious activity is currently not carried out by Italian libraries, with some rare exceptions. At the dawn of the internet, numerous projects concerning the description of web resources (mainly important and authoritative sites) were started, and then abandoned for the poor sustainability, for the difficulty of making the selection and for the instability of the *urls*. The reasons not to indicate the existence on the web of digital texts no longer subject to copyright and freely available on the web may be different. The first is that they are still recoverable through the search engines, motivation certainly correct, but that does not consider how significantly more convenient it would be to be able to find such information in the context that is most appropriate to each individual. It is in the catalogue, in fact, that one can legitimately think of finding books and texts and if access is the immediate one guaranteed by online availability, even better. Failure to report may also be due to the fact that they are not perceived as library resources and that it is therefore not appropriate to devote valuable time to their cataloguing, also in view of the fact that it is impossible to guarantee over time the quality of a web resource, as well as its very existence. In principle, both arguments are correct, even if some digital libraries of texts are now projects of such importance as to guarantee quality and persistence in themselves.

```
<TEI.2 TEIform="TEI.2">
  <teiHeader>
    <fileDesc>
      <titleStm>
        <title>Commedia</title>
        <author>Dante Alighieri</author>
      </titleStm>
      <extent>711 Kb in UTF-8</extent>
      <publicationStm>
        <publisher>Biblioteca Italiana</publisher>
        <pubPlace>Roma</pubPlace>
        <date>2003</date>
        <idno>hbit00019</idno>
      </publicationStm>
      <availability>
        <p>Questa risorsa digitale è liberamente accessibile per uso personale o scientifico. Ogni uso commerciale è vietato</p>
      </availability>
      <seriesStm>
        <title>Collezione BibIt</title>
      </seriesStm>
      <sourceDesc>
        <bibl>
          <title>Le opere</title>
          <title type="part">La Commedia secondo l'antica vulgata</title>
          <author>Alighieri, Dante</author>
          <editor id="ed">Societa dantesca italiana</editor>
          <editor id="ed2">Petrocchi, Giorgio</editor>
          <publisher>Mondadori ; [poi] Le Lettere</publisher>
          <pubPlace>Milano ; [poi] Firenze</pubPlace>
          <date>1994</date>
          <note>Edizione nazionale</note>
        </bibl>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <samplingDecl>
        <p>Tutti i materiali paratestuali della fonte cartacea non riconducibili alla responsabilità dell'autore dell'opera sono stati soppressi nella versione digitale</p>
      </samplingDecl>
      <editorialDecl>
        <correction method="silent" status="medium">
          <p>livello medio: controllo a video con collazione con edizione di riferimento</p>
        </correction>
        <quotation form="data" marks="all">
          <p>I simboli di citazione e di discorso diretto presenti sulla fonte cartacea sono stati rappresentati sulla versione digitale</p>
        </quotation>
        <hyphenation eol="none">
          <p>I trattini di sillabazione a fine riga sono stati soppressi e le parole ricomposte</p>
        </hyphenation>
      </editorialDecl>
  </teiHeader>
```

Fig. 7. The *teiHeader* of the digital edition of a work contains information that should be made clear in the record of the digital edition itself. In the example, taken from the edition of Dante's Comedy published in Biblioteca Italiana, information is provided on the differences between the paper edition used as the source and the digital edition

As far as e-books and digital publishing are concerned, we should consider the fact that, irrespective of the lack of funds, the sense of national bibliography has disappeared and therefore the preservation for the future consultation of the literary and artistic production of the country is no longer protected.

Other resources missing

There is a third category of resources which, apart from a few exceptions, are non-existent in catalogues (in particular in large catalogues): these are online resources which libraries acquire not indefinitely, as the single e-book purchased from the publisher's website, but through annual subscriptions. These are the tens of thousands of databases, electronic periodicals and e-books on which libraries now invest the largest part of their budget. At national level, finding out who has access to a database requires knowledge of the Italian library landscape and, in some cases, a good network of acquaintances working in the field.

There is, in fact, no national catalogue of these resources and those who carry out research without obtaining results could reasonably assume the resource in question is not available in any library. There are many reasons for this state of affairs. First, access to these resources is, in almost all cases, limited to users of the purchasing institution through IP recognition or user through ID and password. One might ask, therefore, what is the point of signalling the possession of a resource that is then inaccessible to most.

Again, to make an irreverent comparison, the same could be said for some ancient or rare books, whose consultation is restricted to a very limited number of experts and scholars. Why describing them in a catalogue open to all if only a few have actually access to them?

A second reason is the volatility of the possession of these resources: in many cases the subscriptions are of annual duration and there is always the risk, for budget cuts or in consideration of the scarce use of a resource, that the subscription is not renewed. All the more so for those e-books, we sometimes speak of tens of thousands of titles, which are purchased in packages pre-established by suppliers. The content of these packages changes from year to year, thanks to policies that allow libraries to select the most popular titles to make them part of the library collection. In order to give appropriate cataloguing relief to these titles, it is unthinkable to proceed to the exemplary description for copy. Instead, it is necessary to obtain from the supplier the corresponding descriptive data, and then upload them massively in the cataloguing database. This activity, however, requires a verification of the quality of the data of the authorities present, to ensure that the syndetic structure of the catalogue is preserved, but also a certain timing. The data must be loaded and replaced within a tight time frame compared to the actual availability of the package, otherwise the operation will be useless.

Of course, the accessibility clause is also valid for electronic books for those who have a link with the institution or institutions that own them. Apart from this, electronic periodicals, which are often described at the cataloguing level only in ad hoc portals, such as ACNP, deserve a special mention, while the consistency and availability of an online version are reported only in some cases, in 'traditional' catalogues.

The resources mentioned above can be considered as the digital equivalents of classes of materials that have long been part of the libraries' assets and for which established descriptive standards already exist. There are other types of resources that deserve to be equally taken into consideration on the basis of the importance that their description can play with respect to their visibility, availability and preservation. However, since they are not yet among materials commonly treated by libraries, shared cataloguing criteria are either missing or not yet widely adopted.

The first consists of the products of the so-called digital humanities, a field of studies developed

in recent years and based on an interdisciplinary approach to research in the humanities and the dissemination of cultural content. Critical editions of texts through computer languages, data visualization, computational linguistics, virtual environments and digital storytelling are just some of the many opportunities of applying computer science to humanities, for example through artificial intelligence techniques such as machine learning to analyze big data and text mining to extract information content from textual content, or semantic web technologies, aimed at improving the understanding of what is asked to the search engine, through associations between information and data.

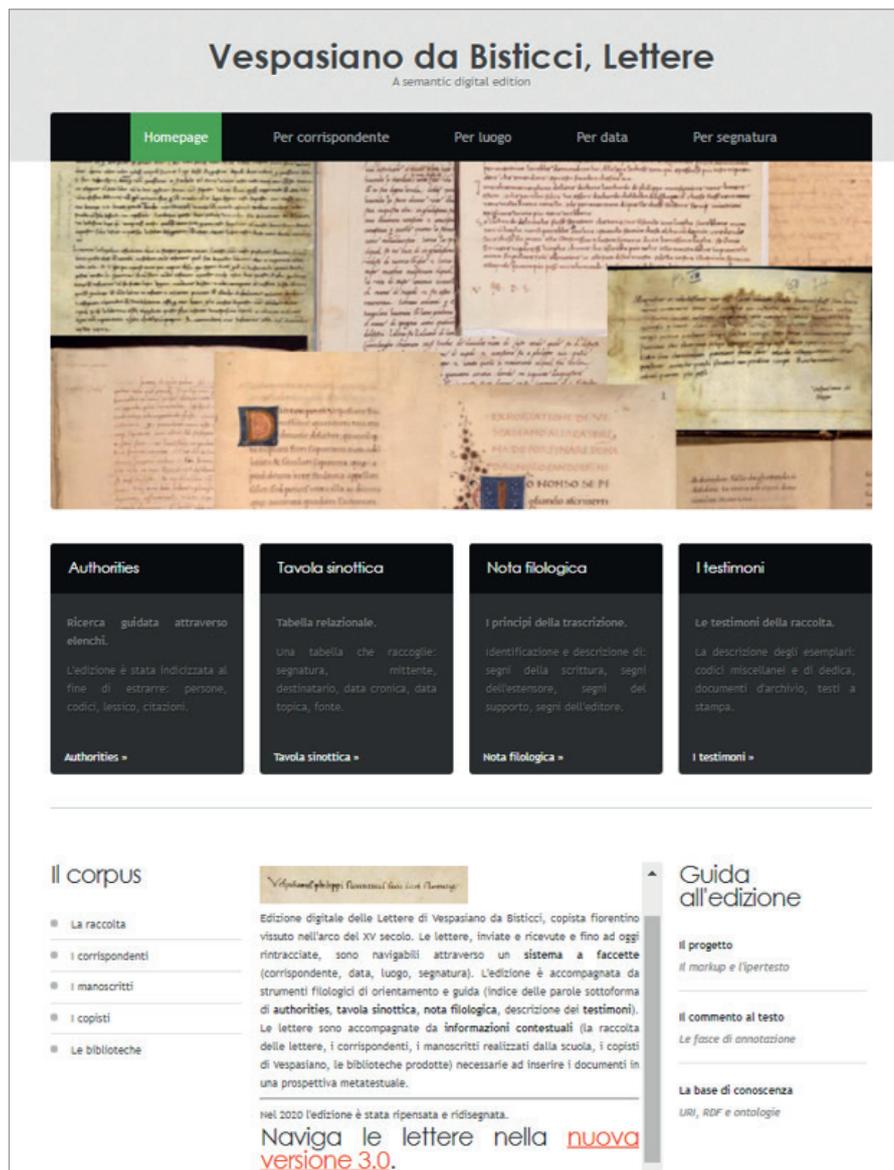


Fig. 8. The digital edition of a work, carried out as part of a Digital Humanities project, represents, for the purposes of the study, even with all the specific characteristics of the digital application, the equivalent of one or more critical essays and as such should be treated in the context of cataloguing to facilitate its knowledge and access (Source: Francesca Tomasi. *Vespasiano da Bisticci, Lettere. A semantic digital edition*. University of Bologna Centro di Risorse per la Ricerca – Multimedia, 2013. <http://vespasianodabisticciletters.unibo.it/#>)

The potential of this area is wide: it allows both to discover new fields of investigation hitherto unexplored, and to expand the public potential of users of humanities through digital technologies, now the main means of production and distribution of knowledge in our society. But for this to happen in a profitable way it is necessary that these achievements, which more and more often constitute the final product of research projects variously financed, are given the same attention as to printed publications. Thus, a number of requirements must be met: the use of open access tools, as well as the adoption of the metadata sets necessary to ensure indexation in cataloguing systems, maintenance and re-use in the later stages of research and storage in long-term repositories

A particular application of digital technology to the publication of reproductions of books, manuscripts and other materials is the International Image Interoperability Framework (IIIF) standard, the purposes of which are described in the following manner:

“The IIIF is driven by a community of research, national and state libraries, museums, companies and image repositories committed to providing access to high quality image resources by defining application programming interfaces that provide a standardised method of describing and delivering images over the web, as well as “presentation based metadata” about structured sequences of images. The standard aims to cultivate shared technologies for both client and server to enable interoperability across repositories, and to foster cooperation among scholars.”⁴

For its specific features, the availability of digital reproductions implementing the IIIF should be made known to readers when describing the digital version of a work.

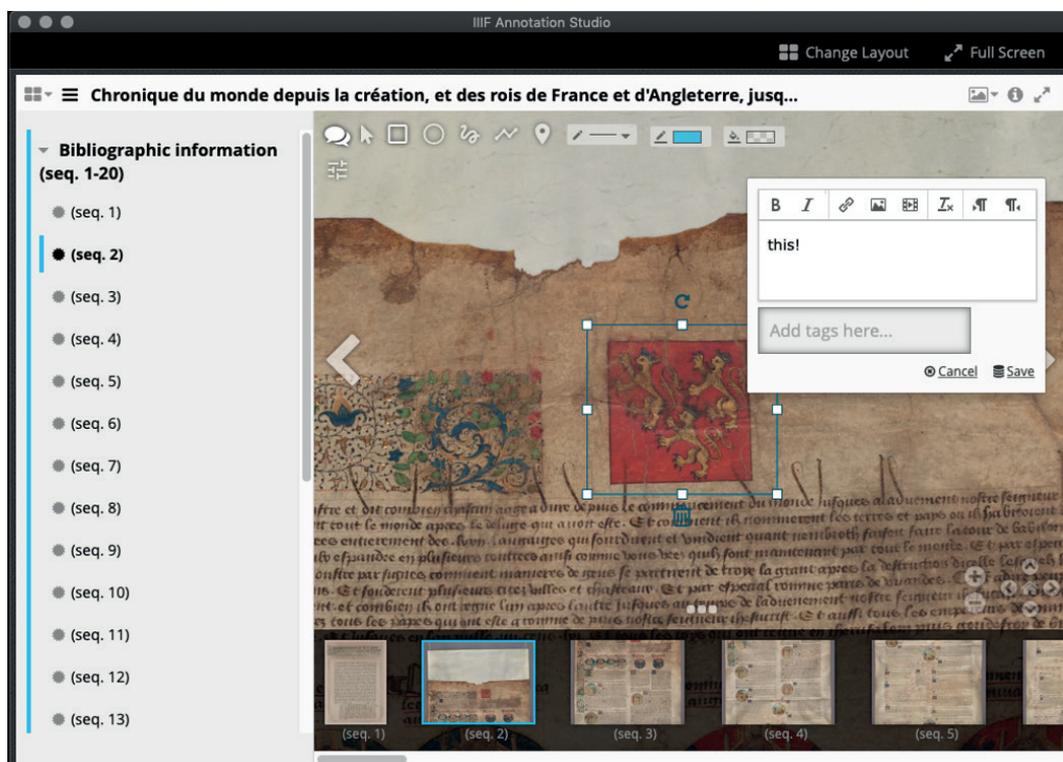


Fig. 9. Users can comment on, transcribe, and draw on image-based resources

⁴ International Image Interoperability Framework. Enabling Richer Access to the World's Images. <https://iiif.io/>.

Other types of resources that should probably be carefully considered are those that have had a considerable boost due to the pandemic. They include, in the first place, the resources created to enable schools and universities to have at their disposal materials useful in supporting teaching and research. Massive open online courses, generally known as MOOCs, online courses aimed at unlimited participation and open access via the Web, were first introduced as early as 2008, and have become over the years a widely researched development in distance education. Aiming at providing open-access features to create virtual environments in which community interactions among students and educators are fostered and supported, MOOCs promoted the reuse and re-mixing of resources such as filmed lectures, readings, data and problems sets. Stemming from this experience, schools and universities have since developed a huge amount of learning objects, “digital self-contained and reusable entities, with a clear educational purpose, with at least three internal and editable components: content, learning activities and elements of context” (Chiappe Laverde, Segovia Cifuentes and Rincón Rodríguez 2007, 8), which require a great deal of investments both in terms of creation and training. To avoid this wealth being dispersed, it is necessary to facilitate their identification, storage and retrieval, through an external information structure consisting of metadata.

Furthermore, universities and professional associations have made extensive use of synchronous and asynchronous streaming sessions, to provide lifelong learning, professional refresher courses and presentations of new products and services. These initiatives too deserve to be preserved, organised, described and made available to the public for future occasions.

The other type of digital resource to consider are podcasts and virtual cultural exhibitions, which in the recent period of the pandemic have enjoyed a large diffusion. Thanks to these initiatives the relationship between people and places of culture and socialization has not weakened, but in some cases has even grown. Libraries, archives, theatres, musical foundations, and opera houses have created a considerable number of products, sometimes revealing a great deal of imagination. Considerable human and professional efforts were required to make all this happen. It would certainly be detrimental not to commit ourselves to preserving and making available this wealth of resources in the future, not just to witness a dramatic event, but as cultural, educational, entertainment, or tourist information materials.

The experience and skills acquired in this last period, together with an interdisciplinary approach commonly referred to as GLAM, and the intersections between the publication formats of the wide variety of classes of digital objects treated, lead to reflect on the way in which digitisation products, the digital libraries, can be reshaped to facilitate their use by users.

It is no coincidence that in the context of the Neustart Kultur Programme, worth almost 1 billion euro, launched by the Federal Government of Germany with the aim of preserving the cultural scene and the cultural infrastructure in the long term, part of the funding has been committed to the programme User-Oriented restructuring of the Deutsche Digitale Bibliothek (DDB). On the assumption that the digitisation projects to which more than five hundred institutions have taken part now give access to 35 million cultural objects in the DDB, 11 million of which are available in digital format, the current programme is aimed at “providing constant free public access to German cultural heritage in digital format still more efficiently. The books, archival materials, photographs, sculptures, paintings, musical works, audio files, films and printed music – in short, the objects – will therefore be linked in such a way that all users of this digital cultural heritage

will be able to explore it using low-barrier search functions and access it in a user-friendly manner. Cultural education using needs-oriented formats will play a key role in this context. Editorially created content containing participative elements will translate DDB objects and collections into narratives, while collections will be contextualised and presented in formats that can be experienced. The outcome will be a range of services that are easily received and used and that promote interactive participation and orientation amid the diversity of the collections accessible through the DDB.”⁵

Changing the paradigm

The presence of digitisation in the catalogues is not a simple cataloguing issue, but it concerns a broader theme, namely the relationship between the world of libraries and the ‘outer world’, or, to put it another way, the positioning of our activities as librarians. In this perspective, a crucial issue concerns policies regarding the inclusion of resources – other than those owned by the institution –, that are freely available on the web and which ought to be described because of their potential usefulness to library users.

The dilemma is linked to the idea of the Library as an institution, its mission, the role and functions it must perform towards users. And when we talk about functions we do not refer only to clarify, and possibly redefine, the relationship and the services that connect the categories of users that each type of library is called to serve. What needs to be identified and possibly reconfirmed is the role – civil, cultural, recreational, social – which the library carries out in the human context, and which justifies its very existence; a role which should give substance to what David Lankester means by stating that libraries are ‘conversations’, participatory realities capable of improving our societies.

In this perspective, the aim is to make the library the place to look for works by using its sophisticated search tools, and to save the user the effort to endlessly repeat the search in the chaotic world of the web. Where, then, is the boundary between the library’s cataloguing needs, which can be exhausted through now traditional rules and practices, and the possibility of exchanging data with the world of the semantic web at the cost of modifying its structure and also its logic?

The complex story of the development of conceptual functional models is aimed, on the one hand, to make the best use of the architectures of the databases and the way in which software programmes treat and structure the data, and, on the other hand, to allow users to comfortably interact with the catalogue. IFLA LRM, approved at the 2017 IFLA Conference and published shortly after, aims to harmonize, within a new modeling that presents higher abstraction levels, the functional models of the FR family (FRBR, FRAD and FRSD), to serve as a theoretical reference for metadatation standards, such as, for example, RDA.

For the fact of having been thought of as a versatile tool ‘usable’ in the semantic web, and consequently based on shared principles and models, independent of the technology used and applicable to any type of medium and resource in any type of cultural institution, RDA raises a number

⁵ Nutzerorientierte Neustrukturierung der Deutschen Digitalen Bibliothek, https://www.dnb.de/DE/Professionell/ProjekteKooperationen/Projekte/NeustartKultur/neustartKultur_node.html.

of issues. In addressing the opportunities offered by RDA, Sardo (2017, 219-225) argues that we are faced with a first step in the direction of a new way of conceiving the activities of cataloguing and of building catalogues that, to deploy its effectiveness, requires the overcoming of a series of significant challenges. First of all, there is the rethinking of the cataloguing data and their organization, which has not yet completely taken place, also because of the huge amount of cataloguing data encoded in ways that are not suitable for the semantic web reality and that cannot always be recoded with automated procedures.

“I wanted librarianship to wake up to the fact that our functional standard was no longer serving us like it should”. This was the point that Tennant (2017) intended to make when, in October 2002, he declared in *Library Journal* that “MARC must die”. “I wasn’t calling for catalogers to go away. I just wanted something better to work with”. That MARC was standing in the libraries’ way more than helping them to survive was not that obvious at the time, nor was it an assertion that would pass unnoticed. Fifteen years later, adds Tennant, “no one seems to think it’s controversial anymore. The Library of Congress has not only admitted that MARC’s days are indeed numbered, they are actively working to develop a linked data replacement. I don’t by any means think that we are out of the woods of making this transition yet, and I also believe it will take many years”.

The process is, indeed, long and painstaking. The term ‘metadata’ is now currently used in literature in place of ‘cataloguing records’, and ‘metadata management’ has replaced ‘cataloguing’ in referring to a much wider application context, far beyond the customary library assets. A report produced by Karen Smith-Yoshimura (2020) sheds light on the results of six years of research and discussions within the OCLC Research Library Partners Metadata Managers Focus Group aiming at clarifying changes in metadata due to the awareness that the time of the bibliographic records hosted in silos is rapidly ending, both conceptually and technically. Meanwhile, innovations in librarianship are putting pressure on metadata management practices to move on as the variety of resources for which metadata sets are required is rapidly growing and libraries are even more involved in cross-sectoral projects, both nationally and internationally. The objective to be achieved is plainly summarised in a document produced by the British Library (2019, 2): “Our vision is that by 2023 the Library’s collection metadata assets will be unified on a single, sustainable, standard-based infrastructure offering improved options for access, collaboration and open reuse”. Expected outcomes of this ambitious process are defined as follows:

- “The complexity of the Library’s collection metadata infrastructure will be reduced by convergence on an agreed set of supported standards and systems
- The unified collection metadata infrastructure will offer new access and processing options enabling a greatly improved user experience of Library services
- Efficient, sustainable collection metadata workflows will match the increasing scale and complexity of collection content via implementation of new techniques for record creation and exploitation of external data source”. (British Library 2019, 8).

Smith-Yoshimura’s report projects these objectives on a much wider scale, to be carried out in countries with very different traditions, organisations, systems of creation and management of data. The question to be addressed as the common starting point of such discussions is: “How do we make the transition to the Next Generation of Metadata happen at the right scale and in a sustainable manner, building an interconnected ecosystem, not a garden of silos?” (Werf 2021).

If collaboration, agreement upon standard outcomes, reuse of data and ontologies are instrumental in reaching the critical mass necessary to create efficiencies and impact and to generate momentum for the picture to change (Dempsey 2019), at the basis of sustainability is knowledge, and therefore professional education of librarians is crucial. Staff fully aware of the potential of linked data and semantic web technologies, totally confident with the data production process, and reassured that no artificial intelligence, no algorithms are going to undermine human intervention in the production of quality data, are key players in a time of transition. According to literature, substantial investment in the professional training of librarians, as opposed to the simple acquisition of the necessary skills for the execution of mechanical procedures, aimed merely at saving time and reducing costs, looks to be a winning strategy in the long run. Guerrini (2020, 13-14) quite correctly explains the reasons:

“Cataloging changes perspective and logic by carrying out metadata creation and management, but remains irreplaceable and maintains the distinctive feature of being an activity primarily cultural and, therefore, technical that reflects the ability to analyse and to represent the resources of the bibliographic universe. [...] The philosophy of the educational approach to cataloguing cannot be characterised by a dogmatic attitude, but, on the contrary, it requires critical sense and recognition of the editorial and historical complexity of the bibliographic object to be described.”⁶

Digital preservation strategies

The question of the relationship between libraries, users and digital resources covers many other aspects that cannot be addressed here, but it certainly cannot be said to be concluded without a reference, however, brief to the issue of preservation, recalling in this regard a thought expressed by Mandillo (2002): “The national collection that is built by law undoubtedly plays a fundamental role in a national policy of freedom of expression and access to information.”⁷

The challenge of ensuring that electronic publications are available for future generations is technically complex and resource intensive in terms of both systems and staff. Digital collecting requires new thinking and new processes, in the first place because digital publications, unlike printed material, can be collected once and made available in multiple locations. This gives libraries the opportunity of sharing, together with the collection, the implementation of other management functions, such as description, storage, preservation and delivery. In an ideal situation, the pooling of human, organisational and infrastructural resources should allow to carry out further collection of digital publications, thus increasing the preserved material ratio.

In highly centralised countries, it was the national library that took on responsibility for preserving digital resources deemed to be of interest for cultural purposes. This situation, however, is

⁶ “La catalogazione cambia prospettiva e logica facendosi metadattazione, ma resta insostituibile e mantiene la caratteristica distintiva di essere un’attività in primis culturale e, quindi, tecnica che rispecchia la capacità di analisi e di rappresentazione delle risorse dell’universo bibliografico. [...] La filosofia dell’approccio formativo alla catalogazione non può essere contraddistinta da uno spirito dogmatico, ma, all’opposto, richiede senso critico e riconoscimento della complessità editoriale e storica dell’oggetto bibliografico da descrivere.”

⁷ “La collezione nazionale che si costruisce per legge gioca indubbiamente un ruolo fondamentale in una politica nazionale di libertà d’espressione e di accesso all’informazione”. On the matter of legal deposit in Italy see (Puglisi 2020).

not very frequent and certainly cannot be the ideal solution in a country such as Italy, where the cultural heritage is dispersed and there are several institutions that have comparable size and history. In addition, almost everywhere there is shortage of staff. Australia⁸ and Germany⁹ have shown the effectiveness of a strategy based on cooperation between institutions characterized by different nature, size, and field of interest. They have undertaken a long process where nothing is improvised, but is the result of the work of various committees and study groups focused on specific issues.

In this respect, the lesson of Luigi Crocetti is as valuable as it usually is: “Preservation without cooperation is still possible; without cooperation it is not possible to make the library a means of communication and information.”¹⁰

⁸ See (Lemon, Blinco and Somes 2020).

⁹ See (Schrimpf and Tunnat 2019).

¹⁰ “Si può conservare senza cooperare; senza cooperare non si può fare della biblioteca uno strumento di comunicazione e d’informazione”.

References¹¹

- British Library. 2019. *Foundations for the Future: The British Library's Collection Metadata Strategy 2019-2023*. London: British Library. <https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2019-2023.pdf>
- Chiappe Laverde, Andrés, Yasbley Segovia Cifuentes, and Helda Yadira Rincón Rodríguez. 2007. "Toward an instructional design model based on learning objects." *Education Technology Research and Development*:671–681. doi:10.1007/s11423-007-9059-0.
- Dempsey, Lorcan. 2019. "What Collaboration Means to Me: Library collaboration is hard; effective collaboration is harder." *Collaborative Librarianship* 10 (4, art. 3):227-233. <https://digitalcommons.du.edu/collaborativelibrarianship/vol10/iss4/3>
- Guerrini, Mauro. 2020. *Dalla catalogazione alla metadattazione. Tracce di un percorso*. Prefazione di Barbara B. Tillett. Postfazione di Giovanni Bergamin. Roma: Associazione italiana biblioteche.
- Lemon, Barbara, Kerry Blinco, and Brendan Somes. 2020. "Building NED: Open Access to Australia's Digital Documentary Heritage" *Publications* 8 (2):19. doi:10.3390/publications8020019.
- Library of Congress. Network development and MARC standards office. 2020. *MARC 21 Format for Bibliographic Data. Update No. 31. 856 – Electronic location and access*. Washington, DC: Library of Congress. < <https://www.loc.gov/marc/bibliographic/bd856.html>>.
- Mandillo, Anna Maria. 2002. "La nuova legge sul deposito legale: una riforma non solo per le biblioteche", *AIB notizie* 14 (3):4-7. <<https://www.aib.it/aib/editoria/n14/02-03mandillo.htm>>.
- Puglisi, Paola. 2020. "Deposito legale quattordici anni dopo: come, quando, 'quanto', e perché" *AIB Studi* 60 (3):591-614. doi:10.2426/aibstudi-12477.
- Sardo, Lucia. 2017. *La catalogazione: storia, tendenze, problemi aperti*. Milano: Editrice bibliografica.
- Schrimpf, Sabine, and Yvonne Tunnat. 2019. "306.2 15 Years of nestor: German Network of Expertise in Digital Preservation (paper Presentation)." OSF. June 20. doi:10.17605/OSF.IO/HA5VN.
- Smith-Yoshimura, Karen. 2020. *Transitioning to the Next Generation of Metadata*. Dublin, OH: OCLC Research. doi:10.25333/rqgd-b343.
- Tennant, Roy. 2017. "'MARC Must Die' 15 Years On" *Hanging together, the OCLC Research blog*, October 15, 2017. < <https://hangingtogether.org/?p=6221>>.
- Trombone, Antonella. 2018. *Principi di catalogazione e rappresentazione delle entità bibliografiche*. Presentazione di Diego Maltese. Roma: Associazione italiana biblioteche.
- Werf, Titia van der. 2021. "Next Generation Metadata... it's getting real!" *Hanging together, the OCLC Research blog*, March 4, 2021. https://hangingtogether.org/?p=8918&utm_campaign=abstracts-6-it&utm_medium=email&utm_source=pardot&utm_content=metadata-opening-plenary-hanging-together-blog-post&utm_term=emea-it-abstracts.

¹¹ Online resources accessed November 11, 2021.