JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12717

JLIS.it

Bibliographic control and institutional repositories: welcome to the jungle

Tessa Piazzini[(a)]

a) Università degli Studi di Firenze, http://orcid.org/0000-0002-8876-371X

Contact: Tessa Piazzini, tessa.piazzini@unifi.it
Received: 3 April 2021; **Accepted:** 23 May 2021; **First Published:** 15 January 2022

ABSTRACT

In 1994 cognitive scientist Stevan Harnad made what he defined a "subversive proposal" to his colleagues: «immediately start serf-archiving their papers on the Internet». Since then, institutional repositories have been chaotically developing, alongside disciplinary repositories. In the early XXI Century the public debate was centered on their purposes and therefore on what they were supposed to contain; librarians joined the discussion and contributed to it by implementing descriptive standards such as Dublin Core and interoperability protocols (OAI-PMH). The themes under discussion were closely related to bibliographic and authority control, given that the quality of metadata has a profound impact on the quality of the services offered to users. Presently, we are still trying to answer some of those old questions: what (or whom) are IRs for? Is bibliographic control so necessary within an environment that has never failed in self-archiving? Can we consider IRs a bibliographic tool? We also need to deal with a wider vision: in a scenario that saw the transition from OPACs (created, managed and controlled by librarians) to current discovery tools (with their information redundancy and the related problems on data correctness and quality control) can librarians still be authoritative and act effectively?

KEYWORDS

Authority control; Institutional repository; Bibliographic control; Metadata.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

1. Introduction

In library sciences we often talk about "ecosystems"[1]. Within this naturalistic metaphor, I used to think of repositories as a jungle: chaotic, dense and impenetrable. Now, however, I look at repositories rather as a rain forest: an equally complex environment full of variety, multi-layered, characterized by a lot of internal and external communication networks and also hidden and visible interdependencies. An environment that rests on a clean, rich soil, on which it is possible to move and walk.

The relatively short history of repositories shows us a great variety in terms of stored material (preprint, research publications, teaching materials, articles and books, theses, multidisciplinary or specialized), in terms of population and organization (self-archiving, batch retrieval, internal collections, librarian mediated insertion) and also in terms of software (Digital Commons, DSpace, Eprints, Fedora…) .

We must not forget that the repositories were born because of the initiative of the scientific community, in particular by the will of single authors; it all started with the "subversive" proposal by Stevan Harnad, professor of cognitive sciences at the Virginia Polytechnic Institute, in 1994: sharing their own research within the institution through the self-archiving of online contributions, in order to make their dissemination more effective. Hence the embryo of a new type of open archive: the institutional archive, promoted and managed by an institution, which goes alongside the disciplinary ones created by aggregation of documents concerning single research areas. We should also remember that "the environment and context in which a repository is situated will unequivocally play a part in the decisions that are made and the quality of metadata that is produced" (Moulaison Sandy and Dykas 2016, 105).

Such repositories, therefore, are fed by the authors themselves through the practice of "self-archiving". It is no coincidence that the elements on which the foundations of repositories rest are the use of minimal metadata sets (such as the Dublin Core Metadata Element Set, adopted by most repository providers) and interoperability standards, such as the OAI-PMH , which go hand in hand with such minimality. According to some scholars, this choice was due to the desire of encouraging the participation of authors in filling the repositories, however it did discourage any activity that could be perceived as a barrier between scholars and their own institutions; this perception has also affected any intervention and quality control measure in the process of metadata creation (Barton, Currier and Hey 2003). Consequently, unlike other more familiar environments, such as OPACs and Discovery Tools, institutional repositories were not created by librarians to organize and make available the bibliographic universe; they are not tools whose birth falls within our sphere, although the contribution of many librarians has undoubtedly been, and still is, vast (for example, in the case of the development of the Dublin Core, despite the "bibliographic control community" seeing it at its birth as "simplistic"[2]).

The awareness that "[for IRs] there is no universally accepted practice or standard defining quality

[1] In this issue many speakers have included this word in the titles of their articles.

[2] In his keynote address to the Library of Congress Bicentennial Conference on Bibliographic Control for the New Millennium, Michael Gorman (2001, xxv) referred to metadata as "a fancy name for an inferior form of cataloguing," and as "unstandardized, uncontrolled, ersatz cataloging." Cited in footnote 6 by Howarth 2005, Metadata and Bibliographic control: soul mates o two solitudes, Cataloging and classification Quarterly, 40 (3-4), 37-56.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

metadata; similarly, there is no set of rules for describing institutional repository materials" (Stein, Applegate and Robbins 2017, 650) marks even more the difference between IRs and actual bibliographic tools.

2. Repositories as rain forests or woodlands?

This awareness leads to a first question: what should we do about these environments? Should we maintain their relatively unorganized nature (with the relative entropy) or should we transform these rain forests into controlled, orderly and organized woodlands, according to our vision of the (bibliographic) world?

Maybe we should be looking for the golden middle way, or the "aurea mediocritas", to quote the Latin poet Horace, who reminded us "est modus in rebus": we should tread carefully and with the respect that is due to an environment with its own characteristics; we should try to intervene most discretely, in order not to deform it, but simply to provide an orientation to its users, so that they not get lost and can appreciate all its beauty.

Therefore, perhaps bibliographic and authority control should also be careful, aiming at providing the users (both humans and machines) with the necessary information in the best possible form, for them to enjoy a pleasant, safe and satisfying journey.

Although born within the academic communities and made their own by the institutions, the repositories, in fact, cater to the widest possible audience, and their main goal is – desired by the authors themselves – not so much to organize and systematize what is produced by an institution (a task that can well be carried out by a catalog), but to ensure the widest visibility for the longest time.

Long-term access and storage can be achieved through two closely linked elements: good metadata and a good repository system; in fact "quality metadata may be underutilized due to weakness in indexing, navigation, and display options"(Moulaison Sandy and Dykas 2016, 103).

Defining what is meant by "quality metadata" is not, however, an easy task in itself: the subjective and local elements remain strong, and the close link between functional requirements and suitability for purpose is often emphasized (Powell, Day and Guy 2004) "by defining both the internal requirements related to the needs of end-users in a local setting and by defining external requirements related to disclose and expose local metadata relating to external service providers" (Park 2009, 214).

At the same time, therefore, it is necessary to guarantee interoperability within a Search Engine Optimization framework, and that implies a much wider series of technical activities and operational choices that should be part of the development plan for a repository.

As librarians we have the task of providing support for the creation of quality metadata, as the NISO (National Information Standards Organization) reminds us in its *A Framework of Guidance for Building Good Digital Collections* (NISO 2007, 61-62), when it says that Good Metadata:

1. Conforms to community standards in a way that is appropriate to the materials in the collection, users of the collection, and current and potential future uses of the collection.
2. Supports interoperability.
3. Uses authority control and content standards to describe objects and collocate related objects.
4. Includes a clear statement of the conditions and terms of use for the digital objects.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

5. Supports the long-term curation and preservation of objects in collection.
6. Good metadata records are objects themselves and therefore should have the qualities of good objects, including authority, authenticity, archivability, persistence, and unique identification.

When speaking of quality of metadata, we cannot, therefore, neglect a fourth aspect besides those of in accuracy, completeness and consistency, which we always find in the literature (Park 2009): their effectiveness in terms of information retrieval, in particular in relation to indexing by search engines. A 2019 study (Mering 2019) conducted upon the University of Nebraska-Lincoln IR tells us that 57% of the collection was accessed via Google Search and only 17% was accessed directly from within the repository. Already in 2012 Arlitsch and O'Brien reminded us that "digital repositories [...] face a common challenge: having their content found by interested users in a crowded sea of information on the Internet" (Arlitsch and O'Brien 2012, 64). Their research also showed that all the analyzed repositories, regardless of the platform, had a low index ratio making them virtually invisible to Google Scholar.

The ultimate goal, however, remains that of ensuring the widest visibility (Swan and Carr, 2008), which is also achieved by making the repository content externally shareable. "Shareability implies an adherence to internal but also extra-organizational standards and best practices; when every repository uses recommended file types, metadata schemas, and the same controlled vocabularies, information is more easily searched and retrieved across them" (Moulaison Sandy and Dykas 2016, 102).

It is therefore important that "for metadata to be effective, enforcement of standards of quality must take place at the community level" (Bruce and Hillman 2004, 3) and that it is necessary to "establish effective policies for the management of authorities in these types of digital collection through cooperative efforts that will permit the development of corpora of authority entries that will aid the processes of cataloging, metadata creation and information retrieval" (Barrionuevo Almuzara, Alvite Díez and Rodríguez Bravo 2012, 101).

Economics, however, teaches us that effectiveness, in cost-benefit analysis, is always accompanied by efficiency, while our daily experience as librarians continually reminds us that bibliographic control activities are very time-consuming and resource-consuming and also require high professional skills.

For example, a survey (Moulaison Sandy and Dykas 2016), conducted in 2015 among some US repository administrators, indicated time limitations and staff hours and skills levels of staff among the greatest obstacles to the provision of high quality metadata, even for those who had self-assessed the quality of their own metadata repository as above average.

3. Different levels of supports for authors and staff in bibliographic control

Hence the need to make prudent choices on how to distribute efforts and how to use all the tools that can help reduce costs.

Now, in the face of the continuous growth of scientific publications – a result also of increasingly pervasive evaluation activities – especially when the main source is self-archiving, one of the tasks of librarians within IRs is to build clear and fast paths, which simultaneously guarantee ease of

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

use and quality of information both for authors and readers, by working on those metadata that have an impact on access.

The first basic form of support consists in offering authors or editors explanatory notes, instructions, drop-down menus or pop-up windows when filling in the bibliographic form online: in this case we certainly cannot speak of bibliographic control; nevertheless, it is a first step which we cannot ignore. Providing clear information, simple tools and technical support goes along with training and education activities, and they all constitute an indispensable step towards creating a community of users who have awareness of the creation of quality metadata, as to reduce *ex ante* the need for subsequent controls.

At the same, initial, layer of support we find the preliminary definition of best practices and guidelines destined to repository managers and staff: "metadata guidelines seem to be fundamental in ensuring a minimum level of consistency in resource description within a collection and across distributed digital repositories" (Park and Tosaka 2010, 711).

A second level consists of recovering quality data from third-party sources through identifiers.

Here is where another naturalistic (or rather, environmentalist) metaphor comes into play: recycling, intended as the possibility of quickly recovering structured and valid information from external sources.

Currently, when it comes to bibliographic and authority control, scholars seem to agree on two main tools:

- Unique publication identifiers for retrieving verified metadata;
- Unique personal identifiers (ORCID, VIAF, ISNI, LoC Authority Name) to ensure the quality control of the authors.

Both tools currently have some limitations and difficulties, but there seems to be a certain agreement among scholars on their effectiveness.

With an eye to economic sustainability, there is a growing tendency for repositories to offer the possibility to retrieve a lot of information from external databases or directly from the publisher through DOI, Scopus identifier or Web of Science Accession number, ISBN, PMID.

A choice of this kind implies on our part, as librarians, the acceptance of delegating the organization and presentation of data to third parties, but this is not new to us: we have done it with derived cataloging, we do it today in part with discovery tools.

If, on the one hand, this can in principle lead to a certain homogeneity in the presentation of information, reducing the risk of typos and errors deriving from manual entry, on the other hand it only partially reduces the need for bibliographic control. This is first of all because publishers and individual databases each have their own metadata and cataloging rules, and secondly because harvesting activities strictly depend on the interoperability protocols applied and on the mapping between the different sets of metadata (Chapman, Reynolds and Shreeves 2009).

For example: the title of an article in the Pubmed database is always presented in English, even if the article is published in another language. In this form the dc.title field of the Core Set Dublin Core is usually imported, where, instead, there should be the title in the original language.

Furthermore, we know well how partial in their coverage large international databases are when it comes to languages or subject matters, and how there are no specific identifiers for certain kinds of publications such as, for example, the essay within a volume – although many publishers are beginning to equip individual book chapters with their own DOI.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

Even in the case of batch uploads carried out by dedicated staff, the quality of the data is not guaranteed, indeed in some cases it seems to be even lower (Stein, Applegate and Robbins 2017), unless an intense metadata cleanup is planned prior to batch ingestion using tools like Open Refine[3].

4. Authority control as part of bibliographic control

We find a similar problem especially in the management of authors' names, whose ever-changing forms have always been a great challenge for the authority control. It could be due to the will of authors (discontinuous use of the middle name or of abbreviated forms, change of surname after marriage...) or for other reasons (different presentations in various sources, linguistic variants, transliterations, etc.).

In order to overcome this problem, as already noted, a first form of assistance is the auto-completion function that – although not present in every relevant software – can kick in during the insertion phase: although this functionality can contribute to the reduction of variant forms, this does not strengthen the authority control.

At a slightly higher level – so much so that we can speak of authority control at a local level – we find the automatic linking of the author's name to the institution identity management system; nevertheless it is evident that this solution does not fully guarantee the interoperability and shareability and "can be, at best, [only] one part of the repository authority control puzzle" (Downey 2019, 130).

In order to achieve this result, unique identifiers of the names are being implemented within the repositories, by linking to external authority schemes.

The interesting aspect, determined by the evolution towards an Linked Open Data model, is the transition from the concept of "name authority work" to that of "identity management", thanks to the association of registered identifiers: "Identity management won't work the same way as the traditional authority control because identity management emphasizes the process of associating a registered identifier (or a URI) with a single entity and the differentiation of names or headings is only of secondary importance in identity management " (Zhu 2019, 227).

The coexistence, however, of numerous projects for the name authority control with the consequent production of different identifiers constitutes an additional element of noise and can lead to the necessity, once again, to make choices.

In 2019 Moira Downey, a colleague from Duke University, published an analysis (Downey 2019) of three among the major international authority sources – Library of Congress Name Authority Files (LCNAF), Virtual International Authority File (VIAF), and Open Researcher and Contributor Identifier (ORCID) – looking to develop a Linked Data authority control within their institutional repository, given the ability of the mentioned systems to provide author URI's via API.

According to the author, LCNAF and above all VIAF, which has developed a cooperative model for the aggregation of authority data from national and regional sources with an intense activity of clustering, merging and deduplication, constitute "a broader step forward in preparing library data

[3] Problem also reported for the management of name entries by Salo Dorothea, 2009. "Name authority control in Institutional repositories", *Cataloging and classification quarterly*, 47 (3-4), 249-261. doi: 10.1080/01639370902737232

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

for better integration with the broader web" (Ibid., 120), but still rely on traditional mechanisms of participatory cataloging and authority control that have an impact on the creation of identifiers, in particular for authors of articles in journals, who happen to represent the category with the biggest presence within many academic repositories, given the hyperproduction of literature in the fields of medical sciences and STEM (Science, Technology, Engineering and Mathematics). Nevertheless, they guarantee reliability on their persistence, thanks to the professionals involved, with the creation of "record in structured, machine-actionable format that did not require additional resources or inferences to ascertain" (Ibid., 131).

ORCID, on the other hand, operates as a "self claim researcher registry", which seems to delegate the authority control traditionally carried out by libraries directly to researchers, also giving them a certain autonomy of choice on their online identity in the universe of academic communication. A leap of faith by librarians or the recognition that we can no longer be the absolute rulers of the organization and presentation of information?

As often happens, the reality lies somewhere in between: ORCID URI's prove to be a good solution for self-archiving, but "the undifferentiated nature of the current ORCID database system seems unhelpful for bulk remediation of existing repository content or for large scale batch operations" (Ibid., 131).

In particular, we have no certainty about the persistence of this identifier, given that any author could decide to remove their profile at any time. We also have the same problem with local registries, which by nature are closely linked to the duration of the author's presence within the institution.

I believe that offering authors a tool, even if not a perfect one, to present themselves is a courageous and intellectually honest choice that we can support by making its use and implementation as easy as possible within "our" repositories, and continuing to invest in a parallel education and information activity on best practices.

There are now numerous experiences in this sense in the world.

For example, in Italy, in 2015, during the national research quality assessment exercise (VQR 2011-2014), the IRIDE project (Italian Research IDentifier for Evaluation) was launched, aiming to equip the Italian academic community (professors, university researchers and research institutions, doctoral students and post-docs) with a persistent ORCID identifier, by activating the registration procedure directly within the repository of their institution. Most of the Italian university repositories, which use proprietary software, have since then allowed, through a push and pull system, a bidirectional communication with their ORCID account.

Equally interesting is the experience (Svantesson and Steletti 2019) of the European University Institute of Fiesole (one of the hosts of 2021 Florence Conference on Bibliographic Control in the Digital Ecosystem) for the integration of its databases – CADMUS, EUI Central Persons Registry – with ORCID, which is also a solution for the authority control over the names of authors that partially compensates for the absence of a CRIS (Current Research Information System). The choice of using the form in the repository as the preferred name is particularly interesting, reminding us that "the criteria of selecting which of the various IDs to use will depend on the stakeholder. Among the factors to be considered is to select the ID system which attracts the "critical mass" representing one's peers" (Smith-Yoshimura et al. 2014, 9).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

5. The challenge of the semantic control

If there is a certain agreement on pursuing these paths, that is not the case regarding the opportunity to invest time and resources for bibliographic control on the semantic component.
In the context of an institutional repository, in many cases a multidisciplinary one, often fed by the authors themselves, the depth, breadth and variety of disciplines means that the use of subject-controlled terms is possible only at a high level, if we are to maintain homogeneity within the repository itself.
If, on the other hand, we want to respect the heterogeneity and we let the communities self-discipline, we will end up with a repository in which the consistency of semantic metadata will be extremely varied: from their total absence to populating via recognized thesauri, and in between the complexity determined by the use of synonyms, homonyms and grammar, spelling and linguistic variants.
We are again faced with the entropy vs. control dilemma: how far must our intervention as repository managers go? Lubas, speaking of PhD theses and dissertation repositories, argued that any subject indexing intervention by staff should have complemented and not replaced the choice of keywords made by the authors, even if this would have led to an inevitable increase in noise (Lubas 2009). A normalization of the keywords, or their mapping on a pre-existing controlled vocabulary would have, in fact, eliminated the unique perspective with which authors refer to their work (Radio 2014).
An interesting study from 2018 (White, Chen and Liu, 2018) tried to analyze the relationship between the presence of some metadata in the Duke University Law School repository and the number of downloads, to understand the effectiveness of the metadata itself.
The results were surprising: the number of co-authors and the presence and the number of keywords (whether they were free text or derived from controlled vocabularies) had a substantially negative correlation with downloads and were not essential for users to reach the content. On the contrary, the presence and length of the abstract had a significantly positive impact.
This partly contradicts our certainty about the importance of subject indexing and its effectiveness. A certainty already undermined over the years by studies which invited us to abandon their use (Calhoun 2006), even within the catalogs. At the same time, other studies confirmed its efficacy (Gross, Taylor and Joudrey 2015). This polarization has now widened with the adoption of Discovery tools by many libraries, which reproduce a sort of "Google-like" environment – in the name of an alleged desire to make the information retrieval experience as satisfying as possible for the user – without, however, the power of Google's Page Rank. Perhaps the term "jungle" is better suited to define such tools, more than IRs.
In order to ensure a balance, a great help could come from Linked Open Data and the Semantic Web, which, with regard to subject indexing, could make an important contribution to the enrichment of contents and bibliographic control, through a simpler management of multiple languages, better linkability of resources and simpler reuse of authority registries in applications. Furthermore, the "semantic search enables a new set of queries that are based on the power of inference engines and are not possible with traditional keyword based search" (Solomou and Koutsomitropoulos 2015, 66).
While repositories go through an inevitable initial effort of adaptation, the choice between differ-

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

ent technologies depends on the complexity and rigor required by the specific environment (Zhu 2019). As already seen at the beginning for quality metadata, also when choosing semantic web tools the relationship between suitability for the purpose and the peculiarities of the environment to which the tools are applied must be addressed, allowing a possible scalarity of choice.
There are also many interesting experiences in this area: in 2017, at the Central University of Gujarat, India, a prototype (Khumar 2018) was developed, in which they linked the Dbpedia knowledge base to a Dspace-based repository, chosen as a linked dataset for its broad disciplinary coverage, for its automated updating mechanisms and its multilingual information support. Equally interesting is the research conducted by Greek scholars on "a transformation engine which can be used to convert an existing institutional repository installation into a Linked Open Data repository: the data that exist in a DSpace repository can be semantically annotated to serve as a Semantic Web (meta) data repository " (Konstantinou, Spanos, Houssos and Mitrou 2013, 834). And it's not the only research of this kind[4].

6. Conclusion

At the end of this absolutely non-exhaustive overview, the conclusion is that there are more questions than answers, more doubts than certainties.
However, it is clear that librarians will not be able to fail in their task of helping to build reliable, rich and "clean" repositories, while exploiting the potential offered by third parties for the creation of quality metadata and bibliographic control.
There are many roads that are being tried to build quality repositories, in which bibliographic control is effective and functional: some will be dead ends, others will become well-marked paths through which librarians and users will be able enjoy the rich rainforest that institutional repositories represent.
All we need to do is to keep on exploring.

Acknowledgements

I woud like to thank my collegue, Paolo Baldi, for the interesting and useful food for thought and Emiliano Wass for his fundamental help for the translation.

[4] See also H. Fari, S. Khan and MY Javed, "Publishing institutional repositories metadata on the semantic web," *Eighth International Conference on Digital Information Management (ICDIM 2013)*, Islamabad, 2013, 79-84, DOI: https://dx.doi.org/10.1109/ICDIM.2013.6694008 and Robert J. Hilliker, Melanie Wacker and Amy L. Nurnberger 2013. "Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons", *Journal of Library Metadata*, 13(2-3), 80-94, DOI: https://doi.org/10.1080/19386389.2013.826036

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

References

Almuzara Barrionuevo, Leticia, Maria Luisa Díez Alvite, and Blanca Rodríguez Bravo. 2012. "A Study Of Authority Control in Spanish University Repositories." *Knowledge Organization* 39 (2): 95-103. https://doi.org/10.5771/0943-7444-2012-2-95-1

Arlitsch, Kenning, and Patrick S. O'Brien. 2012. "Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar." *Library Hi Tech* 30 (1): 60-81. https://doi.org/10.1108/07378831211213210

Barton, Jane, Sarah Currier, and Jessie M. N. Hey. 2003. "Building Quality Assurance into Metadata Creation: An Analysis based on the Learning Objects and e-Prints Communities of Practice." *International Conference on Dublin Core and Metadata Applications; DC-2003--Seattle Proceedings*. Accessed 22 November 2021. https://dcpapers.dublincore.org/pubs/article/view/732

Bruce, Thomas R., and Diane Hillmann. 2004. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*, eds. Diane I. Hillmann and Elaine L. Westbrooks (Chicago: ALA Editions)

Chapman, John W., David Reynolds, and Sarah A. Shreeves. 2009. "Repository Metadata: Approaches and Challenges." *Cataloging & Classification Quarterly* 47 (3-4): 309-325. https://doi.org/10.1080/01639370902735020

Downey, Moira. 2019. "Assessing Author Identifiers: Preparing for a Linked Data Approach to Name Authority Control in an Institutional Repository Context." *Journal of Library Metadata* 19 (1-2): 117-136. https://doi.org/10.1080/19386389.2019.1590936

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53 (1): 1-39. https://doi.org/10.1080/01639374.2014.917447

Karen, Calhoun. 2006. The Changing Nature of the Catalog and its Integration with Other Discovery Tools: Final report. Library of Congress. Accessed 22 November 2021. http://www.loc.gov/catdir/calhoun-report-final.pdf

Konstantinou, Nikolaos, Dimitrios-Emmanuel Spanos, Nikos Houssos, and Nikolaos Mitrou. 2014. "Exposing scholarly information as Linked Open Data: RDFizing DSpace contents." *The Electronic Library* 32 (6): 834-851. https://doi.org/10.1108/EL-12-2012-0156

Kumar, Vinit. 2018. "A Model for Content Enrichment of Institutional Repositories Using Linked Data." *Journal of Web Librarianship* 12 (1): 46-62. https://doi.org/10.1080/19322909.2017.1392271

Lubas, Rebecca L. 2009. "Defining Best Practices in Electronic Thesis and Dissertation Metadata." *Journal of Library Metadata* 9 (3-4): 252-263. https://doi.org/10.1080/19386380903405165

Mering, Margaret. 2019. "Transforming the Quality of Metadata in Institutional Repositories." *The Serials Librarian* 76 (1-4): 79-82. https://doi.org/10.1080/0361526X.2019.1540270

Moulaison Sandy, Heather, and Felicity Dykas. 2016. "High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants." *Cataloging & Classification Quarterly* 54 (2): 101-116. https://doi.org/10.1080/01639374.2015.1116480

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12717

Niso Framework Working Group. 2007. A Framework of Guidance for Building Good Digital Collections. 3rd edition. National Information Standards Organization (NISO). http://www.niso.org/publications/rp/framework3.pdf

Park, Jung-Ran. 2009. "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art." *Cataloging & Classification Quarterly* 47 (3-4):213-228. https://doi.org/10.1080/01639370902737240

Park, Jung-Ran, and Yuji Tosaka. 2010. "Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms." *Cataloging & Classification Quarterly* 48 (8): 696-715. https://doi.org/10.1080/01639374.2010.508711

Powell, Andy, Michael Day, and Marieke Guy. 2004. "Improving the Quality of Metadata in Eprint Archives." *Ariadne* (38). Accessed 22 November 2021. http://www.ariadne.ac.uk/issue/38/guy/

Radio, Erik. 2014. "Information Continuity: A Temporal Approach to Assessing Metadata and Organizational Quality in an Institutional Repository." In *Metadata and Semantics Research*, edited by Sissi Closs, Rudi Studer, Emmanouel Garoufallou and Miguel-Angel Sicilia, 226-237. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-13674-5_22

Smith-Yoshimura, Karen, Micah Altman, Michael Conlon, Ana Lupe Cristán, Laura Dawson, Joanne Dunham, Thom Hickey, Daniel Hook, Wolfram Horstmann, Andrew MacEwan, Philip Schreur, Laura Smart, Melanie Wacker, Saskia Woutersen, and Oclc Research. 2014. *Registering Researchers in Authority Files*. Accessed 22 November 2021. http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf

Solomou, Georgia, and Dimitrios Koutsomitropoulos. 2015. "Towards an evaluation of semantic searching in digital repositories: a DSpace case-study." *Program* 49 (1): 63-90. https://doi.org/10.1108/PROG-07-2013-0037

Stein, Ayla, Kelly J. Applegate, and Seth Robbins. 2017. "Achieving and Maintaining Metadata Quality: Toward a Sustainable Workflow for the IDEALS Institutional Repository." *Cataloging & Classification Quarterly* 55 (7-8): 644-666. https://doi.org/10.1080/01639374.2017.1358786

Svantesson, Lotta, and Monica Steletti. 2019. "DSpace ORCID integration: name authority control solution at the European University Institute." Presented at the The 14th International Conference on Open Repositories (OR2019), Hamburg, Germany https://doi.org/10.5281/ZENODO.3553926

Swan, Alma, and Leslie Carr. 2008. "Institutions, Their Repositories and the Web." *Serials Review* 34 (1): 31-35. https://doi.org/10.1016/j.serrev.2007.12.006

Thomas, R. Bruce, and Hillmann Diane. 2004. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*. Chicago: ALA editions.

White, H. C., S. Chen, and G. Liu. 2018. "Relationships between metadata application and downloads in an institutional repository of an American law school." *LIBRES* 28 (1): 13-24. https://www.libres-ejournal.info/2608/

Zhu, Lihong. 2019. "The Future of Authority Control: Issues and Trends in the Linked Data Environment." *Journal of Library Metadata* 19 (3-4): 215-238. https://doi.org/10.1080/19386389.2019.1688368

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

JLIS.it

In the mangrove society: a collaborative Legal Deposit management hypothesis for the preservation of and permanent access to the national cultural heritage*

Giuliano Genetasio[a], Elda Merenda[b], Chiara Storti[c]

a) Biblioteca Nazionale Centrale di Roma, http://orcid.org/0000-0002-6764-4850
b) Biblioteca Nazionale Centrale di Roma, https://orcid.org/0000-0002-5727-8690
c) Biblioteca Nazionale Centrale di Firenze

Contact: Giuliano Genetasio, giulianogenetasio@gmail.com; Elda Merenda, elda.merenda@beniculturali.it; Chiara Storti, chiara.storti@beniculturali.it
Received: 14 April 2021; **Accepted:** 23 May 2021; **First Published:** 15 January 2022

ABSTRACT

Legal deposit, regulated by Law no. 106 of 15 April 2004 and Presidential Decree no. 252 of 3 May 2006, requires Italian publishers to deposit a copy of the published material with several libraries. Legal deposit involves long-term preservation and access to information on various media, not least computer networks. While traditional media are well regulated, digital legal deposit rules are barely sketched out. The National Central Library of Florence (BNCF), with the National Central Library of Rome (BNCR) and the Marciana National Library of Venice, created Magazzini digitali: a digital legal deposit project that allows harvesting of doctoral theses and e-journals produced by research institutions, in addition to ebooks and commercial journals. Thanks to a collaboration with Horizons and Giunti, BNCR has started an experimental deposit of ebooks through MLOL. While awaiting the regulation on digital legal deposit, it is urgent to reopen the debate on this issue and make more effective the collaboration between institutions involved in the management of the digital library heritage, so as to establish a coordination structure that will define the scientific guidelines and the appropriate technological and service choices.

KEYWORDS

Legal deposit governance; Digital legal deposit; National archive of Italian publishing production; National Central Library of Rome; National Central Library of Florence; Magazzini digitali.

* We would like to give special thanks to Rosa Maiello, Giovanni Bergamin, and Maurizio Messina who shared with us the experience of over 10 years of Magazzini Digitali, and many fundamental reflections for the drafting of our contribution.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

Legal Deposit and bibliographic control in the digital ecosystem

The National Central Library of Rome (BNCR) and the National Central Library of Florence (BNCF) are responsible for the collection, preservation, and cataloguing of Italian publishing production, under the Regulations for State public libraries (Presidential Decree 5 July 1995, no. 417, art. 1, para. 2). This task is possible primarily because of national laws on Legal Deposit (Law 15 April 2004, no. 106) and previous similar legislation in place in many Italian pre-unitary States. The Regulation on Legal Deposit currently in force (Presidential Decree 3 May 2006, no. 252) establishes that all documents of cultural interest which are intended for public use and produced in whole or in part in Italy must be delivered to the two National Central Libraries, for the creation of a National Archive of Italian publishing production, and to a certain number of libraries in the territory to which the publisher or the person responsible for the publication belongs, for the creation of a Regional Archive of Italian publishing production.[1] Legal Deposit constitutes, therefore, the main acquisition channel of publications for the two National Central Libraries. It is instrumental to the bibliographical control and constitutes its necessary premise. For this reason, the control of publishing compliance carried out by the two National Central Libraries, and still aimed only at publications distributed on traditional media, is of fundamental importance. It is interesting to reflect on why even in the digital ecosystem Legal Deposit continues to play an important role. We are immersed in an endless availability of databases, repertories and bibliographical indexes of different nature and origin, which often allow access, under certain conditions, not only to the bibliographical record but to the full resource. An ecosystem in which crowdsourcing appears to be the only way to truly govern the enormous mass of data and information produced by contemporary society. It is precisely these peculiarities of the digital information ecosystem that make Legal Deposit even more important: to preserve and make permanently accessible the national cultural heritage, a public authority must be identified to govern the entire process of acquiring, managing, and making available documents of any nature and on any medium.

Italian and European legislation

The regulatory framework of Legal Deposit in Italy is completed by the Ministerial Decrees of December no. 28, 2007 and December 10, 2009, which identify the beneficiary institutions of regional Legal Deposit, as well as further agreements, notes, and clarifications by the General Directorate for Libraries and Copyright on specific aspects of Legal Deposit. However, the legislation has several grey areas. First, there is the problem of a lack of coordination among depository institutions. Despite the existence of a Commission for Legal Deposit, which should meet periodically to control and monitor the implementation of the law and to define single issues related to it, and even though the General Directorate for Libraries has clarified some doubtful points, there is no coordinating body among depository institutions for the daily activities related to Legal Deposit, either at the national or regional level. There is also a lack of shared databases to monitor publisher activity and compliance. The regulations currently in force present sever-

[1] For previous legislation see (Alloatti 2008, 25–33).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

al critical points linked to the growing phenomenon of paper self-publications:[2] among them, the problem of identifying the subject obliged to make the Legal Deposit stands out, as well as that of partial exemptions, based on the criterion of circulation (publications printed in less than 200 copies are not to be sent to the BNCR). The circulation criterion is difficult to verify both for conventional publishing and self-publications because the circulation is seldom stated in the publication. Finally, another critical point is the total exemptions, which still do not include, even under certain conditions, the exclusion of self-publications. Another problematic aspect is the number of copies for paper Legal Deposit, as many as four between National and Regional Archives of Italian Publishing Production,[3] reduced to three with Decree-Law April 24, 2014, no. 66, for many regions. Even three copies are too many if they involve the multiplication of identical tasks across multiple libraries. Although the legislation provides that some particular types of documents must be deposited in specific institutions designated to manage them, it does not yet apply to all kinds of materials. A further critical point is the monitoring of the fulfillment of the Legal Deposit (Presidential Decree 3 May 2006, no. 252, art. 41), which the Regulation entrusts to generic control instruments of the depository institutions. BNCF and BNCR have databases that cross-reference OPAC and trade book data based on ISBNs and can allow for sufficiently effective monitoring of publishing compliance. At the regional level, however, there do not appear to be similarly adequate monitoring tools. Regional Legal Deposit is an innovative regulatory element because it introduces a partial decentralization of Legal Deposit, but it is also a critical element because the identification of depository libraries and the criteria for dividing the material to be deposited has been difficult. Regional Legal Deposit has not always been successful.[4] Equally thorny is the question of penalties for failure to comply with the Legal Deposit requirement. In the face of widespread evasion, the system of sanctions set out in the legislation involves lengthy and cumbersome procedures.[5] A comparison of documents received for Legal Deposit in 2019 at BNCR and BNCF leaves no doubt. For BNCR: 43205 monographs, 3410 children's publications, 3517 school texts, 387 sheet music and scores, 175 maps, 2329 audiovisuals (CDs, DVDs, and multimedia); the final tally of minor publications is yet to be made. For BNCF: 64184 monographs, 1790 children's, 3311 scholastic, 852 sheet music and musical scores, 123 maps, 2007 audiovisual, 2971 minor publications. These differences show the degree of evasion in both institutions, although partly due to partial exemptions and different categorization criteria. Furthermore, this penalty system tends to criminalize the publisher and thereby make him an enemy, rather than an ally, of the library, with harmful consequences for both parties. Finally, one of the weakest points of the legislation is that of the digital Legal Deposit (Presidential Decree 3 May 2006, no. 252, art. 37), the definition of which is postponed to a further regulation to be formulated – as yet non-existent but soon to be published. Article 37 of the Regulations refers to "voluntary forms of experimentation" of Digital Legal Deposit, through agreements with publishers. Despite the agreement between the Italian

[2] A survey conducted in 2015 by the AIE (Italian Publishers Association) had shown how even then almost 50% of ebooks published in Italy were self-publishing, for a total of over 25,000 titles per year. See («Quasi un titolo ebook su due è nato con il self publishing. L'indagine Aie: grandi numeri, ma il mercato è meno di un terzo» 2016).

[3] Presidential Decree 3 May 2006, no. 252, art. 1, para. 2 and art. 6.

[4] See (AIB Biblioteche e servizi nazionali 2020).

[5] Law 15 April 2004, no. 106, art. 7; Presidential Decree 3 May 2006, no. 252, art. 11, art. 43, 44, and 45.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

Ministry of Cultural Heritage and publishers, and despite *Magazzini digitali* (see below), the Digital Legal Deposit still lacks a regulatory framework that goes beyond the experimental phase. It is no longer possible considering the increasingly important Digital Legal Deposit as the younger brother of the paper Legal Deposit. Today it would be appropriate to overcome this obsolete vision and, on the contrary, to speak no longer of digital (or paper) Legal Deposit but of Legal Deposit *tout court*, since the two aspects are increasingly intertwined. Together with the enactment of the Regulation on the deposit of documents distributed via computer networks and the partial reformulation of Chapter VII of Presidential Decree 3 May 2006, no. 252, the Italian legislator has another opportunity to provide libraries with the appropriate tools to carry out their tasks: the planned stage of transposition into national law of the EU *Directive 2019/790 on Copyright in the Digital Single Market*,[6] approved in February 2019. The Directive has received different opinions, sometimes conflicting, from professionals and associations of libraries and librarians[7] but, certainly, the way in which it will be transposed may influence in one way or another the ability to preserve digital memory and especially the possibility of creating services on the digital documentation deposited and preserved: think of e-lending, but also of the cataloguing and indexing of deposited resources through text and data mining activities. The sustainability of Legal Deposit and bibliographic control services in the near future will depend, to a large extent, on how national and European legislators will integrate the point of view of memory institutions – libraries, archives, and museums – on preservation services and permanent access to information, within the reference legislation, both the one specific to Legal Deposit and the one related to copyright, privacy, and personal data processing.

Preserving digital memory: the experience of Magazzini digitali and the challenges of today

On December 15, 2020, the digital archive of *La Stampa*, one of the most important and long-lived Italian national newspapers, disappeared from the web.[8] The archive was built in 2010 with Flash, a technology that, even then, though widespread, was considered risky, mainly because it was proprietary to Adobe, and therefore incompatible with the paradigms of the web, whose optimal functioning is guaranteed by the use of open standards. At the end of 2020, the archives of *La Stampa* had to "close for maintenance" until February 15th 2021 when, after a re-engineering of the platform, it is available again. In the meantime, because the digitization of the journal had been released under a Creative Commons (CC-BY-NC-ND-it 2.5) license, a team at the Internet Archive downloaded the entire archive and made it available by creating a special collection.[9] We can safely say that that of *La Stampa* is by no means an isolated case. For different reasons, in 2017,

[6] See https://eur-lex.europa.eu/eli/dir/2019/790/oj.

[7] See («Le raccomandazioni della rete MAB per il recepimento della direttiva europea sul copyright» 2020). See also («Copyright reform (archived) - European Bureau of Library Information and Documentation Associations (EBLIDA)» s.d.).

[8] (Tedeschini-Lalli 2021) and («L'archivio storico de La Stampa sarà di nuovo consultabile entro metà febbraio 2021» 2020).

[9] *Archivio storico La Stampa*, "Internet Archive", https://archive.org/details/la-stampa-newspaper (last accessed March 31, 2021).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

the digital archive of *L'Unità* met the same end.[10] We could cite the cases of thousands, or even millions – in the case of websites[11] – of digital resources of extreme interest that have vanished into thin air.[12] The preservation of digital resources and their accessibility in the long term is a de facto necessity of contemporary society, a responsibility that memory institutions cannot shirk. For this reason, despite the absence of regulations in the field of Legal Deposit of publications diffused through computer networks, the prototype experience of *Magazzini Digitali*,[13] born as such on the basis of Presidential Decree 3 May 2006, no. 252, art. 37, has been consolidated over the years to become a real service. As of December 2020, *Magazzini Digitali* includes:

- over 170,000 doctoral dissertations, through harvesting with OAI-PMH protocol from the repositories of 58 Italian universities;
- about 180 open access journals, through harvesting with OAI-PMH protocol from the repositories of 7 Italian universities and other research organizations or associations;
- about 500 ebooks deposited in BagIt format;
- about 80 TB of high-resolution digital copies from library digitization projects (with Google Books Project as the main source).[14]

Besides, beginning in 2018,[15] BNCF launched a Web archiving service joined by approximately 220 institutions for a total of about 300 sites subject to periodic archiving, ensuring long-term access to digital resources is an organizational, management, and technological challenge, even more than preserving them. In 2020 the Web archiving service made important steps:

- Bibliographic records related to doctoral dissertations and open access e-journals have been indexed in the catalogs (OPAC) of BNCF and BNCR, and archived resources made available by the internal networks of the two Institutes;
- the collections of websites archived by BNCF have been made publicly accessible on the Archive-it platform.

BNCR has an experimental service that focuses precisely on user utilization. BNCR and Horizons Unlimited LLC (the Italian company that creates MLOL – MediaLibraryOnLine, an Italian e-lending platform) reached an agreement in 2019 to start experimenting with free ebook deposit, which was initially joined by Giunti publishers and will be joined by Mondadori publishers during 2021. The pilot project foresaw an innovative infrastructure through a new MLOL Reader Desktop App, which uses for the first time in Italy the Readium LCP DRM, the open-source DRM system developed by EDRLab (the European branch of the Readium Foundation and member of

[10] After a pirated copy of the archive was made available on the deep web for a few months, the archive came back online in October 2018, but it is still unknown to this day who and how made it possible to restore the service. See («Riappare online (con un curioso sottotitolo) l'archivio storico de L'Unità. Mistero sull'autore dell'operazione» 2018). *L'Unità* can be consulted at: *L'Unità*, "Internet Archive", https://archive.org/details/lunita_newspaper (last accessed April 1st, 2021)

[11] See also: https://www.internetlivestats.com/total-number-of-websites/ (last accessed 16/12/2020).

[12] See (Laakso, Matthias, e Jahn 2021)the responsibility rested primarily with librarians, but the shift toward digital publishing and, in particular, the introduction of open access (OA.

[13] (Bergamin e Messina 2010, 144–53).

[14] https://www.bncf.firenze.sbn.it/biblioteca/magazzini-digitali/. See also (Storti 2019).

[15] The BNCF carried out an experimental harvesting session in 2006 in which 7 terabytes of data were collected from websites belonging to the .it domain. These data are available through the Archive-it BNCF Collection: https://archive-it.org/home/BNCF. See also (Bergamin 2012, 170–74).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

the W3C for the maintenance of EPUB)[16]. BNCR immediately welcomed the chance of experimenting with a service for the deposit and use of digital content that allows users to have immediate access to documents, without having to wait for the often very long processing times of paper documents.[17] The possible future developments of accessory search and access services (the full text of the documents, the data mining procedures on the books, the possibility of enrichment of the OPAC dialogue with search engines, etc.) are also worthy of interest.

From deposit as a procedure to deposit as a process: a new design for managing complexity

Giovanni Bergamin summarized in 2006 the different positions following the enactment of Law no. 106 of 15 April 2004, which for the first time extended the obligation of Legal Deposit to documents circulated via computer networks: "impossible, useless, civil."[18] Public opinion has had to reconsider the usefulness of such an operation, recognizing that the preservation of digital memory is a duty for any "civilized" society. Instead, there is a debate about the possibility of carrying out this activity in an effective (concerning the available technologies) and efficient (concerning the amount of digital information currently being produced and growing exponentially) way. In addition to the lack of a clear regulatory framework, this possibility still seems remote because there has been too much focus on technology and too little on process management, as has happened in other areas involved in digital transformation. In the words of Luciano Floridi, who inspired the title of this contribution, "the challenge is not technological innovation but digital governance."[19] This governance cannot be implemented by applying or, worse, bending the paradigms of the pre-digital and pre-web world to the web world, but requires a "new design." What is needed is a model that can manage the entire information flow, i.e., the process that enables long-term preservation and access to information, not just individual digital information or a single repository process. It is also an opportunity to reiterate that there is no longer a clear separation between an analog and a digital information flow, as already mentioned. As Luciano Floridi said, we live in a mangrove society[20] in which there is no longer a solution of continuity between offline and online.[21] The new governance model should consider the management of Legal Deposit regardless of the nature of the objects or the form by which information is recorded and transmitted: […] "The term deposit should not be understood in a literal sense (physically bringing an object to a particular location) but in the context of "Legal Deposit" as an "institution" where operation-

[16] See https://www.edrlab.org/readium-lcp/ and (Rosenblatt 2017).
[17] MLOL acted as an intermediary between the Library and the partner publishers, setting up procedures for collecting data and bibliographic metadata of the ebooks, and packaging them in a special deposit package. The BNCR user who accesses the Desktop MLOL Reader App from local workstations can download locally the ebooks equipped with the new Readium LCP copy protection system.
[18] (Bergamin 2012).
[19] The quotations by Luciano Floridi are taken from the conference "For an ethics of technology" - Cubò, Bologna, February 13, 2020. See also (Floridi 2017, 2020).
[20] (Floridi 2018).
[21] See (Floridi 2015).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

al procedures must take into account the nature of the object".[22] To design this new governance model, it is necessary to attempt to briefly reflect on the main individual aspects that make up the Legal Depository process, and that have been affected by the digital transformation.

New roles and new actors

Under the law, responsibility for managing the institution of Legal Deposit, including long-term preservation and access to digital resources, is essentially shared between the State and the Regions. The State exercises it through the National Central Libraries of Florence and Rome, the Regions through the institutes identified by the Ministerial Decrees of 2007 and 2009. The traditional management of Legal Deposit allowed co-responsible institutions to proceed independently, often implementing practices that were supplementary in means and ends. The advent of digital technology makes this form of management impractical or inadvisable. A digital copy can be stored, cataloged, and indexed only once and still be accessible in geographically distant points. Moreover, the traditional management conflicts with other regulatory provisions, particularly the provisions of AGID (Agenzia per l'Italia Digitale) relating to the rationalization of public information and communication technology assets.[23] Until now, preservation of and access to resources has been almost exclusively the responsibility of depository institutions. Today, however, it is necessary to involve producers, distributors, and those responsible for information and records in any capacity, depository institutions, and organizations.[24] This does not mean that depository institutions lose or delegate part of their tasks, but that they rediscover a new role and actively bring new players into the resource management process. It has recently been estimated that the entire digital universe consists of approximately 44 zettabytes of data.[25] Even limiting the responsibility of memory institutions "to documents [of cultural interest] intended for public use […] produced totally or partially in Italy",[26] the mass of information to be managed is enormous and constantly growing. Unlike traditional media, thinking especially of websites, it is increasingly difficult to establish cultural interest from the origin: libraries will play a leading role precisely in the ability to define what is of cultural interest, a role that is not new to the traditional responsibilities and tasks of a library but unprecedented for the quantity and types of material to be selected. Digital and the web have not only changed a large part of the traditional publishing industry and market

[22] (Bergamin 2012).

[23] «Razionalizzazione del Patrimonio ICT|Agenzia per l'Italia digitale». Accessed April 1st, 2021. https://www.agid.gov.it/it/infrastrutture/razionalizzazione-del-patrimonio-ict.

[24] The need for close collaboration between publishers or producers of information and depository institutions was already evident in 2011, during the phase of definition of the parameters of the experimentation of the Digital Legal Deposit service. On July 14 the General Directorate for Libraries and the most representative associations of the publishing industry signed an agreement "for the promotion of the convention for the legal deposit of digital documents and the license for their use". Although the agreement did not have the expected operational results, it constitutes an important model for the collaborative management of digital legal deposit. See also General Directorate for Libraries website: Direzione generale Biblioteche e diritto d'autore. «Accordo MiBAC - Associazioni Editori». Accessed April 1st, 2021. https://www.librari.beniculturali.it/it/notizie/notizia/4ee4df59-4819-11e1-88f7-b7fd06d12128/.

[25] Tremolada, Luca. «Quanti dati sono generati in un giorno?» Info Data (blog - Il Sole 24 ore), 14 maggio 2019. Accessed April 1st, 2021. https://www.infodata.ilsole24ore.com/2019/05/14/quanti-dati-sono-generati-in-un-giorno/.

[26] Law 15 April, 2004, no. 106, art. 1, para. 1 and 3.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

– think of the explosion of the self-publishing and print-on-demand phenomenon – but has also introduced new media or new ways of using existing content: social media, streaming video, podcast platforms, apps, etc. Depository institutions are thus no longer just conservators but selectors of resources. Digital preservation isn't a process that begins the moment the document enters the library's collections: the archivability[27] of a website and, in full, of a digital document must be an original "property" of the documents. The functions of the depository institutions should be reconsidered: the new task will be to define criteria and guidelines for the archivability of documents, taking into account, the theoretical models of reference and the state of the art of technologies on one hand and the real possibilities for information producers to adapt to these models or to adopt standards on the other hand. Similarly, cataloging and indexing can no longer be the exclusive prerogative of libraries, which activate the service as the resources enter the collections. Document producers or distributors should be involved. However, the definition of ontologies for the descriptive, semantic, and managerial metadata remains the responsibility of the depository institutions. Long-term access largely depends on the correct compilation of metadata that describes the policies for access, use, and reuse of documents:[28] cataloging and ordering of resources remains, even in the digital world, the first form of guarantee of access to these resources. Full-text indexing and refinement of search algorithms help facilitate information retrieval, but the number of documents and information that must be retrieved is, as repeatedly stated, increasing. It is therefore clear that the management of Legal Deposit, as a process with the characteristics previously identified, requires, in the first instance, a reorganization of personnel and workflows within the depository institutions. Legal Deposit is a completely new activity in some respects, while in others it continues some of the established services of libraries, which should, however, be revised in light of the changing ecosystem in which it operates. To provide an effective and efficient national service for the preservation and access of digital resources, aligned with international best practices (and indeed capable of constituting a model in its own right), and in keeping with current modes of cultural production, depository institutions should have a sufficient number of professionals capable of fully managing these activities in an integrated manner, not only from a scientific and technological point of view but also from an administrative and organizational one.

New models for acquiring, storing, and accessing resources

Legal Deposit as an institution becomes more understandable in the resource acquisition phase: the sending by the producer or person responsible for the digital publication is considered as a residual modality for the digital deposit, while automatic or semi-automatic harvesting of computer networks becomes the norm. As for the acquisition of resources, memory institutions should define

[27] "Archivability" refers to the set of characteristics that the content, structure, functionalities, and interfaces of a site should possess for the site to be preserved and made accessible over the long term with current web archiving tools. The concept is however extendable to other digital resources whether they are spread on the web or not. See also Biblioteca Nazionale Centrale di Firenze. «Archiviabilità dei siti web». Accessed April 1st, 2021. https://www.bncf.firenze.sbn.it/biblioteca/archiviabilita-dei-siti-web/.

[28] This system should provide for the possibility of managing any changes in rights over time, both those provided for by law (think of the term after which a work falls into the public domain) and for cases in which special or advance licenses are issued.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

clear models[29] that always take into account the sustainability of procedures for all stakeholders and the characteristics of digital documents. In summary, the resource acquisition model should define what resources to acquire and how to acquire them,[30] what formats and what protocols, what descriptive and management metadata are required, what metadata for preservation must be attributed during acquisition.[31] Access models must take into account copyright regulations, policies, and licenses established by the owners of the information, user profiles, indexing systems, and last but not least, the possibilities of replay systems. Finally, storage models should take into account the security and redundancy of data, their growth forecasts, and the ability to move and migrate data and documents as needed. These procedures should be based on globally shared conceptual models, take current technologies into account but not be strictly linked to them, and not disregard the regulatory and application context, here including not least the characteristics of work in and for the public administration.

A hypothesis of collaborative management of the Legal Deposit for the preservation and permanent access to the national cultural heritage

"Unfortunately, even when libraries recognize that they need fresh perspectives, they all too often turn to the academic and library-oriented technical communities that simply reinforce the same problems, rather than widening their reach to the outside world to bring in entirely new ideas and perspectives [...] In the end, libraries have reached an inflection point where they will continue to fade into irrelevance when it comes to web archiving if they are not dragged kicking and screaming out of the third century BC and into the modern world. Such modernization can only come from reaching outside of their traditional confines and engaging in sustained partnerships with the outside technical community, bringing in fresh perspectives and approaches. Until then, our web history continues to rapidly slip away, lost forever".[32]

Let us try to summarize the main peculiarities of the governance model of the national Legal Deposit service that have emerged so far:

- Digital preservation is not just secure backup,[33] so it is not an activity that can be solved with simple storage services;
- The preservation of digital cultural heritage is the task of memory institutions. It differs

[29] The theoretical reference model for digital preservation remains the OAIS model - Open Archival Information System, ISO 14721 standard. "Models" mean in this context the level of application procedures, therefore not strictly related to technologies and formats in current use. See (Michetti 2008, 32–49). About the latest revision of the OAIS standard see «OAIS: entro breve la revisione». ParER - Polo archivistico dell'Emilia-Romagna. Accessed April 1st 2021. https://poloarchivistico.regione.emilia-romagna.it/news-in-evidenza/oais-entro-breve-la-revisione.

[30] "The mere accumulation of data, while waiting for more powerful computers, more sophisticated software, and new human skills, will not work, not least because we do not possess sufficient storage capacity." (Floridi 2017, 18). "In hyperhistory, saving is the default option. The problem becomes what to delete." (Floridi 2017, 22).

[31] The reference standard for the production of preservation metadata is PREMIS (PREservation Metadata: Implementation Strategies), «PREMIS: Preservation Metadata Maintenance Activity (Library of Congress)». Accessed April 1st 2021. http://www.loc.gov/standards/premis/.

[32] (Leetaru 2021).

[33] (Bergamin 2018).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

from "standard preservation"[34] in terms of both the object of the preservation (documents of cultural interest vs. computerized documents) and the objectives of the service (preserving the products of national culture and scientific research vs. guaranteeing the evidentiary value of the documents) while presenting necessary similarities in terms of technological solutions;
- Long term preservation is above all a service that has to do with democracy: access to information with equal opportunities for all citizens can only be guaranteed by public institutions, and cannot be the exclusive prerogative of private companies, even if they have a public purpose;[35]
- Digital management requires a revision and rationalization of the policies of protection, valorization, and access to cultural heritage.

Italian public institutions should find, also at the regulatory level, the definitive confirmation of their role for the fulfillment of the tasks of Legal Deposit, bibliographic control, conservation, and access to documents, in whatever form they are recorded and transmitted. A simple investiture by law, in itself, does not qualify them to exercise this responsibility. In the absence of economic, instrumental, and human resources, no kind of governance is possible: the preservation of and permanent access to collective memory is a strategic national service, a challenge that can only be met if tackled collaboratively. A viable solution to these problems, one which is in line with current laws, is a public company to manage the services of conservation and permanent access to cultural heritage. This society should be shared by the bodies and institutions responsible for Legal Deposit: the State (Ministry of Cultural Heritage and its institutes), the Regions, the libraries of the Constitutional Bodies, which receive the deposit upon request of official publications of the State and other public bodies, and CNR – National Research Council, which receives the deposit of publications in the technical-scientific area. It would be important to establish a greater synergy with the governing bodies of SBN (Italian National Library Service), concerning the managing of digital resources within the SBN catalogue (i.e., the Italian Union Catalog), and AGID. It would be important to involve public players such as the Istituto Poligrafico e Zecca dello Stato, ISTAT (National Institute of Statistics), CRUI (Conference of Rectors of Italian Universities) as representatives of the world of academic research, but also, to mention materials of a different nature, RAI Teche. All these institutions are major producers of cultural information or public sources and bearers of similar instances and needs, as well as domain know-how. On this last point, partnership with private entities is essential and there are many candidates for this role. One possibility is to consolidate long-standing synergies with publishers or digital content distribution platforms: by way of example, Casalini Libri and MLOL. Then, we should turn our attention to companies that deal with the preservation and management of digital information: together with the aforementioned Internet Archive and the Wikimedia movement, whose institutional mission is similar

[34] Legislative Decree 7 March 2005, n. 82, art. 34 para. 1 bis, and the Decree of the President of the Council of Ministers 3 December 2013, art. 5 para. 3.

[35] The most important player in this field is certainly Internet Archive: https://archive.org/about/. For the same reason, a close collaboration with the Wikimedia Foundation, and with the national chapter Wikimedia Italy, would be highly desirable, especially on the side of the definition of tools and methods of access to preserved documentation.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

to that of Memory Institutes, it is certainly equally important to finding forms of collaboration with the so-called Big Players active in this field, such as Google and Amazon. The constitution of a public company would respond more easily to the need to pool stable resources to make existing services more efficient, and would allow the procurement of resources, especially human and instrumental ones, in a manner more in keeping with the rapid development of technological services. Currently, the costs amounted to approximately € 500,000 per year. They had been calculated based on existing contracts and the experience gained in almost 10 years of Magazzini Digitali, net, however, of the investments already made over the years, in particular by BNCF, and those related to personnel and infrastructure currently of the depository institutions. There are also the costs of the traditional legal deposit. BNCR, for example, invests about € 265,000 per year in external support for the management of bibliographic control, made necessary by the severe shortage of staff that forces the use of external collaborators to carry out the service. These costs are those necessary to guarantee the minimum services for the management of the Legal deposit. The procurement of such substantial resources can no longer be solely tied to specific projects or targeted funding, as was the case in the early stages of testing and implementation of the service. A public company would have a stand-alone budget, established through the co-partnership of the various entities responsible for Legal Deposit,[36] and might be able to provide a share of commercial long-term preservation services to third parties beyond the provisions of Legal Deposit. A company with a public shareholding could recruit personnel more easily by selecting the professional skills not available in the roles of the public administration, or not available in sufficient quantity, albeit always with public evidence procedures. A company with public shareholding would have a flexible corporate architecture, allowing the acquisition of instrumental resources in the technological sphere, without partial or total interruption of services, and loss of know-how. The synergy between institutions, professionals, different skills is the only way that can guarantee relevant results, averting the failure of traditional library policies envisaged by Kalev Leetaru in the contribution published in 2017 on Forbes.com, with the significant title *Why Are Libraries Failing At Web Archiving And Are We Losing Our Digital History?*

[36] Regarding the participation of state institutions, resources should be stable and linked to a specific budget chapter.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

References

AIB Biblioteche e servizi nazionali, Commissione nazionale. 2020. «Il deposito legale regionale in Italia: stato dell'arte e risultati di una recente indagine». *AIB Studi* 59 (3): 423–52. doi:10.2426/aibstudi-12019.

Alloatti, Franca. 2008. «L'attuazione della Legge 106 tra incognite e speranze». *Biblioteche oggi* 26 (1): 25–33.

Bergamin, Giovanni. 2012. «La raccolta dei siti web: un test per il dominio "punto it"». *DigItalia* 2 (0): 170–74.

———. 2018. «Conservazione del patrimonio culturale digitale». In *60° Congresso nazionale AIB, 22-23 novembre 2018*.

Bergamin, Giovanni, e Maurizio Messina. 2010. «Magazzini digitali: dal prototipo al servizio». *DigItalia* 2 (0): 144–53.

Biblioteca Nazionale Centrale di Firenze. «Archiviabilità dei siti web». Accessed April 1st, 2021. https://www.bncf.firenze.sbn.it/biblioteca/archiviabilita-dei-siti-web/.

«Copyright reform (archived) - European Bureau of Library Information and Documentation Associations (EBLIDA)». Accessed October 2, 2021. http://www.eblida.org/about-eblida/archive/copyright-reform/.

Direzione generale Biblioteche e diritto d'autore. «Accordo MiBAC - Associazioni Editori». Accessed April 1st, 2021. https://www.librari.beniculturali.it/it/notizie/notizia/4ee4df59-4819-11e1-88f7-b7fd06d12128/.

Floridi, Luciano, a c. di. 2015. *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Cham: Springer International Publishing. doi:10.1007/978-3-319-04093-6.

———. 2017. *La quarta rivoluzione: Come l'infosfera sta trasformando il mondo.* Milano: Raffaello Cortina Editore.

———. 2018. «The good web. Some challenges and strategies to realise it». In *The Web Conference, Lyon, France 23-27 april 2018*.

———. 2020. *Pensare l'infosfera. La filosofia come design concettuale*. Milano: Raffaello Cortina Editore.

Laakso, Mikael, Lisa Matthias, e Najko Jahn. 2021. «Open Is Not Forever: A Study of Vanished Open Access Journals». *Journal of the Association for Information Science and Technology* 72 (9): 1099–1112. doi:10.1002/asi.24460.

«L'archivio storico de La Stampa sarà di nuovo consultabile entro metà febbraio 2021». 2020. *lastampa.it*. dicembre 14. Accessed October 2, 2021. https://www.lastampa.it/rubriche/public-editor/2020/12/14/news/l-archivio-storico-de-la-stampa-sara-di-nuovo-consultabile-entro-meta-febbraio-2021-1.39659298.

«Le raccomandazioni della rete MAB per il recepimento della direttiva europea sul copyright». 2020. *AIB-WEB*. ottobre 18 Accessed October 2, 2021. https://www.aib.it/attivita/mab/2020/85856-raccomandazioni-mab-recepimento-direttiva-europea-copyright/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12732

Leetaru, Kalev. 2021. «Why Are Libraries Failing At Web Archiving And Are We Losing Our Digital History?» *Forbes*. Accessed April 1st, 2021. https://www.forbes.com/sites/kalevleetaru/2017/03/27/why-are-libraries-failing-at-web-archiving-and-are-we-losing-our-digital-history/?sh=f22dacb6ecd4.

Michetti, Giovanni. 2008. «Il modello OAIS». *DigItalia* 1 (0): 32–49.

«OAIS: entro breve la revisione». ParER - Polo archivistico dell'Emilia-Romagna. Accessed April 1st 2021. https://poloarchivistico.regione.emilia-romagna.it/news-in-evidenza/oais-entro-breve-la-revisione.

«PREMIS: Preservation Metadata Maintenance Activity (Library of Congress)». Accessed April 1st 2021. http://www.loc.gov/standards/premis/.

«Quasi un titolo ebook su due è nato con il self publishing. L'indagine Aie: grandi numeri, ma il mercato è meno di un terzo». 2016. *Prima online*. dicembre 10. Accessed October 2, 2021. https://www.primaonline.it/2016/12/10/251136/quasi-un-titolo-ebook-su-due-e-nato-con-il-self-publishing-lindagine-aie-grandi-numeri-ma-il-mercato-e-meno-di-un-terzo/.

«Razionalizzazione del Patrimonio ICT|Agenzia per l'Italia digitale». Accessed April 1st, 2021. https://www.agid.gov.it/it/infrastrutture/razionalizzazione-del-patrimonio-ict.

«Riappare online (con un curioso sottotitolo) l'archivio storico de L'Unità. Mistero sull'autore dell'operazione». 2018. *Prima online*. ottobre 16. Accessed October 2, 2021. https://www.primaonline.it/2018/10/16/279189/riappare-online-con-un-curioso-sottotitolo-larchivio-storico-de-lunita-mistero-sullautore-delloperazione/.

Rosenblatt, Bill. 2017. «Readium LCP Set to Launch». *Copyright and Technology*. marzo 11. Accessed October 2, 2021. https://copyrightandtechnology.com/2017/03/11/readium-lcp-set-to-launch/.

Storti, Chiara. 2019. «Storage, enhancement and preservation of doctoral dissertations in the experience "Magazzini digitali": a contribution to research and access». *JLIS.it* 10 (1): 114–24. doi:10.4403/jlis.it-12526.

Tedeschini-Lalli, Mario. 2021. «Tecnologia Digitale Obsoleta, Un Secolo e Mezzo Di Storia a Rischio». *Medium*. febbraio 23. Accessed October 2, 2021. https://tedeschini.medium.com/tecnologia-digitale-obsoleta-un-secolo-e-mezzo-di-storia-a-rischio-1bb75bf68c2f.

Tremolada, Luca. 2019. «Quanti dati sono generati in un giorno?» Info Data (blog - Il Sole 24 ore), maggio 14. Accessed April 1st, 2021. https://www.infodata.ilsole24ore.com/2019/05/14/quanti-dati-sono-generati-in-un-giorno/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

JLIS.it

Thesauri in the Digital Ecosystem

Anna Lucarelli[(a)]

a) Biblioteca nazionale centrale di Firenze

Contact: Anna Lucarelli, anna.lucarelli@beniculturali.it
Received: 5 May 2021; **Accepted:** 21 May 2021; **First Published:** 15 January 2022

ABSTRACT

In recent years, thesauri have taken on new roles, new functions, and have shown some advantages over other knowledge organitazion systems (KOS). They are increasingly important in the linked data environment of the semantic web. The *Nuovo soggettario*, created and maintained by the National Central Library of Florence, is an example of the changing uses of controlled subject systems, like thesauri and subject heading lists. Thesauri are shown to be dynamic tools, essential components for the integration of data on the web, especially for mapping and to assist with interoperability among heterogeneous resources. With the adoption of formats of the semantic web, such as RDF/SKOS, and following international standards, thesauri have evolved and have proven to be increasingly useful with free reuse and across various frameworks. To varying degrees, they have enabled increased multilingualism and conceptual equivalences, connecting information and metadata produced by institutions of different countries. As authority control systems, they interact with Wikidata and help build 'bridges' between worlds that were too far apart until not long ago, namely libraries, archives, and museums. Will the challenge of search engines, machine learning and artificial intelligence override the thesauri or will it make them even more involved?

KEYWORDS

Bibliographic control; Linked open data; Nuovo soggettario; Thesauri.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Thesauri and bibliographic control

Since the beginning of IFLA's UBC Programme, universal bibliographic control has been primarily focused on the sharing and standardization of descriptive cataloguing. Talking about thesauri gives rise to the following questions:

- the current state and the possible new future of subject indexing of which thesauri are essential components, along with subject heading lists (Petrucciani 2019, 163-173);
- the path followed by subject indexing in recent years; a path that is considered 'autonomous' compared to other cataloguing processes; a path strongly connected to the procedures and to the languages used in various countries, in several cultural, geographical and, above all, linguistic contexts;
- the relationship of thesauri with other knowledge organization systems (KOS), such as ontologies, classifications, taxonomies, and so on (Gnoli 2020);
- the role they play in current bibliographic control, considering that the concept of bibliographic control in the digital ecosystem is changing and is evolving *Dalla catalogazione alla metadazione* (from cataloguing to creating metadata), just to use the title of a recent book (Guerrini 2020) and "transitioning to the next generation of metadata" (Smith-Yoshimura 2020). This is a period when cataloguing tools, data management, and infrastructures are more than ever crossing transitional borders, tied to strategies to make bibliographic data more visible on the web.[1]

We now have an opportunity for a better integration of subject data in universal bibliographic control.
As we will see, thesauri have taken on new roles, new functions and shown some advantages over other knowledge organization systems (KOS). Yet, there are many arguments and confrontations on this issue.
In their multiple types (general or specialized domains; polyhierarchical or monohierarchical, monolingual or multilingual, etc.), thesauri continue to prove their effectiveness compared to simple lexicons or *flat lists*; they have proved to be versatile, usable, both in the framework of the post-coordinated and pre-coordinated languages, in which the rules for the citation order in the subject strings are added to vocabulary control.

Controlled vocabularies are studied by the Subject Analysis and Access Section of IFLA[2], but even by the International Society for Knowledge Organization (ISKO) with its regional chapters.[3] Thesauri are also handled by terminology associations,[4] a transversal discipline. Unfortunately, the communities that are involved are not always interactive among one another. The relationship between terminology experts and librarians engaged in subject indexing still continues to be weak and not as creative as it could be.

[1] For a selective bibliography on the current state of the subject indexing, specifically with regard to the French reality, see: *L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique* 2020.
[2] https://www.ifla.org/subject-analysis-and-access.
[3] https://www.isko.org/.
[4] E.g., Associazione italiana per la terminologia (Ass.I.Term): http://www.assiterm91.it/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Even thanks to their formalized structures, thesauri have been significantly supported by standardization. Subject indexing is one of the rare library tasks specifically regulated by the International Organization for Standardization (ISO). Let's mention the ISO 5963:1985 standard (*Methods for examining documents, determining their subjects, and selecting indexing terms*) on the conceptual analysis, recently validated in 2020, and ISO 25964:2011-2013 (*Thesauri and interoperability with other vocabularies*), just concerning the thesauri themselves, and renewing ISO-5964:1985 and ISO-2788:1986, established before the digital universe existed. However, these are not the only standards about both documentation and terminology. Within the Italian framework, groups and technical committees of Commissione UNI CT/014[5] also deal with them.

Rise or fall of thesauri?

The national libraries have tried to make their vocabularies, used for subject indexing, more 'visible' and usable, through various modes of integration with their own OPACs or with the open data hubs.

Subject indexing in libraries has however suffered a slowdown, not only because the data referred to semantic contents is considered to be less necessary, but also because indexing is a labor-intensive and hence expensive process. The lack of human resources is not only an Italian problem, though.

Nevertheless, in recent years, the development of thesauri has spread everywhere.

This spread is studied by BARTOC, a database developed and maintained at the University of Basel, since 2013. It inventories various systems for the organization of knowledge, including web applications and mapping, so far totalling 3,393 (data as of November 2021).[6]

In particular for thesauri, BARTOC monitors the establishment of thesauri across the world, by describing them by their main features, identifying 781 thesauri to date. Almost all of them are in digital format and freely accessible on the web by open licenses, many of them being multilingual.

Thesauri are inventoried and assessed by librarians and institutes around the world. Assessment includes for their performance, for their compliance with the standards, and for their level of semantic coverage.

They are assessed according to their ability to handle their own terminology expansion, starting with a particular *corpora*, and for their ability to be representative with regard to specialized domains (Folino and Parisi 2020). Such issues in Italy are examined not only by librarians, but also by CNR Institutes and by centres of excellence such as the Laboratorio di Documentazione dell'Università della Calabria.[7]

Thesauri are further assessed for the possibility of being integrated with algorithms employed for the automated indexing.

Ultimately, they are assessed according to the quality of their data (e.g., ability to be re-used) and according to the sustainability of the resulting costs, also considering that the personnel involved in creating and maintaining thesauri are required an indispensable professional development.

[5] https://www.uni.com/index.php?option=com_uniot&view=struct&id=853557&Itemid=2447.
[6] http://bartoc.org/.
[7] https://www.labdoc.it/.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

We can think that the development of these tools may depend on the fact that the terminology has acquired more and more importance within "metadata-ing". Yet, it is not only a matter of this. No longer limited to the library and documentation world, these tools have actually gone beyond the context of subject indexing and of information retrieval (IR); they have been involved in other 'universes'.

Following the standardization established by ISO 25964 and by RDF/SKOS,[8] thesauri organize both the concepts and the terms by which they are represented. The central role of the concept (that is a unit of thought, rather than a lexical element of a specific language), has ensured that the borders which separate thesauri from other knowledge organization systems have become more fluid. There are two reasons for this: for the possibility of the correct *reconciliation* of expressions in different languages, and for the opportunity to compare different systems starting from the conceptual cores on which they are established. It is no coincidence that "metathesauri" have also been set up (e.g., UMLS by the National Library of Medicine in the United States[9]).

An approaching process has been activated among classifications, schemes based on subject headings and ontologies; in the latter, the relationships among concepts are less standardized. The very relational structure of thesauri, when rigorous, encourages their evolution towards the ontologies (Biagetti 2018; Biagetti 2020).

Vanda Broughton, Leonard Will and Stella Dextre Clarke have faced such interesting issues in a recent series of virtual classes organized in 2020-2021 by ISKO UK.[10]

One characteristic of thesauri is that of being dynamic tools, obviously linked to the linguistic fabric of the context in which they are established, yet, often, through multilingual functionalities. In addition, thesauri have shown their capabilities, not only as 'tools of the trade' for librarians but even as tools for users. Yet, we know that this happens if they are provided in the right way, if they are 'well integrated' in OPACs, and if librarians employ them also as a support to reference service and to information literacy (Ballestra 2011, 395-401).

The reason why they are so costly is due to the constant maintenance work they require, along with a careful supervision of the increase mechanisms in relation to the 'literary warrant'.

They also require a continuous assessment of their structural coherence, a monitoring of the semantic relationships, particularly for synonymy and lexical variants on the one side, polysemy and new meanings on the other side.

Languages (of works, of users, of catalogues) quickly evolve, so the work to be carried out on neologisms is continuous. To give a current example, let's think about the importance to 'control' concepts connected to the pandemic we are living in and which are the subjects of works already published. SARS-CoV-2; COVID-19; Social distancing; Confinement; Lockdown; Contact tracing… (see Figures 1-5). These terms were added promptly and captured some of these new concepts from the first months of 2020. Not all vocabularies acquire new concepts with the same timeliness.[11]

[8] https://www.w3.org/2004/02/skos/.

[9] https://www.nlm.nih.gov/research/umls/index.html.

[10] https://www.iskouk.org/KOED.

[11] For a first survey on the terms tied to COVID-19 pandemic, inserted into Thesaurus of *Nuovo soggettario* already since March 2020: Francioni and Lucarelli 2020. On the Italian words about pandemic also Accademia della Crusca: https://accademiadellacrusca.it/it/contenuti/lacruscaacasa-le-parole-della-pandemia/7945.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Examples of new concepts related to the current world health crisis:

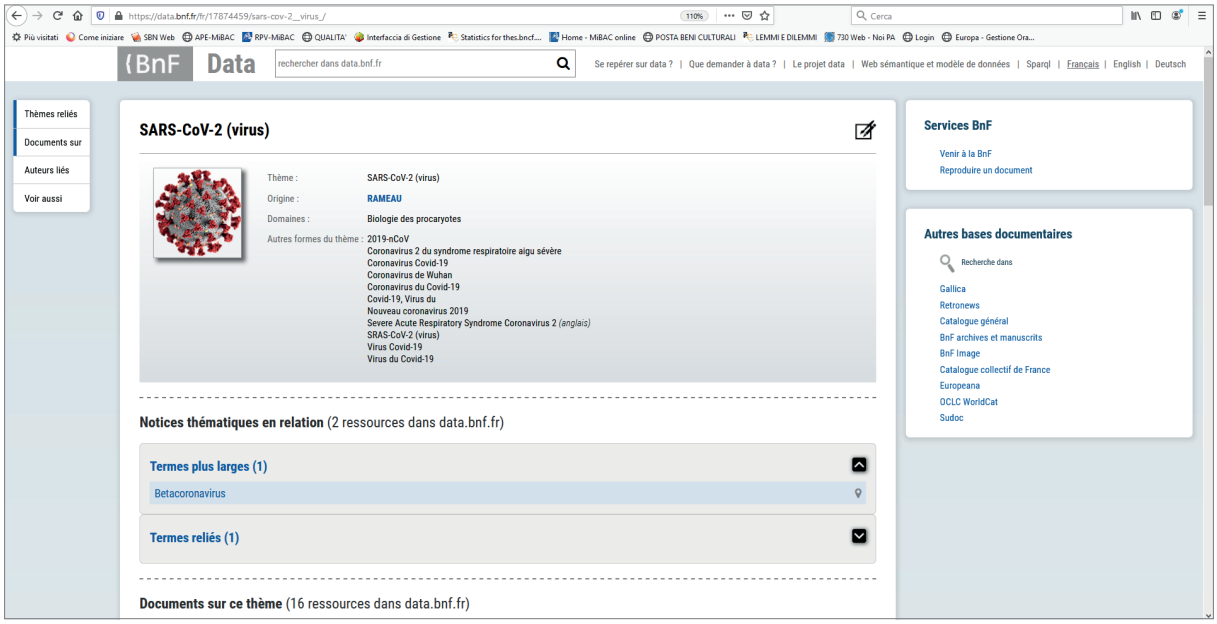

Fig. 1. The concept *SARS-CoV-2 (virus)* in RAMEAU

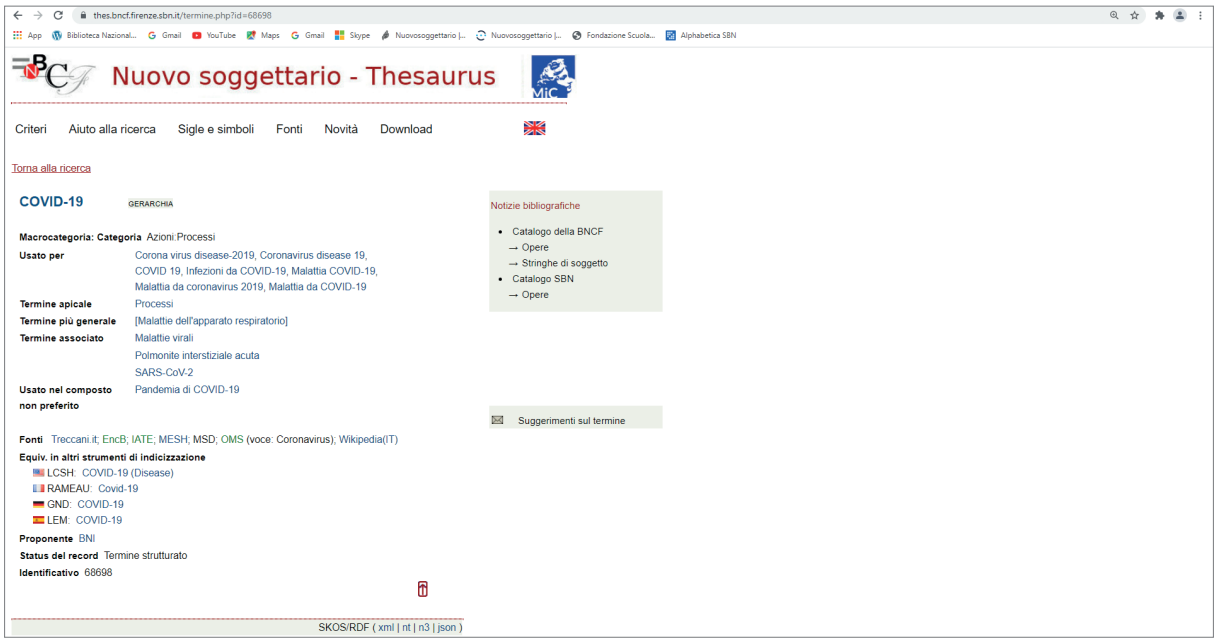

Fig. 2. The concept *COVID-19* in *Nuovo soggettario*

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

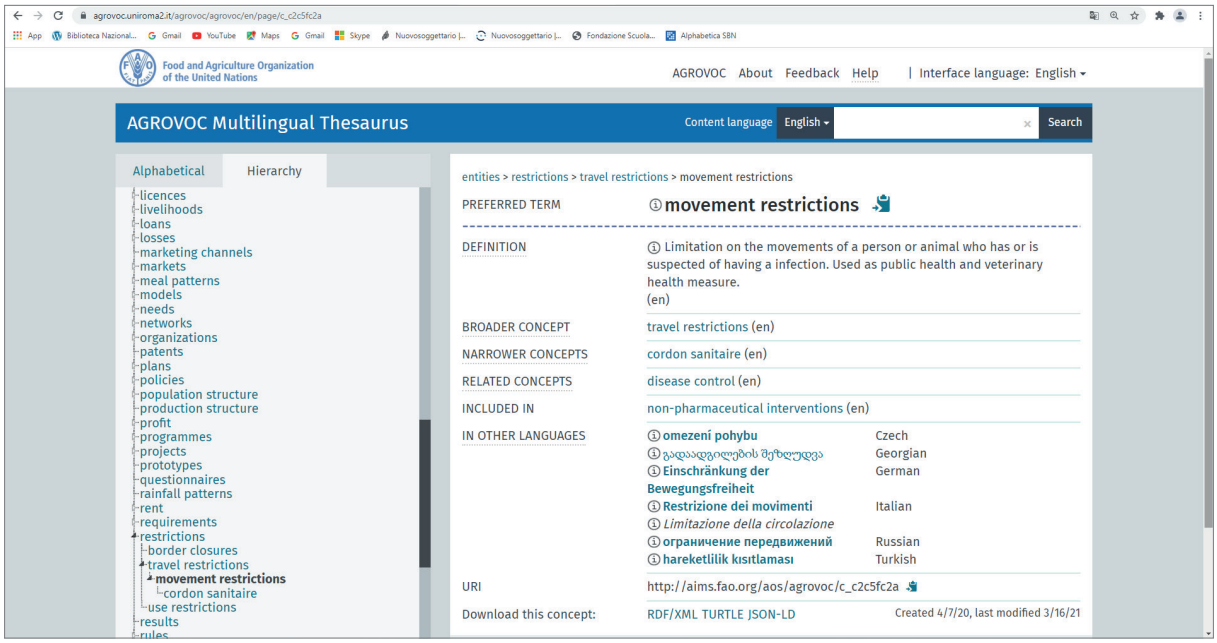

Fig. 3. The concept *Movement restrictions* in AGROVOC

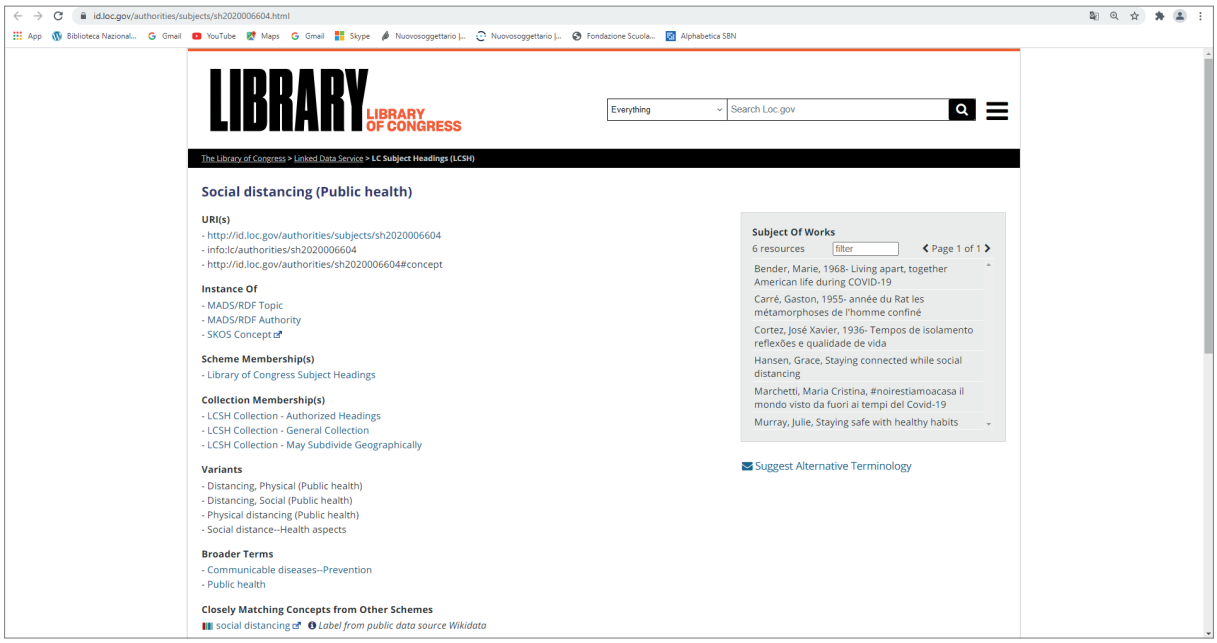

Fig. 4. The concept *Social distancing (Public health)* in LCSH

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

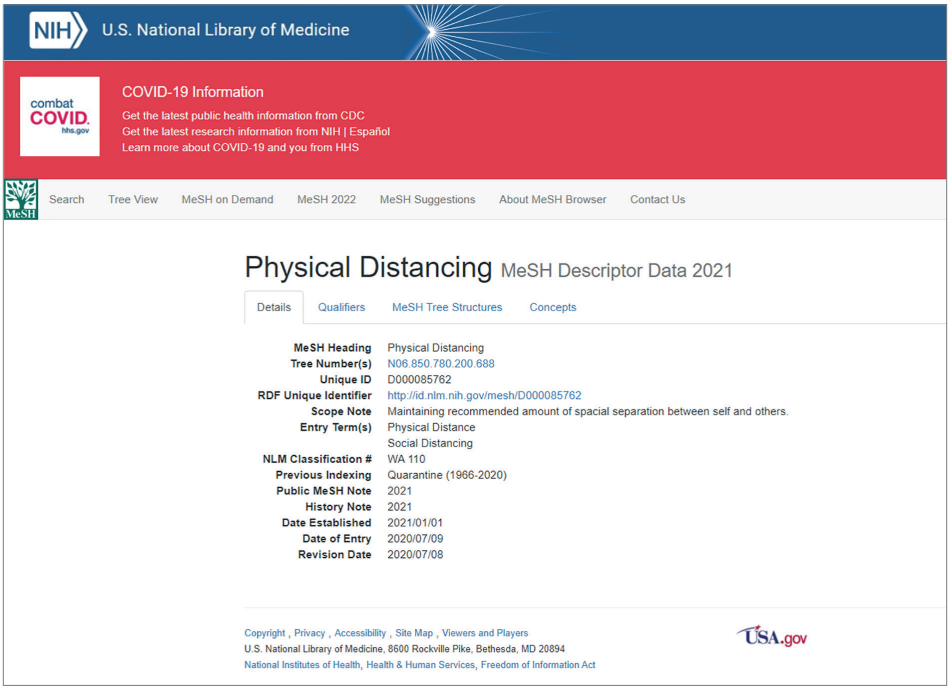

Fig. 5. The concept *Physical Distancing* in MESH

Integration of data on the web

What is important is that thesauri have proved to be the essential components for the integration of data on the web and thus fundamental elements for the affirmation of the semantic web.
We are dealing with the role of the thesauri within the semantic web at various levels and in various contexts and there are many studies on this topic (e.g., Martinez-González and Alvite Díez 2019). We could say that they are among 'the best friends' of the semantic web, for their capability to provide metadata in RDF, that is to say in open formats which allow their re-use in the most varied contexts (not necessarily library ones), because they encourage the development of mapping as well as the interoperability between heterogeneous resources (Zeng 2019, 122-146).
When we wonder which is the most re-used data among those processed by libraries, thesauri are a good example.
Many of them are connected with DbPedia.[12] Tens of other thesauri have recently connected to Wikidata.[13] The Italian Thesaurus of *Nuovo soggettario*[14] – created and maintained by the National Central Library of Florence (BNCF) – has had links with Wikipedia since 2007. Since

[12] https://wiki.dbpedia.org/.
[13] https://www.wikidata.org/w/index.php?title=Special:WhatLinksHere/Q89560413&limit=500.
[14] https://thes.bncf.firenze.sbn.it/ricerca.php. BNCF, with an almost centuries-old tradition for subject indexing (started in 1925), has the institutional task to curate the Italian subject indexing tools. *Nuovo soggettario* contains the concepts/terms employed in the framework of a pre-coordinated language that contemplates also the rules on the construction of the subject strings. Yet, thesaurus is obviously usable also for the post-coordinated indexing. It is employed by the Italian National Bibliography (BNI) and by most libraries of the Italy's National Library Service (SBN). It was also presented during IFLA General Conference 2009 (Cheti, Alberto, Anna Lucarelli, and Federica Paradisi. 2009).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

2013, reverse mapping has been implemented with a mutual browsing mechanism as well as with a synchronization realized through the field P508 (BNCF Thesaurus ID) of Wikidata (Lucarelli 2014).[15]

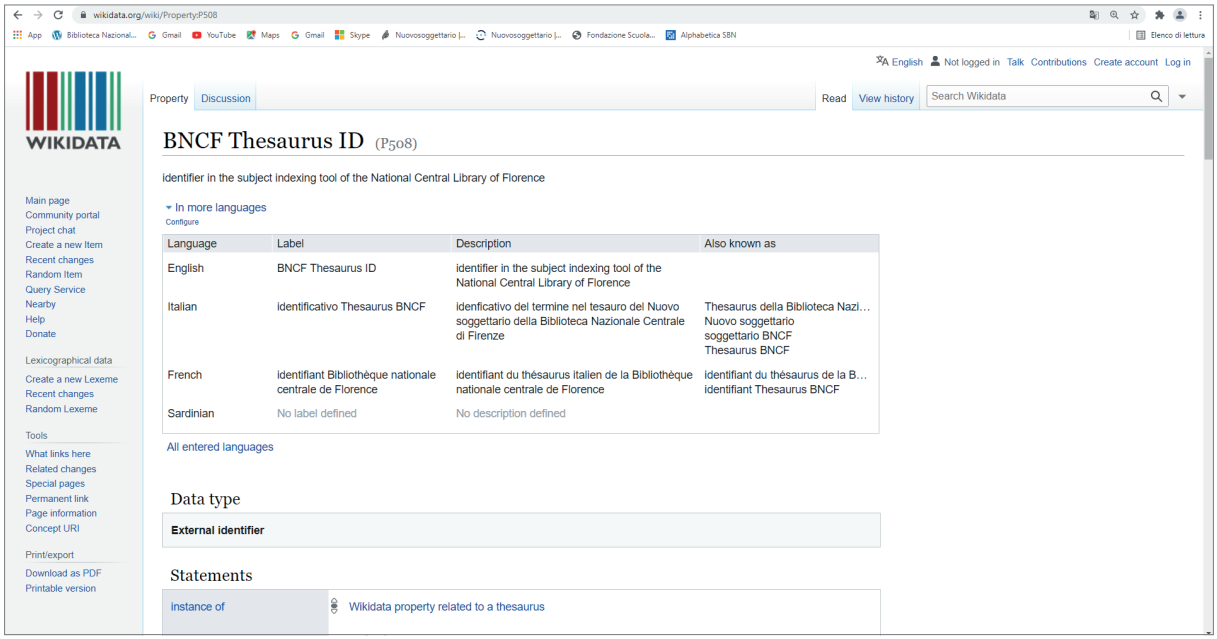

Fig. 6. Thesaurus of the National Central Library of Florence and Wikidata

Since thesauri are among the 'main actors', the interpreters of the semantic web, we must evaluate their costs on the basis of the benefits they bring to the linked open data and on the possibility of creating mapping, as Stella Dextre Clarke has recently reminded us in the above-mentioned virtual classes[16].
The opportunities offered by open data and by mapping, in both the research world and public administrations, are unquestionable. Some examples?
A few years ago, the City of Florence made use of the open data of BNCF's Thesaurus in order to organize the City's open data.
When, in May 2013, the reverse mapping from the Wikipedia entries to the corresponding terms of *Nuovo soggettario* was implemented, the number of visitors to the OPAC of BNCF has increased 28% in only one month.

As we can see from this example, it is no longer possible to talk about thesauri without talking about interoperability. The international standard ISO 25964 dedicates the second of its two parts to the methods for the realization of this interoperability. Furthermore, the interoperability ac-

[15] The *Nuovo soggettario* was the first general thesaurus to activate a form of interlinking with a version of Wikipedia in a specific language, preceded, at an international level, by experiences in specialized sectors as in the case of Thesaurus for Economics of Leibniz-Informationszentrum Wirtschaft (http://zbw.eu/stw/version/latest/about).
[16] About the costs resulted from the procedures of the bibliographic control, also Bergamin 2020, p. 167.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

tivated by thesauri has made them fundamental 'hubs' as well as 'bridges' for the connection between data from different institutions. Mapping also has been realized with particular success in the context of multilingualism. Not only multilingual vocabularies (such as the notable AGROVOC, AAT, EUROVOC, IATE, that are often cited, in a crossed mode, interconnecting one another), but also monolingual vocabularies with equivalences in other languages in the form authorized by those other vocabularies or subject heading schemes. At the same time, Pat Riva explained the importance of multilingualism and of the internationalization of the bibliographic description in order to facilitate access (Riva 2021).

In the revolution of open data, thesauri are thus on the 'front line'. Many of them have implemented new formats for the publications and the exchange of metadata (i.e., SKOS) by exceeding the previous ones (i.e., Zthes). They have become structures "of" the web.
In the linked open data cloud, many controlled vocabularies are represented, including those created and maintained by the national libraries.[17]
The Thesaurus of *Nuovo soggettario* has been in SKOS since 2010 and has achieved the 'five stars' of Tim Berners-Lee.[18] It can also be found in the hub *dati.beniculturali* of the Ministero della Cultura.[19]

The initiatives of national libraries and national bibliographies

Since the publication of IFLA's *Guidelines for subject access in National Bibliographies* ten years have passed, but many indicated best practices are still valid.[20] Following these guidelines, both national libraries and national bibliographies that are assigned to the bibliographic control of our countries, have implemented important choices in the field of subject indexing.
Many national libraries have updated their bibliographic tools to follow the latest standards and entered the world wide web of data, following new 'conceptual models'.
For some of these institutions it has been a period of reforms, like for the Bibliothèque Nationale de France which, in 2019, made public its *Réforme de Rameau.*[21]
Regardless of the subject indexing language used, the national libraries continue to benefit from the controlled vocabularies even when indexing graphic resources, audio resources, ancient works and, in certain countries, works of fiction as well. They even use controlled vocabularies when providing Genre/Form descriptions, a practice that is also supported by IFLA.[22] In some cases, they use expressly dedicated thesauri, for the indexing of particular types of resources, for instance, the *Library of Congress Genre/Form Terms for Library and Archival Materials* (LC-GFT).[23]

[17] https://lod-cloud.net/clouds/publications-lod.svg.
[18] https://lod-cloud.net/dataset/bncf-ns.
[19] https://dati.beniculturali.it/altri-linked-open-data-del-mibact/.
[20] https://www.ifla.org/publications/ifla-series-on-bibliographic-control-45.
[21] https://rameau.bnf.fr/syntaxe.
[22] https://www.ifla.org/node/8526.
[23] https://id.loc.gov/authorities/genreForms.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

National libraries generally use these vocabularies for projects of automated indexing or semi-automated indexing of online resources, by having them interact with implemented algorithms. For example, this is part of the subject cataloguing policies of the Deutsche Nationalbibliothek, as explained by Ulrike Junger since the beginning (Junger 2018), and also more recently described by Mödden and Suominen (Mödden 2021; Suominen 2021).
In the name of the data quality, the use of vocabularies continues to rely on uncontrolled keywords. Thanks to mapping to RDF and to open data's hubs, the national libraries' vocabularies encourage a connection among different OPACs, which hopefully is a prelude to additional future forms of connections; some of these connections were originated from the project named MACS (Multilingual Access to Subjects) which was exceptionally innovative and whose operational phase started in 2005.[24]

As we can see in the figures below, starting a search with the subject term employed by the Deutsche Nationalbibliothek, one sees the connected publications, but it is also possible to move to the French equivalence of data.bnf, where resources on that topic can be explored. Through the correspondent RAMEAU page, it is possible to browse towards *Library of Congress Subject Headings* (LCSH) where works about the same topic can be explored in the catalogue of the Library of Congress. The equivalences are generally ensured by the form of the closely matching concepts from other schemes, as well evidenced by LCSH.

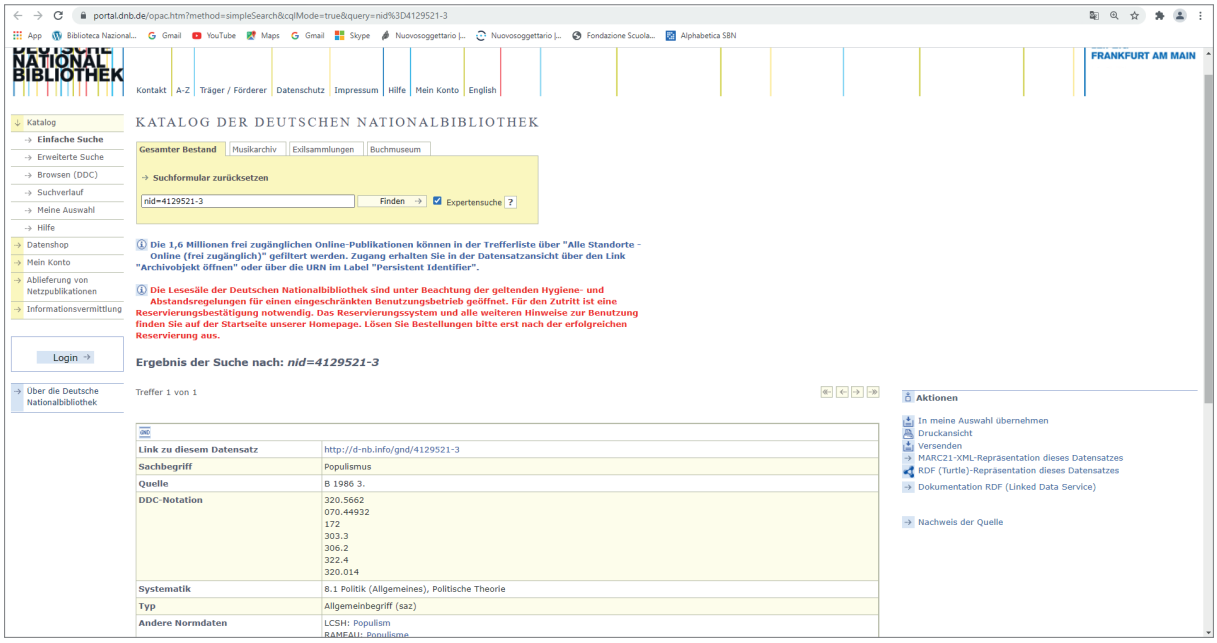

Fig. 7. Concept with equivalences in Gemeinsame Normdatei (GND) and links

[24] https://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/9041.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

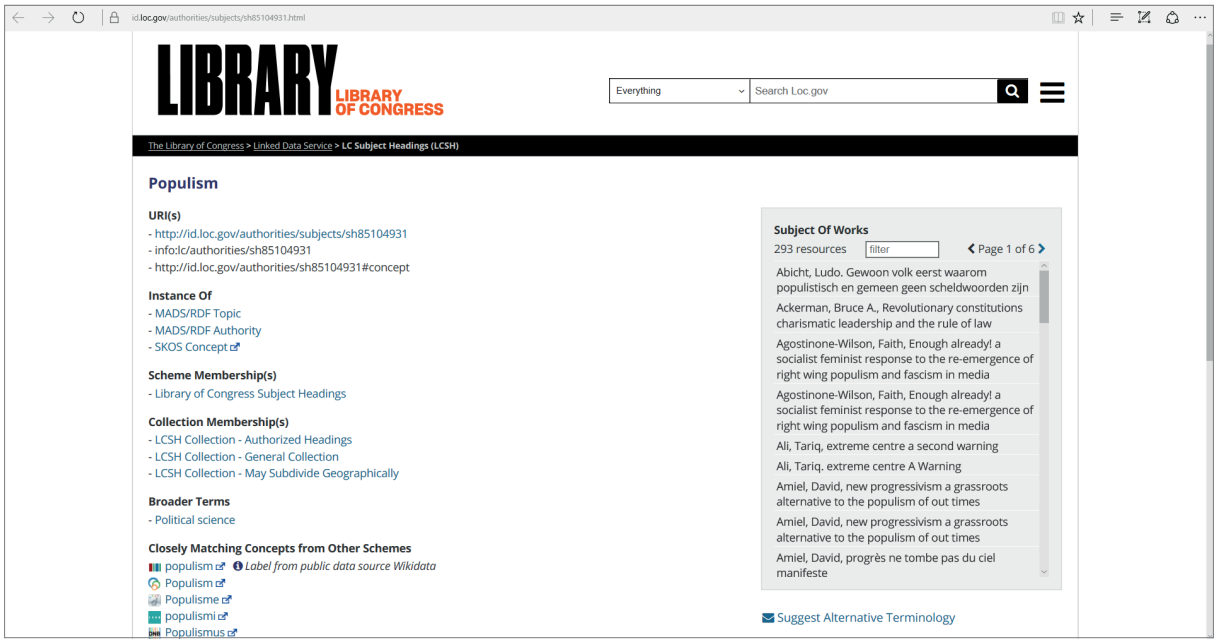

Fig. 8. Concept with equivalences in *Library of Congress Subject Headings* (LCSH) and links

Likewise the Thesaurus of *Nuovo soggettario* has been connected to the works described in the online catalogues of the National Central Library of Florence and Italy's Servizio Bibliotecario Nazionale (SBN), as shown in Figure 9, it has also been possible to navigate to Datos. BNE, that is, to the controlled equivalents of the Biblioteca Nacional de España, and, from there, it has been possible to explore the *Obras* on the same subject in the BNE catalogue.[25]

[25] https://datos.bne.es/tema/XX525409.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

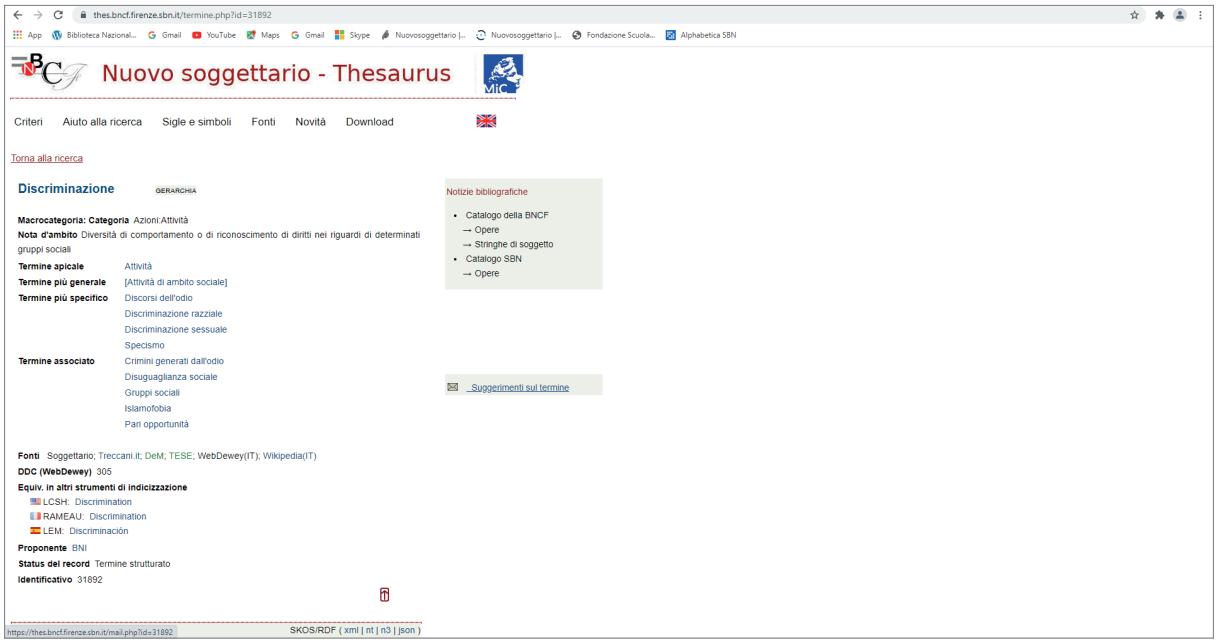

Fig. 9. Concept with equivalences in *Nuovo soggettario* and links[26]

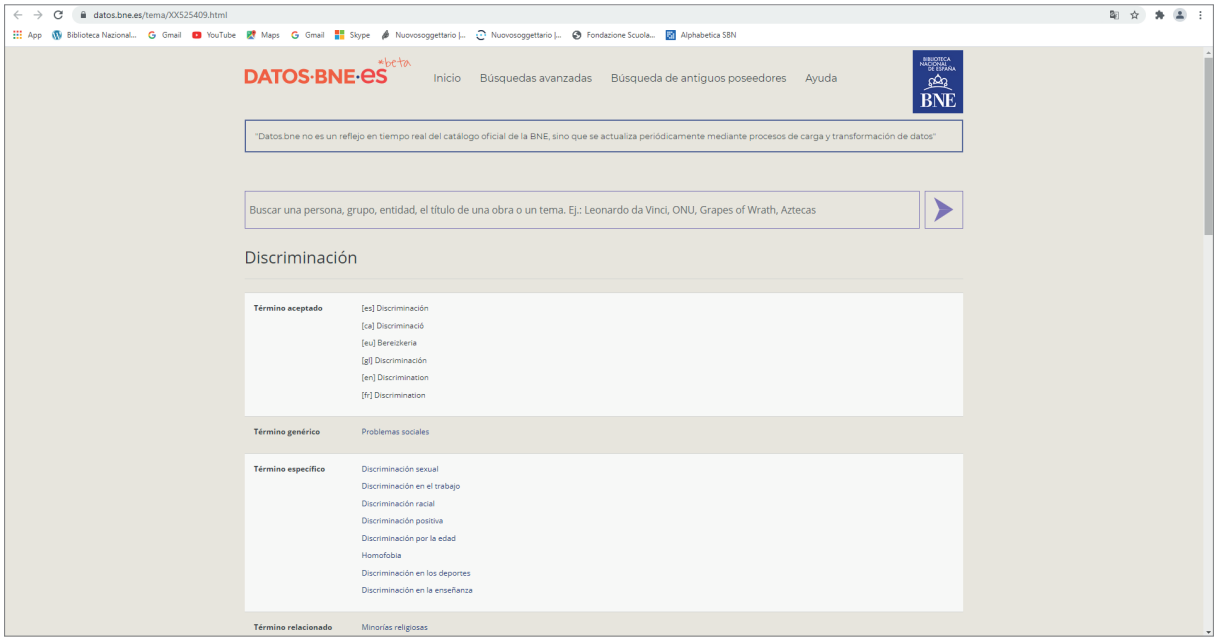

Fig. 10. Concept with equivalences in EMBNE and links

[26] The figure refers to the result of the research carried out on the Thesaurus at the time of the Conference (8th-12th February 2021) . For current results, see: https://thes.bncf.firenze.sbn.it/termine.php?id=31892

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

In fact, over the years, the Italian Thesaurus of *Nuovo soggettario* has considerably increased the mapping with other KOS and with equivalents of other vocabularies and continues to link to more equivalences (Viti 2017, 624-637). The number of links with LCSH has increased from 390 in 2011 to the current 14,970; with the French terms of RAMEAU from 380 in 2012 to the current 13,380 links; with the German terms from 130 in 2018 to the current 2,200; with the Spanish terms from 300 in 2019 to the current 2,270.[27]

Making such links is challenging work, requiring careful mapping and not without problems. For instance, there are challenges about the level of equivalences among concepts, especially across languages. This was explained by Pino Buizza in one of his latest papers on mapping between the Thesaurus of *Nuovo soggettario*, in Italian, and the two subject heading lists produced by national bibliographic agencies in the United States and in France: the *Library of Congress Subject Headings*, in English, and the *Repertoire d'autorité-matière encyclopédique et alphabétique unifié*, in French (Buizza 2020, [59]-68).

The equivalences found in *Nuovo soggettario*, when downloaded through SKOS, can activate mutual connections or give rise to the indication of the variant in Italian, as data.bnf shows in Figure 11 under the term *Épidémies*, among the *Autres forms du thème*.

Such initiatives demonstrate the importance that policies be activated among national libraries.

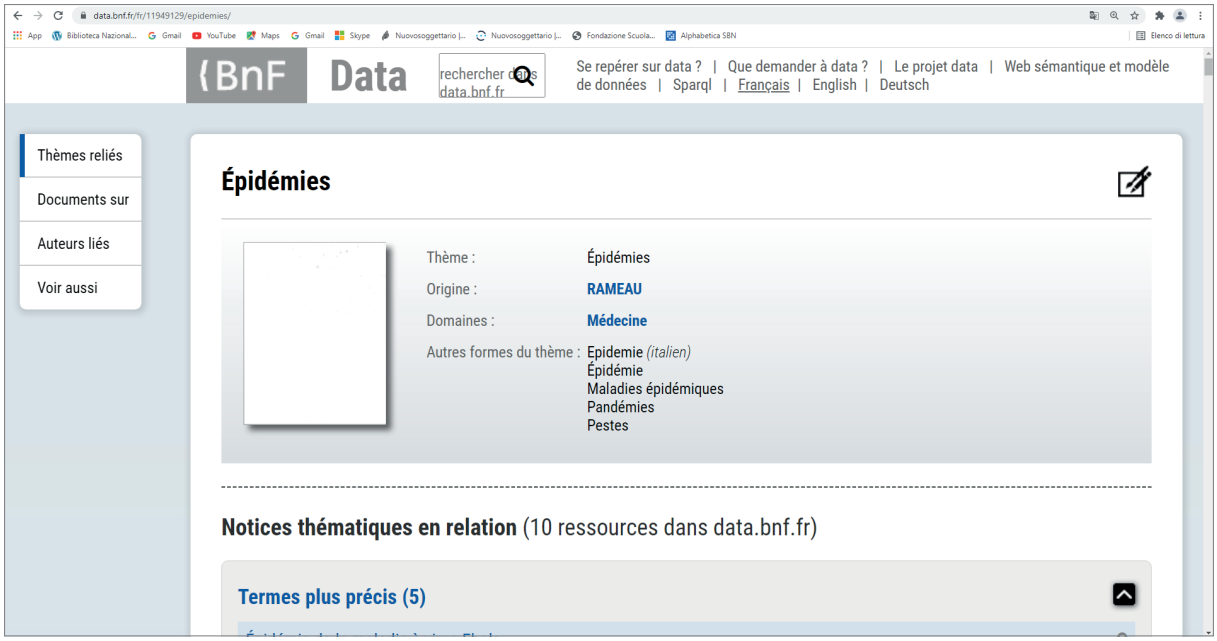

Fig. 11. Mutual connections or the indexing of the variant in Italian

[27] The comprehensive data on the trend of the equivalences in other languages are visible in: https://thes.bncf.firenze.sbn.it/stat.php.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Thesauri and Authority control: connection with other interlocutors

Collaborating on policies does not mean that the different indexing languages used by national libraries and connected through the respective vocabularies must have the same characteristics, the same syntactic rules. Not all tools have the same compliance with the standards, the same structure or functionality. Not all of them are polyhierarchical. Not all of them have comprehensive hierarchies up to the top term. Not all receive both common and proper names. Not all have the same integration with an OPAC and open data.

What brings them together is the progressive alignment among one another, the fact that they achieve common features, for example, to be integrated within Wikidata, so they are all visible on Wikipedia.

Who would have imagined that an encyclopedia would connect its own entries with the most important controlled vocabularies created by national libraries for the purpose of the bibliographic control? At the bottom of the Wikipedia page, you can find the box 'Authority control' with the relevant links.[28]

We know that libraries are not the only producers of bibliographic data, and that other operators are involved in universal bibliographic control. Yet, the data produced by 'certain' major libraries keep reflecting the highest level of quality.

Thesauri and recent features in today's context

Other issues related to thesauri within the digital ecosystem might be added to the above-described panorama. I take a cue from the *Nuovo soggettario* to outline some particularly interesting ones:

1. It has grown in size.
   Nuovo soggettario, in compliance with ISO 25964, has so far had a remarkable quantifiable increase: from 13,000 terms of the prototype to the current 67,000 terms.

2. It has a new interface.
   Since 2020, it has had a new, more user-friendly interface, implemented during the development of BNCF's new web site.

3. It interfaces with classification systems.
   Beyond the above-mentioned multilingualism, *Nuovo soggettario* maps with the Italian WebDewey (Crociani, Giunti, and Viti 2016), etc.

4. It has increased coverage in various subject domains.
   Thanks to the institutions that collaborate with BNCF,[29] it has largely enhanced its general coverage and expanded coverage in specific domains.

[28] See, for instance, the connections to the main thesauri at the footnotes of *Arredo urbano* of the Wikipedia in Italian language through the "Controllo di autorità": https://it.wikipedia.org/wiki/Arredo_urbano.
[29] https://thes.bncf.firenze.sbn.it/enti.htm.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

5. It can be employed for the Genre/Form indexing.
   This will be possible once our OPACs implement the tag MARC 655.

Also, the Thesaurus of *Nuovo soggettario* is employed apart from BNCF for the subject indexing of specialized resources:

- for audio and audiovisual resources, as for instance, in projects on oral sources of the Istituto centrale per i beni sonori e audiovisivi (ICBSA) (Magrini 2021);
- for graphic resources, for instance for photographs, also in BNCF but additionally in photographic libraries, for example, in the Fototeca - Biblioteca Panizzi;[30] for iconographic resources and maps, for instance in the Museo Galileo (Pocci 2020);[31]
- for archival resources, for example, for the documents indexed in projects of BNCF in collaboration with both Soprintendenza archivistica e bibliografica della Toscana,[32] and Historical Archives of the European Union.[33]

Integration of the *Nuovo soggettario* with databases of archives and museums

This connection of the Thesaurus of *Nuovo soggettario* with databases of archives and museums is quite interesting.

Let's look first at the Gallerie degli Uffizi, one of the most important museums in the world.[34] In 2019, BNCF started a partnership, a "Research pact," with the Uffizi.[35]

From *Violini* of *Nuovo soggettario* [36] it is possible to browse through the records of the Gallerie degli Uffizi catalogue thanks to the connection with *Violino* of the "Scheda OA" (Opere/oggetti d'arte) for the object's definition.[37] A reverse connection can also be seen from the record of the Museum. When the concept from the *Nuovo soggettario* indicates an iconographic subject (for instance *Albero della vita* [tree of life]), the link is to the Uffizi works that represent that subject, as shown in Figure 12.

From some terms, for example *Sestanti* [Sextant] (as shown in Figure 13), it is possible to view the resources of both the Gallerie degli Uffizi and the Museo Galileo.[38]

[30] http://panizzi.comune.re.it/Sezione.jsp?titolo=Fototeca&idSezione=233.

[31] https://www.museogalileo.it/it/biblioteca-e-istituto-di-ricerca/biblioteca-digitale/collezioni-tematiche/747-bibliotheca-perspectivae.html.

[32] http://sa-toscana.beniculturali.it/index.php?id=2.

[33] https://www.eui.eu/en/academic-units/historical-archives-of-the-european-union.

[34] https://www.uffizi.it/.

[35] https://www.beniculturali.it/comunicato/uffizi-e-biblioteca-nazionale-di-firenze-patto-per-la-ricerca.

[36] https://thes.bncf.firenze.sbn.it/termine.php?id=17664.

[37] http://www.iccd.beniculturali.it/it/ricercanormative/29/oa-opere-oggetti-d-arte-3_00.

[38] https://thes.bncf.firenze.sbn.it/termine.php?id=30976; http://catalogo.uffizi.it/it/29/ricerca/iccd/?search=*&fromRA=true&filter_OGTD-words=%3D&filter_OGTD=Sestante; https://catalogo.museogalileo.it/oggetto/Sestante.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

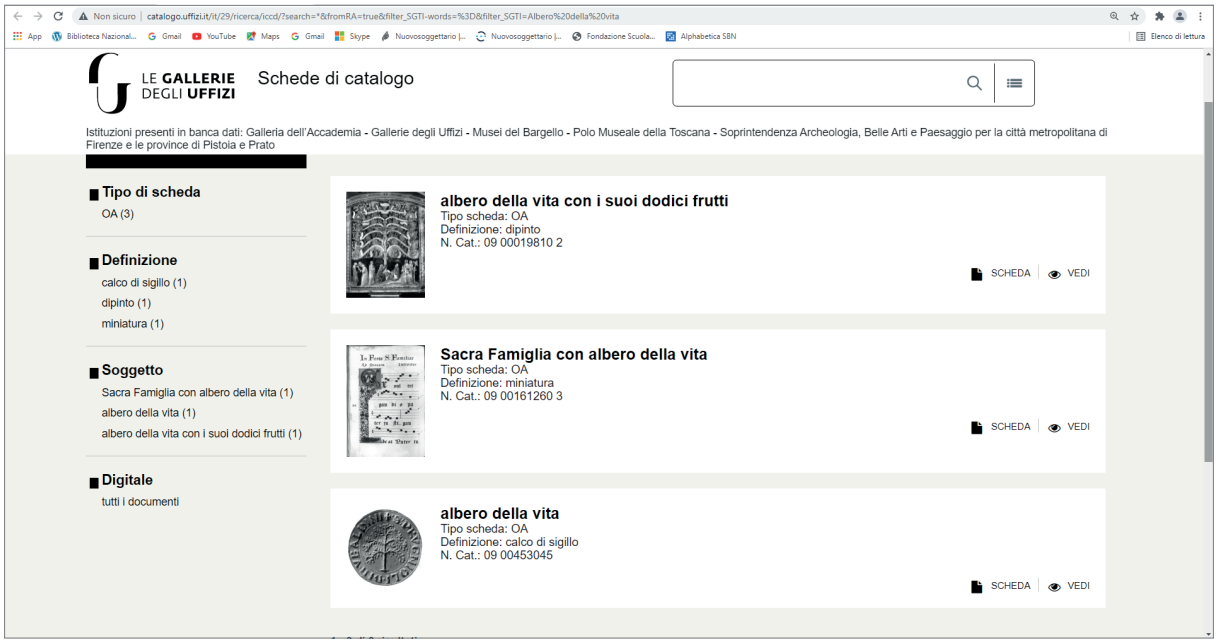

Fig. 12. The Uffizi works on *Albero della vita*

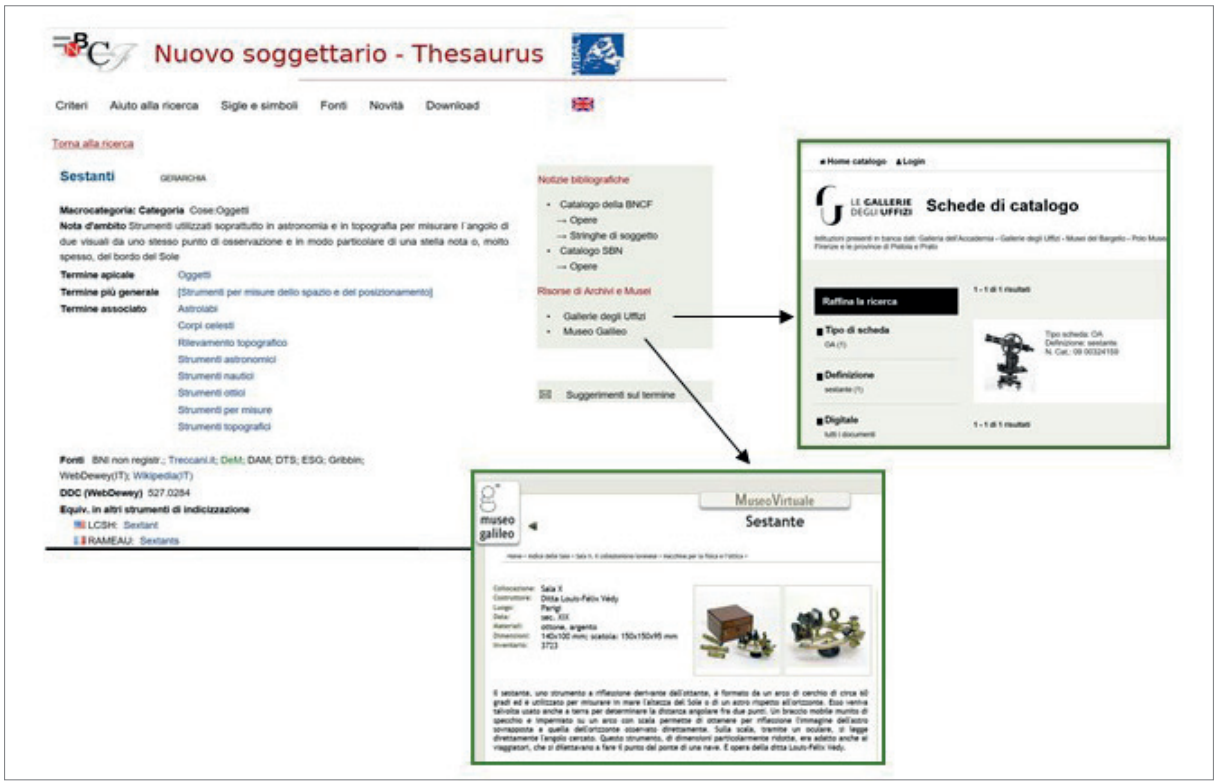

Fig. 13. The term *Sestanti* as seen in A. the *Nuovo soggettario*, B. the Gallerie degli Uffizi's Schede di catalogo, and C. the Museo Galileo's Museo virtuale

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

An example of links with archives can be seen in Figure 14, where *Federalisti europei* in the *Nuovo soggettario* is linked with the Ernesto Rossi fund of the Historical Archives of the European Union.[39]

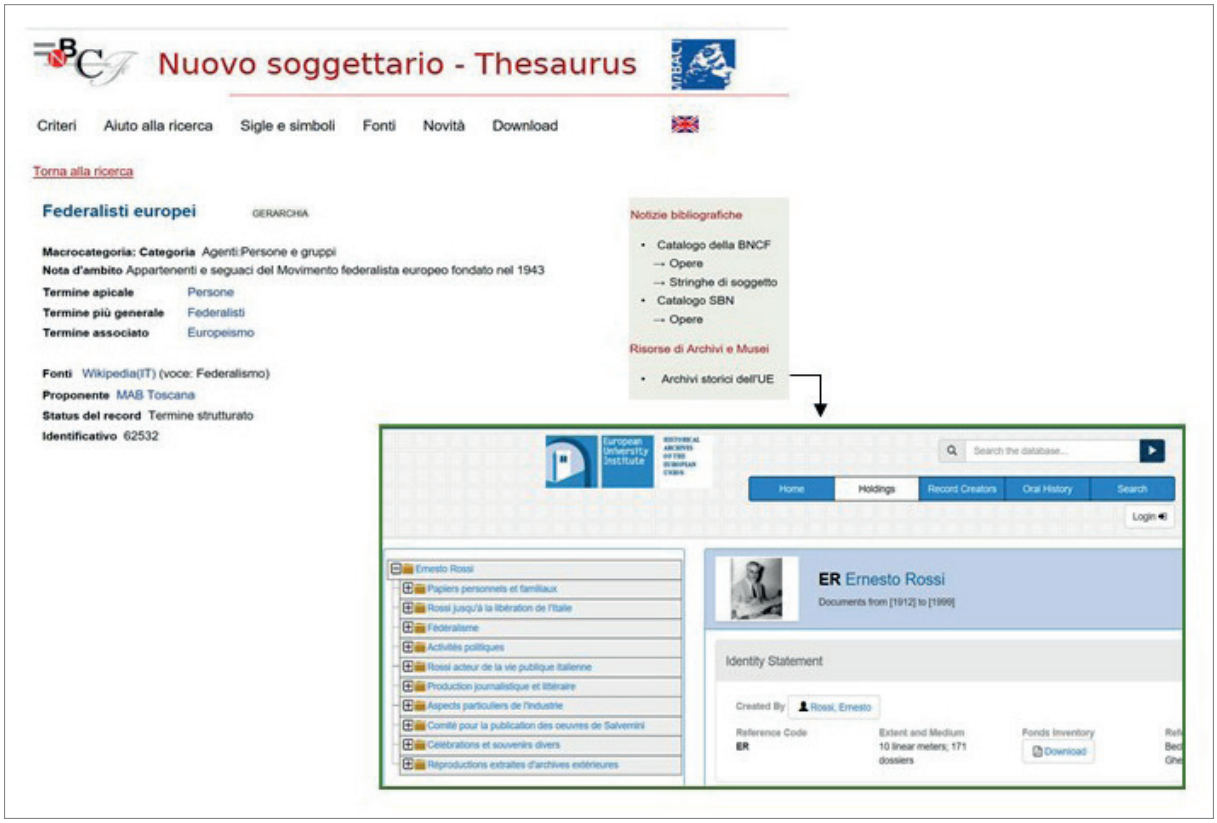

Fig. 14. Links between *Nuovo soggettario* and the Historical Archives of the European Union

These examples of the GLAM (Galleries, Libraries, Archives and Museums) perspective are also promoted by the Wikipedia universe,[40] and in Italy by the MAB (Musei Archivi Biblioteche) projects in which BNCF has participated while joining various research efforts.[41] Likewise, it is hoped there will be future possible connections, for example, with the controlled vocabularies of the Sistema Archivistico Nazionale (SAN)[42] or collaborations with institutions dealing with the standardization of the terminology employed for the cataloguing of the cultural heritage, such as the Istituto Centrale per il Catalogo e la Documentazione (ICCD) (Birrozzi et al. 2020).

[39] https://thes.bncf.firenze.sbn.it/termine.php?id=62532; https://archives.eui.eu/en/fonds/115005?item=ER. About the experimentation on subject indexing of Ernesto Rossi Fund: Becherucci et al. 2019, 24-48.

[40] For example, see the recent Gruppo Wikidata per Musei, Archivi e Biblioteche, https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche.

[41] https://www.aib.it/attivita/mab-italia/.

[42] http://san.beniculturali.it/web/san/home;jsessionid=66BD6878BF6E20807ACABB005C45C7CE.sanapp01_portal.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

The perspectives of machine learning, artificial intelligence, automated subject indexing

In which directions will the future of the *Nuovo soggettario* go? Its challenges are not that different from those of other thesauri.

Within the current context, which has much changed due to the predominant role of the Internet, subject indexing is interacting with the semantic capabilities of search engines, such as Google, with the development of both artificial intelligence and machine learning and, of course, with the dissemination of the ever increasing number of digital resources.

At the same time, we know that it is wrong to assume that sources transmitting information be only those 'hooked' by Google, just as it is wrong to confuse the functions of our catalogues with those of other tools for access to information.

However, as often pointed out by the indexing experts, such as Stella Dextre, one must also be aware that libraries and other institutions, dealing with information retrieval, have much fewer resources to be earmarked for 'manual' subject indexing, that is 'intellectual' indexing, as compared to Google's algorithms.

Despite the presence of search engines and their powerful automatic and semi-automatic indexing, the role of thesauri does not seem to be outdated.

For instance, Birger Hjørland, professor at the Royal School of Library and Information Science of Copenhagen, has very recently questioned about the reasons why the search engines, despite they apply principles of semantic type, do not make knowledge organization (KO) and mapping of the relationships among concepts superfluous at all (Hjørland 2021).

'Human' taxonomists working for Google, support the well-known Google Knowledge Graph, which is connected with DBpedia, Wikidata and with the linked data. This is a project about which very many reservations have been expressed.[43]

The procedures for automated and semi-automated translation/indexing are dealt with within IFLA[44] but also within countless other frameworks; to give some examples, in Italy these procedures are studied at the Istituto di linguistica computazionale di Pisa, at the Universities of Padua and Udine. In 2011, BNCF took its first steps by starting a project for the semi-automated indexing of digital doctoral theses. At that time, we used MAUI and other open source software. Should we have the resources and the possibility to restart this project, we could build on the important experiences of other national libraries, such as the Deutsche Nationalbibliothek or utilize tools like those implemented by National Library of Finland.

Studies will continue on machine learning, knowledge graphs like Google's, *corpora* of terms, and the benefits that thesauri can bring to our users, because not only the artificial intelligence world will benefit from such insights, but also libraries and the national bibliograpies world in their mission for the dissemination of knowledge.

In closing, here are some Keywords for the future of thesauri and for their challenges: creativity, versatility, sharing.

A special thank goes to Barbara Tillett who sent me many comments and suggestions.

[43] https://en.wikipedia.org/wiki/Google_Knowledge_Graph.

[44] Automated subject analysis and access Working Group, https://www.ifla.org/node/92551.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

References

(Last consultation of the websites: 15 July 2021).

Ballestra, Laura. 2011. "Information literacy education in Italian libraries: evidence from an Italian University." Bibliothek Forschung und Praxis 35, no. 3 (December):395-401.

Becherucci, Andrea, Silvia Bruni, Benedetta Calonaci, Emilio Capannelli, Walter Fochesato, Anna Lucarelli, and Sonia Puccetti. 2019. "Libri per gli internati militari italiani durante la Seconda guerra mondiale: un inedito di Ernesto Rossi." Biblioteche oggi 37, (May):4-48. DOI: http://dx.doi.org/10.3302/0392-8586-201904-024-1.

Bergamin, Giovanni. 2020. "Postfazione." In Guerrini, Mauro. 2020. Dalla catalogazione alla metadazione. Tracce di un percorso, 167-168. Roma: Associazione italiana biblioteche.

Biagetti, Maria Teresa. 2018. "A comparative analysis and evaluation of bibliographic ontologies." In Challenges and opportunities for knowledge organization in the digital age. Proceedings of the Fifteenth International ISKO Conference 9-11 July 2018 Porto, Portugal, edited by Fernanda Ribeiro, Maria Elisa Cerveira, 501-510. Baden Baden: Ergon.

Biagetti, Maria Teresa. 2020. "Ontologies (as knowledge organization systems)." In ISKO Encyclopedia of Knowledge Organization, edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/ontologies.

Birrozzi, Carlo, Barbara Barbaro, Maria Letizia Mancinelli, Antonella Negri, Elena Plances, and Chiara Veninata. 2020. "Catalogare nel 2020. La digitalizzazione del patrimonio culturale." Aedon. Rivista di arti e diritto on line no. 3. http://www.aedon.mulino.it/archivio/2020/3/birrozzi.htm.

Broughton, Vanda. 2020. "General principles underlying knowledge organization systems (KOS)." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4025408

Buizza, Pino. 2020. "Thesaurus and heading lists: equivalence and asymmetry." In Knowledge Organization at the Interface. Proceedings of the Sixteenth International ISKO Conference, 2020 Aalborg, Denmark, herausgegeben von International Society for Knowledge Organization (ISKO), prof. Marianne Lykke, prof. Tanja Svarre, prof. Mette Skov, Daniel Martinez Avila, [59]-68. Baden-Baden: Ergon.

Cheti, Alberto, Anna Lucarelli, and Federica Paradisi. 2009. "Subject indexing in Italy: recent advances and future perspectives.". https://www.ifla.org/past-wlic/2009/200-lucarelli-en.pdf.

Clarke, Stella Dextre. 2020 "How should today's thesaurus earn its keep?." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4048801

Clarke, Stella Dextre. 2020 "What is a thesaurus? How and Why so?." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4048800

Crociani, Laura, Maria Chiara Giunti, and Elisabetta Viti. 2016. "Trent'anni di Dewey in Italia: il ruolo della Biblioteca nazionale centrale di Firenze e i nuovi sviluppi sul fronte dell'interoperabilità con altri strumenti di indicizzazione semantica." AIB studi 56, no. 1 (January/April):87-101. DOI: https://doi.org/10.2426/aibstudi-11408.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Folino, Antonietta and Francesca Parisi. 2020. "Rappresentatività e copertura semantica dei KOS." AIDAinformazioni 38, no. 3/4:93-112.

Francioni, Elisabetta and Anna Lucarelli. 2020. "Nuovi concetti, nuovi termini ai tempi del Coronavirus." Bibelot: notizie dalle biblioteche toscane 26, no. 1 (January/April). https://riviste.aib.it/index.php/bibelot/article/view/12038.

Gnoli, Claudio. 2020. Introduction to Knowledge Organization. London: Facet Publishing.

Guerrini, Mauro. 2020. Dalla catalogazione alla metadatazione. Tracce di un percorso; prefazione di Barbara B. Tillet; postfazione di Giovanni Bergamin. Roma: Associazione italiana biblioteche.

Hjørland, Birger. 2021. "Search engines and Knowledge Organization (or why we still need Knowledge Organization).". KO-ED Theoretical Perspectives. https://www.iskouk.org/event-4058726.

L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique, sous la direction d'Etienne Cavalié. 2020. https://www.bnf.fr/sites/default/files/2020-03/biblio%20indexation%20matiere%2011mars20.pdf.

Junger, Ulrike. 2018. "Automation first – the subject cataloguing policy of the Deutsche Nationalbibliothek.". http://library.ifla.org/2213/1/115-junger-en.pdf.

Lucarelli, Anna. 2014. "«Wikipedia loves libraries»: in Italia è un amore corrisposto..." AIB studi 54, no. 2/3 (May/December). DOI: https://doi.org/10.2426/aibstudi-10108.

Magrini, Sabina. 2021. "«Ti racconto in italiano»: management, description and indexing of oral sources. A project by the ICBSA (Istituto Centrale per I Beni Sonori e Audiovisivi)." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=10942&end=12772.

Martínez-González, M. Mercedes, and María Luisa Alvite Díez. 2019. "Thesauri and semantic web: discussion of the evolution of thesauri toward their integration with the semantic web." IEEE Access, 7. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8873649.

Mödden, Elisabeth. 2021. "Artificial intelligence, machine learning and DDC Short Numbers." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=124&end=1523.

Petrucciani, Alberto. 2019. "C'è un futuro per l'indicizzazione?." In Viaggi a bordo di una parola. Scritti sull'indicizzazione semantica in onore di Alberto Cheti, a cura di Anna Lucarelli, Alberto Petrucciani, Elisabetta Viti; presentazione di Rosa Maiello, 163-173. Roma: Associazione italiana biblioteche.

Pocci, Adele. 2020. "Bibliotheca perspectivae: una sperimentazione del Nuovo soggettario nell'ambito specialistico dell'iconografia scientifica." Bibelot: notizie dalle biblioteche toscane 26, no. 3 (September/December). https://riviste.aib.it/index.php/bibelot/article/view/12798.

Riva, Pat. 2021. "The multilingual challenge in bibliographic description and access." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi72E1T4?start=12826&end=14355.

Smith-Yoshimura, Karen. 2020. Transitioning to the Next Generation of Metadata. Dublin, OH: OCLC Research. https://doi.org/10.25333/rqgd-b343.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12744

Suominen, Osma. 2021. "Annif and Finto AI: developing and implementing automated subject indexing." In Conference BC 2021. Video. https://www.youtube.com/embed/Yo6Vi-72E1T4?start=1892&end=2953.

Viti, Elisabetta. 2017. "My First Ten Years: Nuovo soggettario growing, development and integration with other Knowledge Organization Systems." Knowledge Organization 44, 8:624-637.

Will, Leonard. 2020. "From concepts to knowledge organization systems." In *KO-ED Introduction to Knowledge Organization*. https://www.iskouk.org/event-4043820

Zeng, Marcia Lei. 2019. "Interoperability." Knowledge Organization 46, 2:122-146. https://doi.org/10.5771/0943-7444-2019-2-122.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

JLIS.it

How to build an «Identifiers' policy»: the BnF use case

Vincent Boulet[(a)]

a) Bibliothèque nationale de France

Contact: Vincent Boulet, vincent.boulet@bnf.fr
Received: 15 July 2021; **Accepted:** 12 September 2021; **First Published:** 15 January 2022

ABSTRACT

Identifiers are at the crossroads of two interconnected, major evolutions which heavily impact national libraries: the massification of dataflow, redrawing the place libraries occupy within the global and national data ecosystem in a shared environment, and the strategic shift towards entity management underlying behind the new professional practices and standards. Based on the experience and maturation libraries are gaining in this field, the time maybe has come to formalize them and to highlight the impressive strike force libraries could have in a highly competitive landscape. This is the aim the Bibliothèque nationale de France is trying to reach by publishing an identifiers' policy. It comes as the last part of a triptych after the new cataloguing policy (2016, including the indexing policy published in 2017) and the quality policy (2019). This identifiers' policy is intended to clarify why and on what grounds a national library could, more or less, get involved in a given identifier, taking into account the diversity of scope, governance structure and business model of identifiers, be they international (for instance: ISNI, ISSN, ARK) or local (for instance: the BnF proper identifiers). Therefore, the identifiers' policy highlights why it is necessary to use permanent, trustworthy identifiers and to what extent they are helpful in the daily working and quality control processes led by cataloguers. This is why the identifiers' policy is not limited to principles, but has a very concrete dimension, both for internal and external issues.

KEYWORDS

Identifiers; ISNI; BnF.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

Why an "identifiers' policy" now ?

Libraries' web presence now makes them familiar with identifiers and their uses. This presence poses for them several major and well-known challenges. We can summarize them as follows:

1. The adaptation of their system and data model to the requirements of research and "findability" of their resources *in* the Web. This global framework implies and fuels a needed, major shift of the data structuring, from a world where libraries used to standardize records for making them exchangeable into a world where libraries, along with other players, have to structure data for making them sharable. This issue is at the heart of the crucial problematic of the future of the bibliographic control and has many, crucial implications. For instance, the division of the bibliographic world into bibliographic records and authority records is now close to an end. Therefore the emerging international standards go with the flow, be it the IFLA-LRM data model published by IFLA in 2017 or the new version RDA reshaped by the "3-R project" which became the official version of the RDA international cataloguing code last December. Both have the same underlying principle, namely an entity/relations-based overall model. It means, from the authority control point of view, to switch to logic based on entity management.
2. Resources in a digital world are increasingly more agile and more scalable, due to changes in research and uses' practices. Have we to describe serials or articles published on several platforms? Have to describe coherent set of musical works or a given piece of music diffused by various platforms under various formats? That issue has major implications on legal deposit for digital sound, books and movies. This complex reality challenges the new, above-mentioned library models and cataloguing codes, as they have to take into account changing resources which do not necessary feel part of any idealistic pyramidal model. What is recorded now should not be considered as permanent.
3. The data flows are becoming more and more massive, as the metadata accompanying them. This is also a challenge both for the bibliographic control and for the consistency of library databases. It actually raises the question of how applicable cataloguing rules are for the whole data set libraries deal with. Here is the issue of quality control processes and quality policy, because quality processes can be applied differently according to different data sources and subsets. This makes the question of sourcing data crucial, both for data flows reused by libraries, and for data flows libraries disseminate to end-users.
4. The technical and legal opening of datasets and catalogues is one of the points to be considered for having really sharable data. It may also be a political, strategic commitment taken by public administration towards citizens. As far as the legal opening is concerned, it may also put on the table the issue of mentioning the source of the data and keeping it associated with the metadata produced by a given player.

All these challenges are well-known for the future of the bibliographic control and we have to draw consequences from them. The entity management is unthinkable and impossible without any identifier management and identifiers' policy. The shift from labels (different forms of a name for a person for instance) to identifiers provides less ambiguous data and a kind of stability. Identifiers allow access points or labels to be treated as entities being differently usable according to

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

context and needs : this is the key-principle of the "nomen" entity in IFLA-LRM. This shift also improves interoperability of data in regards with different contexts[1].

Beyond these principles and opportunities, libraries have nowadays to deal with a very scattered landscape, due to the wide variety in nature offered by identifiers they are using or have intention to use. We can distinguish:

- The global identifiers supported by an ISO standard, which ISO signs an agreement with an international agency about. They correspond to a specific business model and global governance, whose libraries are a part of, along with other players, like music and cultural industry or copyright management firms. Libraries take part in a global business and scientific framework deciding on attribution and possible uses of a given identifier, and they can act as basic members, or a registration center for a given community or a specific field. This is, for instance, the case for ISNI (ISO standard 27729:2012), ISSN (ISO standard 3297), and ISAN (ISO standard 15706-2). For instance, BnF hosts the French national ISSN center and has an official, nationalwide responsibility on this identifier. BnF is, furthermore, an ISNI registration agency since 2014 for a specific dataset, corresponding to the scope of the French legal deposit and national bibliography. But BnF has no special responsibility on ISBN.
- The global identifiers which could be assimilated to a de facto standard, or are being engaged in a standardizing process, and which are supported by an users' community. For instance, ARK (*Archival Resource Key*), an identifier created by *California Digital Library* (CDL), intending for identifying all resources, both physical or digital, records from catalogues or even immaterial resources as concepts. ARK is based on some key-principles and on a community of players engaged to maintain them ("naming authorities", being able to attribute ARK to their resources, and "addressing authorities", being able to resolve the identifier in order to give through it access to resources, by applying a policy of permanence). Moreover, the ARK identifiers have an explicit structure, which make them a de facto standard. So, about ARK, BnF respects an engagement framework with an users' community.
- The specific identifiers BnF has itself set up and is maintaining for internal uses and management of its databases, as for instance internal numbers of bibliographic and authority records (for instance: FRBNF identifiers). But external players can reuse them when reusing these records. So, even if these identifiers have been designed for internal uses at the time of catalogues' automatization, they are also de facto external. BnF keeps the complete control on their maintenance.

So, this short review shows that, over time, successive projects and needs, identifiers have been piled up one on another. In the same time, we have been gaining gradually more maturity and more experience on the overall identifiers' issue.

Managing identifiers doesn't fall from the Jabal Musa as Ten Commandments, but is highly de-

[1] Gordon Dunsire and Mirna Wilner, "Authority versus authenticity: the shift from labels to identifiers". In: Authority, provenance, authenticity, evidence: selected papers from the conference and school Authority, provenance, authenticity, evidence, Zadar, Croatia, October 2016. Edited by Mirna Willer, Anne J. Gilliland and Marijana Tomić. Zadar : Sveučilište u Zadru, 2018. p. 87-113.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

pending on human and IT resources, on transparency in how these identifiers are managed and on what libraries intend to do with them. Nevertheless, the way of dealing with identifiers as a whole, of choosing them, of handling with them must be consistent with best practices given from a global perspective. We have already framework documents for them, endorsed by W3C[2], by IFLA[3] or by other international authoritative bodies. But, the question, for a given library, could be raised from another perspective. From the point of view of a given institution, to what extent using and disseminating identifiers can be helpful for addressing its own role and tasks? What criteria can be used strategically to justify the commitment of the library in one or more identifiers, and, possibly, its non-involvement? Here is the aim of an identifiers' policy.

Identifiers: a commitment story

Using identifiers highly depends on how committed or engaged libraries want to be. We can easily assume an activist aspect for conceiving and implementing policies. In its identifiers' policy, BnF defines the idea of « engagement » as following:

- For a given and explicit dataset, BnF integrates identifiers in its dataflow and in its development policy regarding metadata (for instance : ARK for every resource, and ISNI for "agent" entities). This is why the identifiers' policy is a follow-up of the BnF quality policy. Identifiers are a tool to delineate specific data subset on which a specific quality control can be applied. It is also helpful to automatize some data processing, by helping interconnections of data. For instance, one of the projects BnF ISNI registration agency is developing is to propose alignments between EAN and ISNI so as to help cataloguers to create links between bibliographic and authority records (and, tomorrow, between manifestations, works and agents).
- BnF ensures, through identifiers, persistence of accessibility to its resources, in a broader meaning of the word: physical resources, digital resources (both digital version of physical documents, and natively digital resources), metadata describing and identifying resources. Identifiers ensure how trustworthy resources are identified for end-users.
- BnF builds up for end-users specific services and transactions thanks to identifiers, being based on its status of national bibliographic agency. For example, the BnF ISNI registration agency has built some transactions with the French book supply chain to register and disseminate ISNIs for their authors.
- BnF disseminates identifiers and resources for free, thanks to legal and technical opening. From this regard, the idenfiers' policy is a follow-up of the open data policy BnF has set up as early as 2011 for data.bnf.fr and as 2014 for every resource.

In other words, the identifiers' policy ensures: to have easily disseminated resources, for the broadest communities, to have traceable, linkable, visible and discoverable resources.
This is the reason why the identifiers' policy is at the crossroads of the strategic shift made by BnF

[2] Data on the Web Best Practices, W3C recommendation, 31st January 2017 (https://www.w3.org/TR/2017/REC-dwbp-20170131/)
[3] Best Practice for National Bibliographic Agencies in a Digital Age, https://www.ifla.org/FR/node/8786

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

under the name of « bibliographic transition »[4], and embodied by several strategic documents : the statement on open data (2014), the cataloguing policy (2017)[5], the indexing policy (2018)[6] and the quality policy (2019)[7]. A global metadata policy is under preparation and should be published this year.

Negotiating tensions

An identifiers' policy has to deal with three major tensions.

The first tension is the relationship between principles and concrete work and data libraries have to handle with. An identifiers' policy should be intended to give a general framework to action and to concrete involvement on identifiers, both internally, by integrating the identifiers management to the concrete dataflows and cataloguers' work, and externally, for end-users. The question is not to add even more practices for a given identifier, but give practices global framework and direction. In other words, an identifiers' policy finds its role somewhere between a statement of principles on the one hand, and concrete practices and using in the other hand.

The second tension regards the relationship between a common policy and the diversity of identifiers, as said above. It means handling with the diversity of identifiers themselves, and the diversity of how libraries can exercise some responsibility on them. Libraries can only use identifiers in their dataflows, without any significant role ; or they can attribute them ; or they can maintain alignments, or they can build up services for third parties, for instance for the library national community, or the book supply chain.

Here are, for instance, the different role BnF exercises, or intends to exercise on identifiers.

BnF role	**International ISO identifiers**	**Identifiers with an international audience**	**Local identifiers**
Attribution or registration responsability	ISSN, ISNI	ARK	FRBNF
Identifiers BnF doesn't attribute, but BnF uses and builds services for the community on.	ISBN	EAN	
Identifiers which BnF develops alignments with		LCSH, MESH, GND, datos.bne.es, VIAF, NOMISNA, Geonames, Agrovoc, Wikidata	
Identifiers integrated in dataflows	ISAN	EIDR	

[4] For more details on the « Bibliographic transition » national programme, see : https://www.transition-bibliographique.fr/enjeux/bibliographic-transition-in-france/

[5] https://www.bnf.fr/fr/politique-de-catalogage-dans-bnf-catalogue-general

[6] https://www.bnf.fr/fr/politique-dindexation

[7] https://www.bnf.fr/fr/politique-de-qualite-des-donnees

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

We should distinguish "registration" from "attribution". "Attribution" means that library attributes directly a given identifier, following international policies and rules. This is the case for ISSN, through ISSN French Centre, which *attributes* ISSN identifiers following rules and policies validated by the ISSN International Centre and ISSN international network. "Registration" means sending data for asking attribution to international authoritative body. For instance, BnF *registers* ISNI by sending authority records for names of persons and the bibliographic records linked to them to the ISNI International Attribution Agency, by getting back ISNIs attributed on its own data by this attribution agency, and by disseminating ISNIs through the book supply chain and the library community in France.
The third tension regards the necessity to keep a two-fold diachronic, dynamic approach. On the one hand, the international landscape of identifiers is moving. On the second one, the responsibility libraries can take on one given identifier can move, too. For instance, BnF is thinking about taking more responsibility on ISAN, ISWC and ISRC, depending on their business model, legal structure, on the one hand, and on resources BnF can invest on them, on the other hand.
Therefore, setting up an identifiers' policy means to declare principles, on which BnF can commit itself, by taking into accounts these tensions, and concrete conditions allowing such a commitment by a State and non-for-profit institution to be concretely achieved.

The policy content

The key-principle is permanence. The identifier shall give guarantees on permanence, which concretely implies for it to be based on shared, transparent governance, broad and, if possible, global community, sustainable business model, as for the identifier itself, as for the community using it, and a standardizing process. All these elements create trust in the opportunity of consuming human and financial resources to integrate the identifier in the library dataflows and in the services and engagement the library agrees on with other players.
We have also formulated four main conditions to make these principles concretely applied.

1. The identifiers must benefit from a broad and stable community or inter-community commitment It implies that the identifier is part of a normative strategy:
   - either because it corresponds to an ISO standard (for example: ISO 27729: 2012 for the ISNI identifier; ISO 15706: 2002 and ISO 15706-2 for the ISAN identifier; ISO 3297 for the ISSN) and undergoes the international consultation process applied to periodically revised ISO standards;
   - or because it is part of a strategic standardizing process (for example: ARK[8])

   The identifier must therefore benefit from support of an international community or of a cross-domain commitment, depending on its scope of use. Its use must also be recognized and promoted by one or more communities. The governance of the identifier, whether at a national or international level, must be based on a written contract and allow the community or communities to be represented in decision-making bodies and to contribute to the technical and strategic orientations of the identifier.

[8] See above

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

The identifier must be based on a negotiated, transparent, stable, contractual and sustainable economic model for a public institution. Business model should enable the BnF to develop a medium and long-term policy of use and services for the communities it serves. It must also be balanced in order to provide guarantees of financial stability in the medium term. This is the case, for example, for the ISNI business model, which allows libraries overall business model of this identifier.

2. The identifier must have a clear and explicit application policy, in other terms, we must clearly know what does identify the identifier. For instance, we know to what entity ISNI is applied for, as described in the ISO corresponding standard, which put forward the concept of "public identity", more or less similar to the concept of "bibliographic identity" libraries are familiar with.
   The identifier must respect the principle of uniqueness. An identifier relates to one and only one resource. When a resource is stable, so is the identifier. When a resource changes to become something else, a new identifier must be assigned. Similarity and duplication issues need to be identified and addressed. For ARK, BnF is developing practices of redirection when merging two duplicates, for instance. The question is more sensible for concepts and remains under discussion for now, because two concepts are never exactly similar.
   The identifier data model must be defined, documented and transparent. The attribution policy and the scope of data and resources to which the identifier applies must be stable, unambiguous and explicit. The conditions for attributing the identifier must be clear and explicit so as to control the mechanism and scope of their attribution, as well as their non-reassignment. This is why BnF has made explicit the scope of ARK, and has recently extend it to records for archives and manuscripts, so as to make every BnF resource covered by this identifier, without any regard to the data base describing it.
3. The identifier must be technically sustainable. The ID is built to last.
   That means:
   - It must be independent from the technical protocols to ensure its attribution and management, as well as of the authority that technically ensures its attribution. The guarantees of technical sustainability must be made explicit in the contractual commitments binding the national or international governance body on the one hand and the BnF on the other. This is the case for ISNI, for instance.
   - The link between the identifier and the resource described must be permanent. The existence of the identified resource must be certified. We are developing the scope of the future French National Entity file (FNE), to be published around 2024, in this direction. The entity, and the identifiers associated to this must correspond to a real resource belonging to a member of the FNE network.
   - An identifier must be maintained during and beyond the life of the resource that it identifies. If the resource or entity evolves, the persistent identifier must ensure a redirection to the most recent version of the resource or of the description of the entity to which it returns. The user must be informed of any significant change in the identified resource: deletions, replacements, merges, substantial modifications of the scope of the resource. The memory of the assignment of the identifier must thus be preserved.
   - An identifier is never and under no circumstances reassigned.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12768

- In accordance with W3C best practices, it is better for an identifier to be expressed as an URI, as allowed by ARK, for instance.

4. The identifier must be open and neutral politically and technically.
   This means :
   - The identifier must be administered by an independent body contributing to the neutrality and uniqueness of the Web. It does not depend on exclusive mercantile interests that unilaterally could impose objectives, governance and an economic model incompatible with the requirements of a public institution. Dedicated and trained teams follow the attribution and registration procedures. This is the case with the ISNI governance structure and Quality Team.
   - The BnF favors identifiers that are opaque in their meaning in order to avoid the temptation to modify them if the resource or entity they identify changes and to allow their widest distribution.

Conclusion: audience and next steps

The identifiers' policy is intended to have both an internal and external audience. It aims at explaining cataloguers' and librarians the main directions BnF is implementing, and at committing BnF in its coming discussions with end-users and management bodies of identifiers. The next steps are to concretely develop this policy for the identifiers already used in the workflow.
An identifiers' policy shows how important identifiers are for the future of bibliographic control, by accelerating and making consistent the overall shift of data structure towards entity management. We could say it is both a tool for managing this shift and the aim this shift is supposed to achieve, because it is a tool to redraw the library role and place in the global data ecosystem. It supposes not to have a defensive approach but to elaborate strategic orientations for making libraries not a customer or a victim, but a genuine player in this shift.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

JLIS.it

The International Standard Name Identifier: extending identity management across the global metadata supply chain

Andrew MacEwan[(a)]

a) The British Library

Contact: Andrew MacEwan, andrew.macewan@bl.uk
Received: 12 April 2021; **Accepted:** 3 June 2021; **First Published:** 15 January 2022

ABSTRACT

This article describes how ISNI is being adopted as a common identifier across disparate sectors of publishing. Whether publishing and distributing recorded music, film or text ISNI is making good identity management a staple element in the global metadata supply chain. As the content creation industries become more engaged with the value of embedding good metadata from the point of publication libraries can look forward to benefitting from a truly global revolution in the metadata supply flow. A case study describes how a British Library project has taken ISNIs already in the British National Bibliography and cross-matched them with data from UK publishers' own databases to embed ISNIs into the book supply chain. It also describes plans for ongoing publisher engagement through implementation of ISNI assignment into its cataloguing-in-publication workflows for UK legal deposit.

KEYWORDS

Authority control; Identity management; Identifiers; Names.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

Introduction

According to the standard ISO 27729 the International Standard Name Identifier was originally conceived as a "bridge identifier" with the ambition that it would be used for the identification of public identities of parties involved throughout the media content industries in the creation, production, management, and content distribution chains. This paper provides a brief update on how this ambition is beginning to be realised through the growth in adoption of ISNI in different publishing supply chains. Whilst this is important for the growing utility of ISNI in breaking down metadata silos in relation to efficient name identification it is also important to contextualise this as part of a broader trend that is seeing the business of producing well-controlled metadata become part of the business of publishing in the age of digital supply and demand. This paper, however, will focus on ISNI as an exemplar of this trend and will report in particular on a British Library case study describing our engagement with a group of UK book publishers and other agencies to embed ISNIs in the book supply chain.

Metadata silos and the supply chains

Different forms of creative content are distributed in supply chain metadata silos specific to each content type. The standards followed in each supply chain are well documented on websites promoting their use. Text publishing is supported by metadata supplied in the ONIX schema, with enhanced subject access through THEMA subject codes and additional product control provided in the form of trade identifiers: ISBN, ISSN, EAN barcodes, DOI, etc., as described at the EDItEUR website (EDItEUR, n.d). The music industry mirrors this with the DDEX schema standard, underpinned by the use of identifiers to express products at varying levels of granularity: ISWC, ISRC, RIN, RDR, etc. all described at the DDEX website. Metadata standards for the film industry are described most comprehensively at the website for the Entertainment Industry Identifier Registry (EIDR, n.d). Library standards have the advantage of attempting to accommodate and describe different content types in common standards, but even so libraries too have also worked in their disconnected silos reflecting historical divisions in curation of different content types. At the British Library our Sound Archive, our general catalogue, and our manuscripts and archives are catalogued in separate databases that reflect the major differences in the types of content and the standards that we use to describe them.

Library metadata itself exists in a silo in the context of the global supply chains. We rely on crosswalks and mappings, such as ONIX to MARC, to re-use data from the supply chain in our library based schemas. We also rely heavily on industry standard identifiers like the ISBN and the ISSN to build efficient automated workflows that allow machine matching based data enhancements from multiple sources. Co-operative cataloguing, the efficient re-use and sharing of metadata between libraries, where possible via automated workflows, is a staple activity fundamental to the efficient realisation of bibliographic control in the library world. In recent years the same theme of better metadata standards to support efficiency, automation and re-use have become a hot topic in every commercial supply chain in the publishing world. There is interest both in improving end-to-end metadata supply chains within each content industry and in building crosswalks between supply chains where appropriate commonality exists. In a Whitepaper on identifiers for artists

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

(Movielabs, 2019), the company Movielabs reviewed existing approaches to name identification such as VIAF, ORCID and ISNI as potential models for managing identities for the film industry. The paper notes that the widespread adoption of ISNI in the music industry is a factor recommending ISNI adoption in the film industry, given high levels of commonality linking the sectors, rather than pursuing invention of another name identification standard.

An early example of building better metadata solutions around commonality was the collaboration on the "RDA/ONIX Framework for Resource Categorisation" (JSC-AACR, 2006) that connected the work of the revision of the Anglo-American Cataloguing Rules with the development of the ONIX standard in the publishing industry. In 2014 the Linked Content Coalition published a paper, "Principles of Identification" (Paskin & Rust, 2014) that highlighted the content neutral potential of ISNI as a name identifier that could be used across multiple supply chains. Most recently the UK standards body, Book Industry Communications, has launched a Metadata Capability Directory (Matthews, 2020) to promote and improve the use of metadata standards in the end-to-end text publishing supply chain. The Directory is intended to be a platform where the use of standards across the supply chain can be compared, deficiencies and opportunities identified, and collaboration on solutions initiated. In the music industry the by-line on the DDEX website perhaps best summarises the conversations and initiatives that are taking place in every supply chain: "DDEX is a standards setting organisation focused on the creation of digital value chain standards to make the exchange of data and information across the music industry more efficient." (DDEX, 2021)

This brief outline of the wider supply chain serves to highlight ISNIs place in the digital ecosystem of the global supply chains, but it also serves as a reminder that library metadata exists in the context of those supply chains and has the potential to benefit from the growing commercial interest in making metadata work better.

ISNI's place in the supply chain

The focus of the rest of this paper is on ISNI as a specific exemplar of a content neutral standard for name disambiguation that is starting to fulfil its purpose as a bridge identifier across sector specific silos for metadata. The foundation of ISNI in library metadata means that it already provides identification for authors, musicians, actors, editors, producers, artists and supports identification of both individuals and groups or organisations. In recent years, adoption has been strongest in the library sector and the music sector, with building blocks in place to encourage more widespread use in the book supply chain. ISNI's ability to work across so many specialist domains is based on a hub and spoke model in which Registration Agencies and Members provide sector expertise but work with a common database in the ISNI Assignment System, maintained by OCLC.

ISNI at work in the supply chain

In the music industry the ISNI membership list is growing. YouTube, Apple, Spotify and both major and minor record labels are set to be users of ISNIs and a growing network of music metadata

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

organisations specializing in rights, credit and attribution of content to artists and performers are providing the engine rooms for the supply of ISNIs to the music industry. Currently listed on the ISNI website from the music sector (alongside YouTube, Apple and Spotify) are SoundExchange, Quansic, Qanawat, Consolidated Independent, Jaxsta, @Musiekweb, Muso.AI, The ISRC Team and Soundways. (ISNI, n.d.) The last of these, Soundways, is a sound engineering company that has built an ISNI Registration Service within its Sound Credit system. Soundways describe it themselves on the ISNI website: "Sound Credit's ISNI registration system is part of its larger system for music crediting, using Sound Credit's new massive cloud profile feature. Once music creators and engineers set up a free profile, they can be instantly credited simply by entering an email address, swiping a card at a kiosk, or selecting their profile in an app. Any credited profile in Sound Credit will automatically attribute their ISNI code to every project involving that creator, along with other identifier codes such as the IPI/CAE or IPN that users can optionally enter" (Sound Credit, 2020). The interface with the ISNI central database emphasizes search and entering rich metadata to ensure that each ISNI is unique in the central database whilst local control of identities is maintained in the Sound Credit system itself.

An example of similar intention in the book publishing industry came in January 2020, when the Frankfurt-based technology and information provider MVB took on the role of an ISNI RAG operating in Germany, Austria and Switzerland. The first step will be to assign automatically an ISNI to all creators listed in the Verzeichnis Lieferbarer Bücher (VLB), the books-in-print catalogue used in the German-speaking world. In a second step, publishers whose books are listed in the VLB will be able to register new ISNIs for the creators of their works – directly from the catalogue, and free of charge. (MVB, 2020)

The British Library and ISNI

The British Library has a long standing involvement with ISNI from being a member of the ISO 27729 International Standard Name Identifier Committee to draft the standard to becoming one of the Founding Members of ISNI acting jointly with the Bibliothèque nationale de France to co-represent the Conference of European National Librarians (CENL) on the ISNI Board. Working with the Bibliothèque nationale and OCLC we supported the foundational work to build the initial ISNI database from VIAF and other data sources. The BL and the BnF have continued to provide quality assurance services to the ISNI International Agency for the ongoing maintenance of the ISNI database.

When the British Library became an ISNI Registration Agency in its own right this marked a strategic shift in our goals for authority control away from name disambiguation in the British National Bibliography (BNB) and in our catalogues towards bridging data silos and exploiting the potential of a numeric identifier to build and embed identity management into the supply chain. There are three guiding principles for our implementation of ISNI:

1. Embed ISNI in all our cataloguing workflows
2. Automate processes as far as possible
3. Engage with the supply chain

Pursuing these principles involves overcoming significant challenges. The British Library's cata-

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

loguing workflows with regard to authority control use the LC/NACO file. We hold a complete mirror copy of the LC/NACO file in our Aleph cataloguing system and maintain currency with the other LC/NACO nodes through daily file exchanges. Integrating ISNI into our authority control workflows will require ISNIs to be uploaded into this LC/NACO shared resource. Conversations and planning for this to happen at scale are ongoing with the Library of Congress and the Program for Cooperative Cataloging, but it is evident that capturing and loading all the ISNIs already associated with NACO records within the ISNI database will take place over an extensive time period. In the meantime we have focused on getting ISNIs into our legacy bibliographic data and engaging with the UK publishing supply chain. Happily these two endeavours have worked in concert as will be described below.

A British Library case study in supply chain engagement.

Serious engagement with publishers and other actors in the UK supply chain was initiated in two facilitated meetings in early 2018. In January 2018 Publisher Licensing Services, an organization providing collective licensing and rights management services for the publishing sector, and an ISNI member organization, hosted a meeting for publishers to discuss the potential use of ISNI for improving identification of publishers and imprints in the supply chain. This discussion led to a follow up meeting in March hosted by Book Industry Communication to explore the wider topic of ISNI for authors, publishers and imprints. A colleague from the Bibliothèque nationale joined this meeting to give a presentation on their integration of ISNI into their cataloguing-in-publication workflows for French legal deposit. Thanks to further advocacy and promotion by EDItEUR the interest sparked by both these meetings led to the establishment of an informal UK Publishers Interest Group comprising the following organisations:

Bibliographic Data Services (BL's CIP subcontractor)
Book Industry Communication
British Library
Cambridge University Press
EDItEUR
Hachette UK
International ISBN Agency
Harper Collins
ISNI International Agency
Nielsen Book (UK ISBN Agency)
Pan Macmillan
Penguin/Random House
Publisher Licensing Services
Bloomsbury

Early on the group settled on a remit to explore practical solutions for disseminating ISNIs that were already established in the ISNI database into bibliographic product records that were already held in common by publishers and aggregators and the British National Bibliography. It was agreed that the quickest way to demonstrate value at scale and to introduce ISNIs into the supply

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

chain was to exploit what ISNI had already achieved in building its database of identifiers. Since the group as a whole had many different levels of capability for handling varieties of ONIX and MARC data it was also settled upon to make CSV files the medium of exchanging data between the British Library and the publishers themselves.

The starting point for the work was to get ISNIs into the British National Bibliography. Names in records in the BNB are the established name forms found in the LC/NACO file. We already had staff experienced in working with the Virtual International Authority File (VIAF) to associate VIAF and NACO IDs with the Linked Open Data version of the BNB. We also already had established links from ISBNs for product records, names in those records and LC/NACO IDs. By using the VIAF links we were able to pull across all ISNI-LC/NACO associations already established in VIAF clusters and bring the ISNIs back into the BNB. This provided us with a base file of 3,160,908 names in BNB records with assigned ISNIs for working with publishers' product data.

Each of the publishers in the working group provided us with sample files and later full back files as we developed the matching processes. Publishers provided us with a name string, their proprietary in house author ID, and its associated products. ISBNs were the key match point for identifying the target records and our staff developed algorithms to ensure we associated only confident matches between the LC/NACO name string and the publisher's name string to assign the corresponding ISNI. Differences between original publisher data and BNB catalogued data meant there were a variety of issues to work with: different name forms, punctuation and character set issues, reverse name forms, presence or absence of names for translators or illustrators, multiplicity of product ISBNs for the same work. The process was refined over time. Early results were quite variable between publishers and percentages of assignment relatively low in the first round of work. After several iterations and an expanded group of publishers' files to work with the latest results are as given in the table below.

Publisher	Number of names	Number of matches	Success rate
Atlantic	1,201	954	79%
Bloomsbury	43,558	28,420	65%
BurleighDodds	681	35	5%
ChannelView	1.392	1,146	82%
Canongate	521	363	70%
Cambridge University Press	21,298	16,292	76%
Dorling Kindersley	2,103	1,409	67%
Hachette	10,857	7,820	72%
Harper Collins	13,406	8,107	60%
Kogan Page	1,117	708	63%
Liverpool University Press	1,498	1,064	71%
PanMacmillan	1,642	1,332	81%
Penguin	14,297	9,638	67%

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

Publisher	Number of names	Number of matches	Success rate
Pluto	1,589	1,107	70%
Random House	24,060	16,127	67%
Taylor&Francis	107,871	68,878	64%
Total	**247,091**	**163,400**	**66%**

Fig. 1. Publishers' Author Name Matching Results

Generally, we have achieved a high level of consistency in the results and feedback from those publishers who have integrated the ISNIs into their own databases has confirmed the accuracy of the assignments from their side. An additional benefit that has come out of the work is cross deduplication of authors between publishers and in some instances deduplication within a publisher's own author file. The figures for deduplication are as given in the table below (Figure 2).

Publisher	Number of de-duplicated IDs (across all publishers)	Number of de-duplicated IDs (within publisher)
Atlantic	0	12
Bloomsbury	5409	619
BurleighDodds	0	0
ChannelView	0	0
Canongate	139	0
Cambridge University Press	3363	2225
Dorling Kindersley	398	62
Hachette	1940	746
Harper Collins	2350	119
Kogan Page	0	0
Liverpool University Press	0	4
PanMacmillan	524	33
Penguin	3449	846
Pluto	0	4
Random House	4386	1788
Taylor&Francis	8650	9916
Total	30608	16374

Fig. 2. Publishers' Authors Names Deduplication Results

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

Whilst the deduplication across publishers was an anticipated benefit of sharing a common supply chain author identifier, the cleanup of duplicates within a publisher's own data was an unexpected bonus, but one that demonstrated additional value in working across data silos. A further early bonus of this project with publisher data is the first example of a provided ISNI being re-used by Harper Collins in an ONIX record for a new publication by one of their authors. (Figure 3)

```
<TitleElement>
    <TitleElementLevel>01</TitleElementLevel>
    <NoPrefix/>
    <TitleWithoutPrefix textcase="02">Boy Giant</TitleWithoutPrefix>
    <Subtitle>Son of Gulliver</Subtitle>
</TitleElement>
<TitleStatement>Boy Giant: Son of Gulliver</TitleStatement>

<Contributor>
    <SequenceNumber>1</SequenceNumber>
    <ContributorRole>A01</ContributorRole>
    <NameIdentifier>
        <NameIDType>01</NameIDType>
        <IDTypeName>HCP UK Author ID</IDTypeName>
        <IDValue>4121</IDValue>
    </NameIdentifier>
    <NameIdentifier>
        <NameIDType>16</NameIDType>
        <IDValue>0000000121251907</IDValue>
    </NameIdentifier>
    <PersonName>Michael Morpurgo</PersonName>
    <PersonNameInverted>Morpurgo, Michael</PersonNameInverted>
    <NamesBeforeKey>Michael</NamesBeforeKey>
    <KeyNames>Morpurgo</KeyNames>
</Contributor>

<Contributor>
    <SequenceNumber>2</SequenceNumber>
    <ContributorRole>A12</ContributorRole>
    <NameIdentifier>
        <NameIDType>01</NameIDType>
        <IDTypeName>HCP UK Author ID</IDTypeName>
        <IDValue>1897</IDValue>
    </NameIdentifier>
    <NameIdentifier>
        <NameIDType>16</NameIDType>
        <IDValue>000000012147035X</IDValue>
    </NameIdentifier>
    <PersonName>Michael Foreman</PersonName>
    <PersonNameInverted>Foreman, Micha
        el</PersonNameInverted>
    <NamesBeforeKey>Michael</NamesBeforeKey>
    <KeyNames>Foreman</KeyNames>
</Contributor>
```

Fig. 3. Example ONIX record containing ISNIs

Future work with publishers

The above results reflect the work we have achieved so far but the UK Publishers Interest Group continues to meet and we have more work to do. Although we do not think we can achieve much more improvement in the match rates through further improvements to our matching processes there may be improvements to be gained via more direct work with the ISNI database itself. Although the ISNI database began its life with a series of regular uploads of relevant records from the full VIAF database the last of these took place in 2016. Since then ISNI has worked with direct authority file loads from the increasing numbers of national libraries who have joined the ranks of the ISNI membership. The British Library has recently completed work on preparing an update file from its own copy of the LC/NACO database from 2016 to the present for submission to the ISNI database to bring LC/NACO up to date in the ISNI assignment system. Where possible this was enriched by associating title and ISBN data with the LC/NACO records extracted from the BL's own catalogues and the LC Books All file to facilitate the matching and the rich record assignment processes in the ISNI Assignment System. Following this load the total number of assigned ISNIs associated with a LC/NACO identity stands at 5,553,823 persons and 602,288

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

organisations. The results from this update will be used to re-run and fill some of the gaps in the publishers' results.
Following the above step the dialogue with the publisher group will move onto another stage. The high assignment rates already achieved have already opened up a conversation around and an appetite for 100% ISNI coverage in publishers' data. There are several avenues to explore for achieving this. The work to date has been an experimental project with the goal of seeding ISNIs at scale into the databases at the beginning of the supply chain. It has also been a mutually beneficial project to both publishers and the British National Bibliography. If we have run out of automated means to populate both the BNB and the publishers' databases then one option to provide a more intensive, manual level of intervention to fill the gaps could be a priced service for the remaining legacy data, acting in our role as an ISNI Registration Agency.
The other unresolved question though is the provision of new ISNIs for future authors. We are working on two solutions for this. One is the provision of a Registration Service portal for individual ISNI assignment requests. The second is the integration of ISNI assignment into our CIP workflows for our Legal Deposit intake and the development of a feedback loop to publishers along the lines already pioneered by the Bibliothèque nationale. Since the integration option accords with our lead principle of embedding ISNI into our workflows the steps to that are described first before concluding with an outline of the functionality of the Registration Service Portal.

Building a CIP workflow for ISNI assignment – next steps

As already noted earlier the British Library's cataloguing-in-publication programme has been contracted out and for many years has been provided by a company called Bibliographic Data Services (BDS). BDS have been a member of the ISNI UK Publishers Interest Group since its inception and they are closely engaged with the goal of embedding ISNI as a supply chain identifier. As part of the current CIP workflow BDS use and supply name headings from the LC/NACO file in their records. As a precursor to implementation of ISNI the British Library has already supplied a reconciliation file for corresponding LC/NACO – ISNI equivalents for BDS to use to facilitate automatic assignment on the back of their use of LC/NACO. The next step will be to update this file with an additional correspondence file based on the forthcoming update of the LC/NACO file in the ISNI database. Once this is in place BDS systems and workflows are primed and ready for implementation. At this point in time the details of a feedback loop to the publishers supplying BDS with pre-publication information to inform CIP work has yet to be determined, as does the potential role for BDS acting as a Registration Agency for original assignments, but the workflows as building blocks to inform those decisions will be in place.

Providing an ISNI Registration Self-Service Portal

The final piece of the British Library's engagement with the supply chain has been the development of an online service for individual requests. As part of our provision of quality assurance services to the ISNI-IA we respond to user queries and feedback, often leading to requests for updates and additions to existing records and requests for new assignments. We have firsthand experience of a wide

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

level of interest in ISNI amongst smaller publishers and directly from authors, artists, and performers across all repertoires of creative content. We are also acutely aware that at this level of interest and engagement the fact that only ISNI Registration Agencies and ISNI Members can register new ISNIs through privileged access to the ISNI assignment system interfaces is a barrier for those without the means to engage at the membership level. Since we are dealing with requests of all kinds at an individual level that is growing alongside the broadening engagement in ISNI we also know it is cumbersome and costly to deal with these individual requests sent to us through system-generated emails.
As part of another project, involving cross-sector engagement with the music industry initiated by the British Library's National Sound Archive, the Mellon Foundation provided us with specific funding to build an End User Portal for ISNI assignment requests. We have now completed development of thus system and it became operational in February 2020. The portal supports three main functions: Search, Request, and Add Data. The portal mediates these functions to interact directly with the ISNI system. The first two of these mirror the capability developed recently by Soundways in their Sound Credit system described above. As with the Sound Credit system the BL Service requires users to register on the system to access the functionality. Search is the critical first step to ensure that a pre-existing ISNI is not over looked before submitting a request for a new ISNI. As a second check, when a request is submitted, it passes through the matching algorithms in the ISNI system in case a similar name identity does already exist. The Add Data function is an additional aspect of the service that will allow the many end users who want to enrich an ISNI record with additional titles or links to do so directly and easily. Editing existing data is not permitted because the ISNI database is built from the metadata of its Members and contributors and only members can edit their own data in an ISNI record. The British Library will regularly monitor and quality assure all activity and transactions that go through the portal.

Concluding reflections

This paper has sought to present a short update on the growing adoption of ISNI as a name identifier supporting different metadata supply chains. Although only a selection from all the activities going on across the ISNI network of 65+ Agencies and Members, it has provided examples that highlight drivers behind the interest from the supply chains. Drivers that position engagement with ISNI in the broader context of a more developed interest in the value of high quality metadata as an essential component in supply chain management for commerce, discovery and the attribution of rights. It has sought to contextualize the implications of this for libraries and our common interest in bibliographic control by show casing just one approach, developed by the British Library, at engaging with the UK book publishing supply chain. We depend much on the supply of publishers' metadata but we have only had limited influence on bringing it into convergence with libraries' metadata requirements. Although authority control is only a single component of library metadata it has long been one of our most expensive metadata creation activities. Shifting the task of authority control into simultaneous management of ISNIs in the supply chain provides us with an opportunity to share that cost and to share its value. Metadata conceived and developed in the library sector, redefined as identity management, becomes a shared, common goal and the global supply chain becomes part of the solution.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12728

References

DDEX, n.d. Accessed June 2021. https://ddex.net/

EDItEUR, n.d. Accessed June 2021. https://www.editeur.org/

EIDR, n.d. Accessed June 2021. https://www.eidr.org/standards-and-interoperability/

ISNI, n.d. Accessed May 2021. https://isni.org/

Joint Steering Committee for Review of AACR, "RDA/ONIX Framework for Resource Categorisation". Last modified august 3, 2006. https://www.loc.gov/marc/marbi/2007/5chair10.pdf

Matthews, Peter. 2020, "Introducing the BICMetadata Capability Directory". EDItEUR website last accessed June 2021. https://www.editeur.org/3/Events/Event-Details/561

MVB. 2020. "MVB becomes an ISNI Registration Agency". Press release posted on ISNI website, January 2, 2020. https://isni.org/page/article-detail/mvb-becomes-an-isni-registration-agency/

Movielabs. 2019. "Creating a Talent Identifier for the Entertainment Industry". Last modified August 28, 2019. https://movielabs.com/talentid/Talent_ID.pdf

Paskin, N & Rust, G. April 2014. "Linked Content Coalition Principles of Identification". Linked Content Coalition website. http://doi.org/10.1000/287

Sound Credit. 2020. "Music industry ISNI registrations now free and automated". Press release posted on ISNI website October 23, 2020. https://isni.org/page/article-detail/music-industry-isni-registrations-now-free-and-automated/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

JLIS.it

VIAF and the linked data ecosystem

Nathan Putnam[(a)]

a) OCLC, http://orcid.org/0000-0002-3984-3035

Contact: Nathan Putnam, putnamn@oclc.org
Received: 30 April 2021; **Accepted:** 14 June 2021; **First Published:** 15 January 2022

ABSTRACT

This article reviews the founding, current state, and potential future of VIAF®, the Virtual International Authority File. VIAF consists of an aggregation of bibliographic and authority data from over 50 national agencies and infrastructures, systems that follow different cataloging practices and contain hundreds of languages. After a short history of the project, the results of surveys for implementers of linked data projects on the use of VIAF data and provides suggestions for future use and sustainability.

KEYWORDS

RDF; Library Linked Data; VIAF; History.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

The Virtual International Authority File, known as VIAF, provides cultural heritage institutions and users with access to a combined, single international authority file with data from national libraries and infrastructures worldwide. VIAF contributors supply authority data that is matched, linked, and clustered with existing VIAF entities. VIAF allows researchers to identify names, locations, works, and expressions while preserving the regional language, spelling, and script preferences. There are more than 50 VIAF contributors from over 30 countries. The VIAF Council governs VIAF, which includes representatives from the contributor organizations and provides guidance on the policies, practices, and operations of VIAF.
This article begins with a short history of VIAF, including some current statistical information. It then discusses VIAF within the larger linked data ecosystem through several surveys conducted by OCLC. It concludes with a discussion regarding OCLC's continued support for VIAF, its use and potential integration into OCLC's shared entity management infrastructure, and recommendations for further research and investigation.

VIAF history and current use

History

In April 1998, the United States Library of Congress (LC), the German National Library (Deutsche Nationalbibliothek, or DNB), and OCLC wanted to test linking to each other's authority records for personal names as a proof-of-concept project. In August 2003, the LC, the DNB, and OCLC formed the VIAF Consortium in a written agreement during the International Federation of Library Associations and Institutions (IFLA) conference in Berlin, Germany. In October 2007, the National Library of France (Bibliothèque nationale de France, or BnF) joined the consortium. After this, the four organizations, assuming the role of Principals, had joint responsibility for VIAF with the three libraries contributing authority and bibliographic content, while OCLC supported the software and infrastructure. Other organizations later joined the consortium as Contributors, providing source files and expertise to advance the state of VIAF. Due to the proof-of-concept success, the Principles and Contributors sought a suitable long-term organizational arrangement for VIAF. After considering several options, the Principals and Contributors agreed to transition VIAF to OCLC, which was completed in April 2012 (Murphy, 2012).
As of September 2020, VIAF receives data from 56 sources and includes 172 million bibliographic records, 87 million authority records, and 33 million cluster records. Records are clustered, i.e., grouped together, when they represent the same thing but with data from different sources. VIAF stores both the source record and the aggregated cluster record. Figure 1 provides detailed statistics on source information, authority records by type, and top language representation. The comparison year in the figure uses OCLC's fiscal year, which runs from July to June. Interestingly, the top three languages within VIAF are of the three Principal institutions, LC, DNB, and BnF. VIAF also contains a range of authority records from the sources, including 10.5 million corporate authorities, 10.9 million geographic authorities, approximately 60 million personal name authorities, and 5.7 million title authorities, totaling 87 million.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

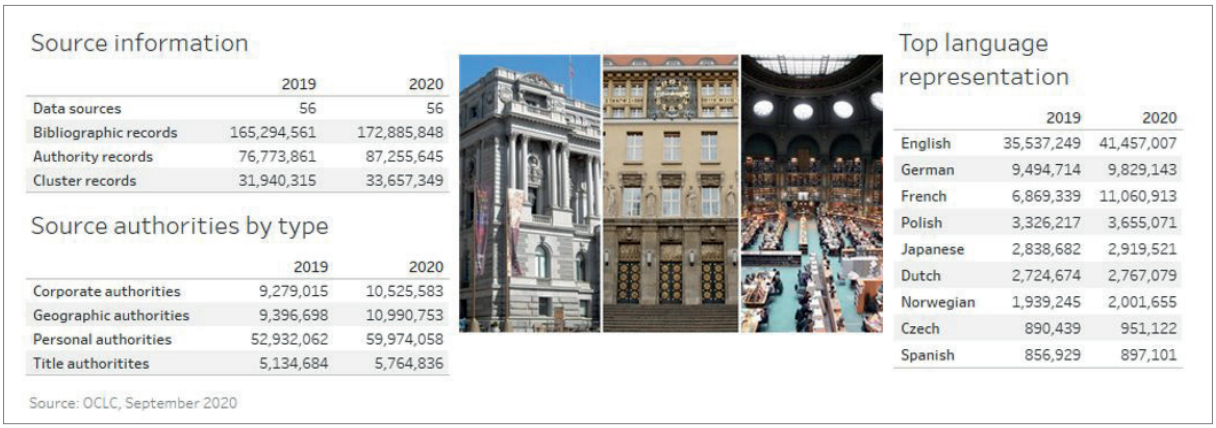

Fig. 1. Source, type, and language representation

VIAF continually adds data from existing and new sources. As seen in Figure 2, data clusters continue to grow, and the cluster types for personal, corporate, work, expression, and geographic authority records. While there are considerably more personal name clusters within VIAF, OCLC believes that the other types will increase in importance as existing and new users consume the VIAF data.

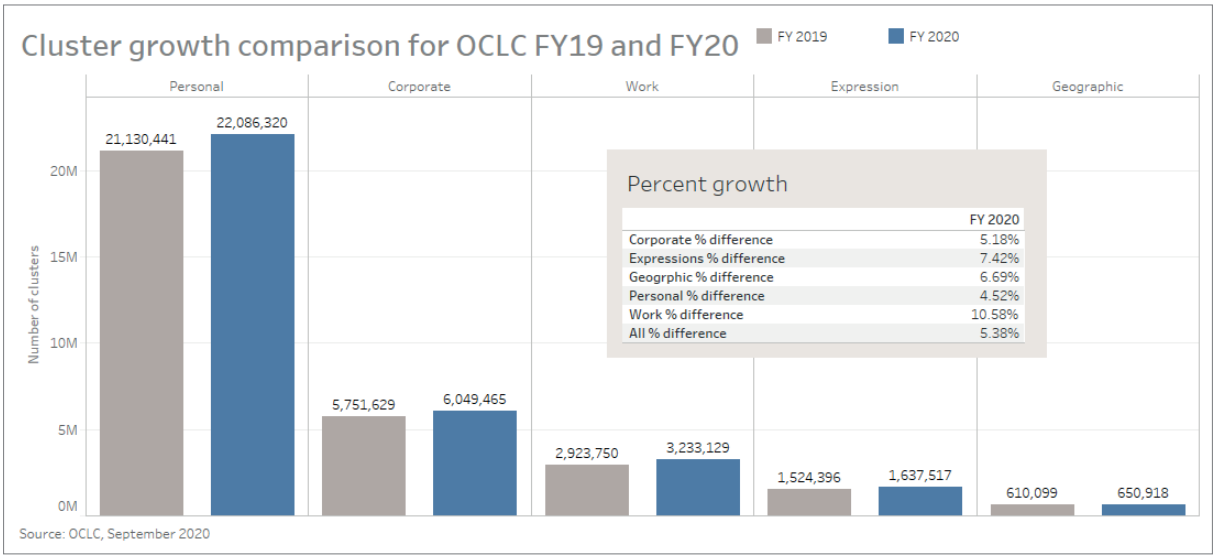

Fig. 2. Cluster comparison between OCLC Fiscal Year 2019 and Fiscal Year 2020

International linked data survey for implementers

During the past seven years, OCLC Research conducted surveys on implementing various linked data tasks and uses. Building upon the interest of the OCLC Research Library Partners Metadata Managers Focus Group, OCLC Research conducted the first "International Linked Data Survey for Implementers" between 7 July and 15 August 2014. This followed updates to the original survey in 2015 and 2018. This article discusses the results regarding the use of VIAF data, but analyses and

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

results are available on the OCLC Research Linked data webpage (OCLC Research, n.d.). Interested persons can access the data directly on the OCLC Research linked data pages or through several articles written by Karen Smith-Yoshimura, including her discussion and analysis of the results[1].
Many institution types participated in the surveys including research libraries, national libraries, research institutions, library networks, governments, service providers[2], public libraries, museums, and a few classified as other. While research libraries continue to have many responders, Figure 3 shows the growing interest in different groups like national libraries, research institutions, and government institutions. Even on the lower end of the responder spectrum, public libraries and museums have seen a slight growth.

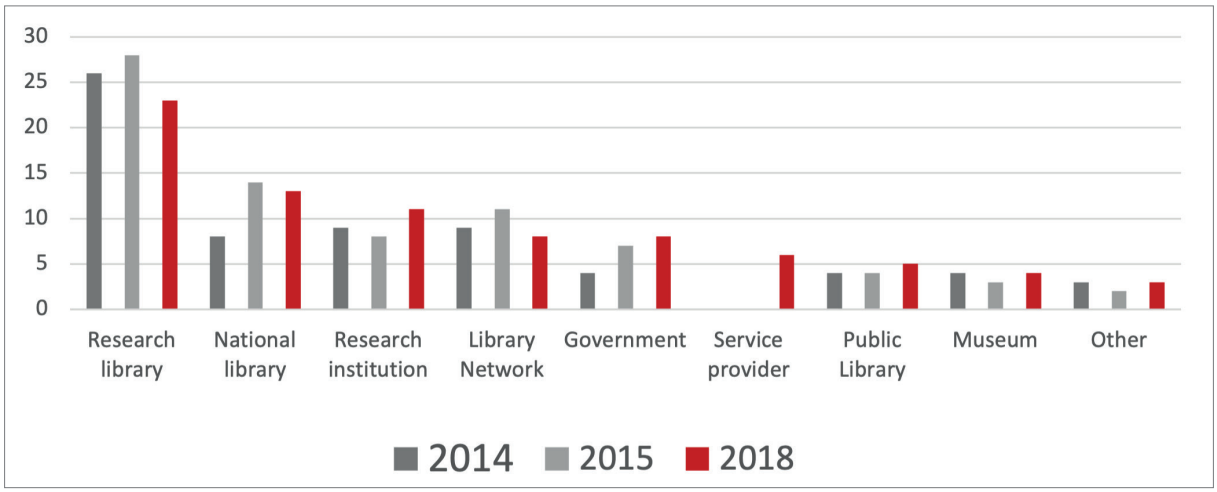

Fig. 3. Reponding insitutions by type

As the ecosystem matures and adoption increases, the participation of these groups will continue to grow. While there was a slight drop in the number of institutions answering how long they have had a linked data project or service in production, the number of projects and the time they have remained active continue to grow. 75% of the linked data projects/services described in 2018 are in production, slightly higher than the 67% reported in 2015. 40% of the linked data projects/services described in 2018 have been in production for more than four years.
The 2018 survey highlights the top seven linked data implementations. "Most used" is measured by the average number of requests per day, with all services reporting over 100,000 requests per day. All eight services have also been in production for more than four decades and include:

- American Numismatic Society's nomisma – a thesaurus of numismatic concepts[3]
- Bibliothèque nationale de France's data.bnf.fr – provides access to the BnF's collections and is a hub among different resources[4]

[1] See https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-survey.html for a complete listing of Karen's publications and presentations
[2] Service providers responded only to the 2018 survey
[3] http://nomisma.org/
[4] https://data.bnf.fr/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

- Europeana – an aggregation of metadata for digital objects from museums, archives, and audiovisual archives across Europe[5]
- Library of Congress Linked Data Service – provides access to over 50 vocabularies[6]
- National Diet Library's NDL Search – provides access to bibliographic data from Japanese libraries, archives, museums, and academic research institutions[7]
- North Rhine-Westphalian Library Service Center (hbz) Linked Open Data service – provides access to bibliographic resources, libraries and related organizations, and authority data[8]
- OCLC's Virtual International Authority File (VIAF) – an aggregation of over 50 authority files from different countries and regions[9]

Figure 4 shows the top ten linked data sources consumed by the 2018 survey respondents compared to 2015. The count of respondents in the 2018 and 2015 surveys was the same, 69 and 68, respectively. Six of the ten sources dropped between the 2015 and 2018 surveys while the other four grew. The most considerable change was in the increased use of Wikidata. And even though VIAF dropped between the two surveys, it is still ranked relatively high, coming in second after the Library of Congress's ID service.

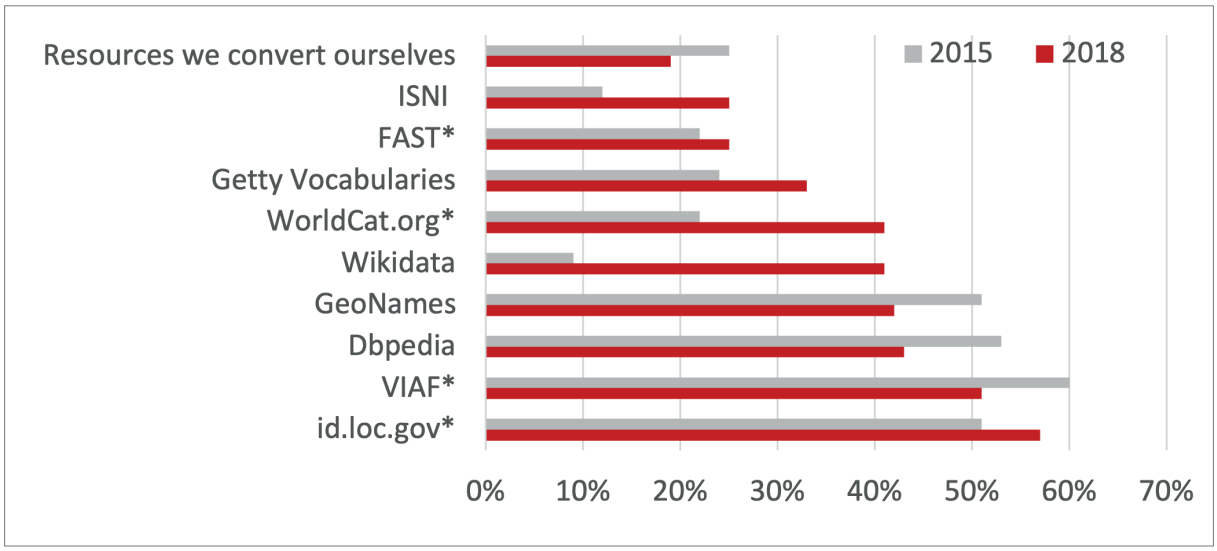

Fig. 4. A comparison of linked data sources consumed in 2015 and 2018

[5] https://www.europeana.eu/
[6] https://id.loc.gov/
[7] https://iss.ndl.go.jp/
[8] http://lobid.org/
[9] http://viaf.org/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

The potential for VIAF

VIAF continues to be supported by OCLC and, as discussed earlier, continues to add new sources and data. The ongoing success of VIAF for various consumers, including OCLC, will depend on greater integration into the linked data environment. Success includes transforming VIAF from a primarily MARC-based system to native RDF and integrating with RDF services and support.
With financial support from the Andrew W. Mellon Foundation, OCLC is building a shared entity management infrastructure for library linked data. When completed in December 2021, this infrastructure will include authoritative descriptions of several types of entities, including works and persons, and will be enhanced and managed by the library and OCLC. Connections to other external vocabularies will place library collections in a broader context across the web.[10]
VIAF plays an integral role in the entity infrastructure, especially during the infrastructure's initial development phase. The grant-funded portion using VIAF entities to connect person entities within the infrastructure. During the first six months of data loading, selected VIAF clusters had connections to either WorldCat® works or Wikidata entities. The key to the initial phase was that the entities had built-in relationships with other entities that provided an enriched experience. The second six-month checkpoint continued the enrichment by adding additional entities. The second six-month checkpoint, which ended December 2020, included personal name entities from VIAF and work entities from WorldCat FRBR clusters. While not part of the grant requirements, it also had place entities from a separate linked data pilot for OCLC CONTENTdm®[11] with data from GeoNames, a database of geographical place names[12].

Areas for further investigation

The existing VIAF infrastructure continues to meet the goals of the original Principals and current Contributors. As with any library data project, continued usefulness will require change.
Two key areas to help determine the future of VIAF include running the implementers survey in the coming year and continued integration within the entity infrastructure. The implementer survey would indicate the continued use of VIAF within the larger linked data ecosystem and the probable and continued growth of Wikidata. Implementing VIAF into the infrastructure will help ensure stability and continuity as the ecosystem moves from record-based description to graph-based. Note that OCLC remains committed to providing a level of free access to those that wish to use the VIAF data regardless of in which ecosystem it finds itself. Additional areas for consideration include continued work with Wikidata partners to find solutions to challenges and the ever-on-going issues revolving around data quality and integrity.

[10] More information on the entity management infrastructure can be found at oc.lc/sharedentitymgmt.
[11] More information on the CONTENTdm Linked Data Pilot can be found at oc.lc/transform-linked-data.
[12] https://www.geonames.org/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12749

References

Murphy, Bob. 2012. Virtual International Authority File service transitions to OCLC; contributing institutions continue to shape direction through VIAF Council. 4 April. Accessed January 24, 2021. https://worldcat.org/arcviewer/7/OCC/2015/03/19/H1426803137790/viewer/file1365.html.

OCLC Research. n.d. Linked data from OCLC Research. Accessed January 24, 2021. https://www.oclc.org/research/areas/data-science/linkeddata/linked-data-survey.html.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

JLIS.it

Call me by your name: towards an authority data control shared between archives and libraries

Pierluigi Feliciati[(a)]

a) Università degli studi di Macerata, Dipartimento di Scienze della Formazione, dei Beni Culturali e del Turismo, http://orcid.org/0000-0002-2499-8528

Contact: Pierluigi Feliciati, pierluigi.feliciati@unimc.it
Received: 14 April 2021; **Accepted:** 11 May 2021; **First Published:** 15 January 2022

ABSTRACT

An important and not often addressed topic – considering the issues opened by cross-disciplinary projects – is the shared control of authority records, or better authority metadata, extended to other documentary and cultural heritage sciences. This paper will examine the potential opened by multi-dimensional and networked logics in the representation of entities in the form of data towards which the document communities are converging. This approach is even more valid if we consider the users' point of view, presently forced to jump from one information environment to another, and confront different names, forms and attributes for the same entities. The core entities to work on are persons, corporate bodies, places, chronological contexts, events, qualifying their relationships. After a brief resume of archival description's peculiarity, the paper highlights the updated standards available, mostly IFLA-LRM and RiC, precious documents to start from and stimulate an active collaboration. To facilitate the sharing, control, and enrichment of authority data in the form of RDF assertions, librarians and archivists may follow several pathways: matching the existing conceptual models, converging on a shared data playground like Wikidata, and developing foundational meta-ontology.

KEYWORDS

Archival description; Semantic web; Wikidata; Authority data; IFLA-LRM; RiC.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

Introduction: convergences between archives and libraries

In the digital era we live in, and after centuries of applying the profession in archives and libraries, documentary disciplines share some fundamental lines. For example, for preserving paper documents and records, quality and digital resources management, digital preservation, administrative metadata. However, there are traditionally few convergences about principles, methods, and informational approaches. The description seems to be the crucial activity that keeps the two professions furthest away, especially in Europe and mainly in Italy. Whether some bridges were more comfortable to be built, the informational approaches are commonly distinct because of the objects' nature, the separated communities and projects, and the awkwardness in converging towards shared goals. Nevertheless, this paper argues that it is impossible to postpone the goal of a shared, integrated control of authority data, extending the most up-to-date approaches to all the areas of documentary and cultural heritage disciplines. This paper focuses on the potentials of collaboration opened by the multi-dimensional and networked logic in representing information entities towards which the documentation communities are converging. Moving from the presentation of archival description peculiarity, matched with the recent evolutions for bibliographic catalogues, this paper will try to shape the future possibilities to activate the development and control of shared authority datasets.

Archival description and authority control

Traditionally, archival description produces closed information pieces, inventories, or finding aids, representing individual archival fonds, informing about their provenance and internal logical partitions (Duranti 1992). The descriptive standards released by the ICA-International Council of Archives from the 90s to 2008 formalized this approach at the international level. The sage, secular principle of *respect des fonds* provides that every fond has to be managed and described separately, as a particular case due to its creator's unique activity. Moreover, the multilevel description rules state that each level of description has to give «information for the parts being described», and archivists should «present the resulting descriptions in a hierarchical part-to-whole relationship proceeding from the broadest (fonds) to the more specific» (ICA 2000, 12). The context prevails over the content, and rarely inventories reach the item level, offering data about records. This model has necessarily held back any connection among descriptions, isolating every pair creator/fond as a unique informational resource. These standard-compliant descriptions are produced mostly adopting relational databases and made accessible through a textual search on descriptive fields. Consequently, the archival description has not easily followed the World Wide Web's evolutions, markedly in the new century, whether the archival information entities are shaped as definitive records with closed hierarchical relations and hardly could keep the form of graphs, neither hypertextual, nor semantic-based.

Regarding the authority records, the archival access points according to the ICA descriptive standards are referred just to archives' creators (corporate bodies, persons or families). They have to be «based upon the elements of description» and their informational value «is enhanced through authority control» (ICA 2000, 9). The ISAAR(CPF) rules guide archivists in editing authority records, even establishing relations between them, under some defined categories:

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

hierarchical, temporal, associative, family (ICA 2003, 21-22). We had to underline that those standards' combined effect led to the loss of the access points included in the traditional archival finding aids: personal or corporate bodies' names, places, subjects (*notable things*). Indeed, some crucial elements like names, dates, events, and places were conceived just as attributes of the units of description.

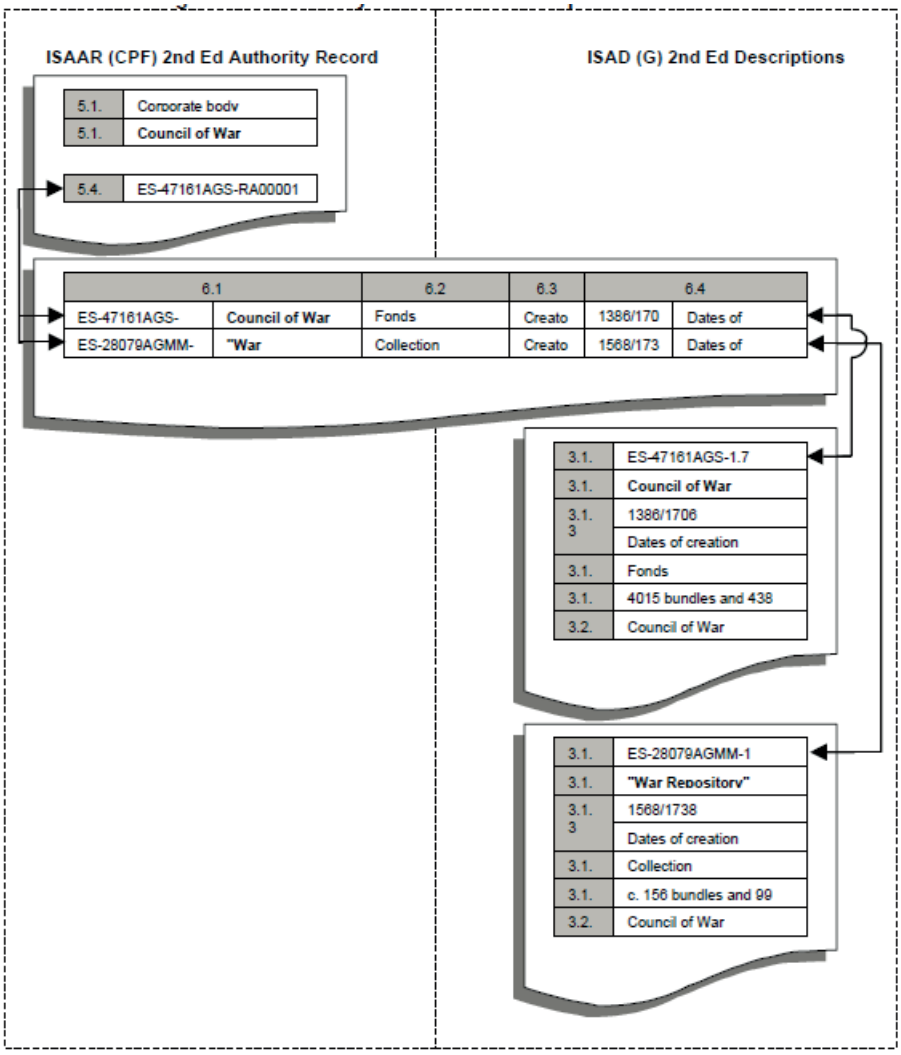

Fig. 1. Representation of how archival authority records can be linked with descriptions of archival materials (ICA 2003, 29)

Furthermore, names (i.e., units of description' titles) have to be extracted from archival files and other sources related to creators' internal organization, with the indication of limiting their normalization as much as possible. To explain better this traditional practice: suppose that the original name of an archival series is "Biccherne" (the magistrate or chancellery of finance from the 13th to the 14th century for Siena, Italy). Archivists are asked to describe this entity under "a formal title or a concise supplied title in accordance with the rules of multilevel description and national conventions." (ICA 2000, 14). This practice underlies two kinds of problems:

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

1. interoperability and Autority Control: every description can only be checked by those who produce it, in possession of the bibliographical reference and especially of the wisdom arising from the heuristic study of the fond. It is almost impossible to build distributed authority control features, and the centralized control is allowed to verify the respect of formal rules;
2. users' friendiness: users may not know the fond or series's original name but are forced to query a database adopting Google-like behaviours. The adoption of relational databases caused the prevalence of searching vs. browsing services, and not-expert users may be deluded or lost while performing their research, as some studies clearly demonstrated (Duff, Stoyanova, 1998; Yakel, 2003; Chapman, 2010).

Nevertheless, recently some Linked Open Data data extraction experiments from archival DBs based on ICA standards were provided. Unfortunately, the assertions produced are not easy to be integrated into the Semantic Web *info-verse* because the ontologies adopted are local, representing specific data models, and could not be standardized in the absence of a shared Conceptual Model.

Archives in the *info-verse*: Records in Contexts

The new ICA standard RiC – *Records in Contexts,* defined by the EGAD – Experts Group on Archival Description from 2012 to 2016 turned upside-down the hierarchic and mono-dimensional logics of ISAD(G) and ISAAR(CPF). Proposing a multi-dimensional description, RiC Conceptual Model aims to be the reference for producing graphs of linked information entities instead of hierarchic or bare database rows connections. The 0.1 draft version of the Conceptual Model was published in August 2016 (ICA 2016) and questioned deeply by the international community (Bunn 2016; Duranti 2016; ANAI-ICAR 2017; SAA 2018). The recommendations covered several aspects: the "western" composition of EGAD, and the request to open RiC to existing ontologies like IFLA LRM (Riva et al. 2017), CIDOC-CRM (CIDOC CRM 2021), PREMIS (LoC 2018), and PROV-O (W3C 2013).

The draft version of RiC-CM was then updated in December 2019, publishing another draft version, the RiC-CM 0.2 (ICA 2019a), on which the RiC Ontology 0.1 (ICA 2019b), developed by the EGAD RiC-O team,[1] was based. Recently, in February 2021, the RiC-0 0.2 was released, compliant with the latest version of RiC-CM, 0.2, released in July 2021, and slightly different from RiC-CM 0.2 preview[2]. Again, a draft version explicitly to be corrected and enriched, in the perspective of the release of RiC-O 1.0. First of all, it "does not include the Conceptual Model Introduction, diagrams, or appendices". Moreover, it has to be quoted the absence of any explicit reference to the acceptance of the community's observations to the 2016 consultation draft and to the methodology adopted in the development process. As regards RiC-O 0.2, it lacks examples and tutorials, and it is explicitly declared that it "will continue to evolve, the next milestone being the release of RiC-O 1.0, which will probably take place by the end of 2021, at the same time as RiC-CM 1.0".[3]

[1] The EGAD RiC-O team is coordinated by Florence Clavaud (Archives nationales de France) and composed by Daniel Pitti (University of Virginia, USA), Aaron Rubinstein (University of Massachusetts Amherst, USA), Tobias Wildi (Docuteam GmbH, Switzerland) and Miia Herrala (National Archives of Finland).

[2] See https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf, accessed November 11, 2021.

[3] See https://www.ica.org/standards/RiC/RiC-O_v0-2.html, accessed November 11, 2021.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

Ric-CM 0.2 deeply changed the entities articulation present in version 0.1, adopting a four-level hierarchical logic: the macro-entity RiC-E01 *Thing* (first level) includes the entities RiC-E02 *Record Resource* (containing RiC-E03 *Record Set*, RiC-E04 *Record* e RiC-E05 *Record Part*), RiC-E06 *Instantiation*, RiC-E07 *Agent* (containing RiC-E08 *Person*, RiC-E09 *Group*, articulated in RiC-E10 *Family* and RiC-E11 *Corporate Body*, RiC-E12 *Position* e RiC-E13 *Mechanism*), RiC-E14 *Event*5 (specifiable with RiC-E15 *Activity*), RiC-E16 *Rule* (specificabile con RiC-E17 *Mandate*), RiC-E18 *Date* (specifiable with RiC-E19 *Single Date*, RiC-E20 *Date Range* or RiC-E21 *Date Set*), and RiC-E22 *Place* (see fig. 2). The entities and sub-entities of RiC-O are expressed as classes, and the properties are detailed in the datatypes. It has to be noted that the Internationalized Resource Identifier of RiC-O is not yet active, so it is not possible to refer to the namespace and allow applications to be automatically processed. This draft state of the new standard, and the Experts Group on Archival Description's isolation from the international community cannot help slow down the development of description tools based on RiC, any projects of conversion of existing catalogues, and the availability of archival linked triples in the semantic info-verse. Anyway, some isolated experiments, not ascribable directly to EGAD, started. We can quote the case presented in a spanish paper (Llanes-Padrón, Pastor-Sánchez and Juan-Antonio, 2017), the French proof of concept PIAAF, *Pilote d'interopérabilité pour les Autorités Archivistiques françaises* (Clavaud 2018),[4] and the Matterhorn RDF Data Model, based on RiC but open to existing ontologies (Dubois, Nef, 2017).
Another archival ontology to consider is the EAC-CPF Ontology (Mazzini, Ricci, 2011), based on the XML schema maintained by the Society of American Archivists with the Berlin State Library. It is used for encoding contextual information about persons, corporate bodies, and families related to archival materials, encoding the rules published in ISAAR(CPF). Some updated archival description applications are offering the export feature of RiC-like RDF triples, converting the hierarchical descriptive structures into multi-dimensional graphs. Nevertheless, nowadays, archives' global semantic interoperability is quite tricky without a wide-accepted, stable and accessible ontology.

Metadata integration between archives and libraries

The notion of catalogue could be taken in its broadest sense: ordered and systematic collection or record of items. Its function could not be reduced to the retrieval and identification of a single item, having the role of activating unexpected connections between different items:

> *Functions of the Catalogue*: The catalogue should be an efficient instrument for ascertaining 2.1 whether the library contains a particular book […] and 2.2 (a) which works by a particular author and (b) which editions of a particular work are in the library. (Statements 1961, 1).

Adopting this broad notion of catalogue, archival finding aids can also be considered catalogues (term commonly used in English). This phenomenon is even more reasonable considering that the outlines of informative objects tend to blur on the web, and in the web of data they are reduced to minimal assertions[5]. Considering the present tendencies in the archival and bibliographic de-

[4] See also http://piaaf.demo.logilab.fr/, accessed april 7, 2021.
[5] See Michetti 2020, 28, note 9.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

scription, we may dare to say that both inventories and catalogues are conceptually and technically outdated. The documentary communities are asked to produce, control, share, monitor and enrich pieces of data, no more deep-web records, entrusting them to be accessed in the *infoverse*, understood, used and re-launched by human or web agents.

Authority control represents an important function to ensure the quality of linked open meta/data, produced through the intermediation of libraries networks but more e more in collaboration with the other memory institutions such as Archives and Museums. Firstly, it is no longer sustainable the management of authority control just at a local or national level. Then, the perspective must be broadened beyond the provenance descriptions, bibliographic, archival or relating to other human artifacts, such as artworks. While respecting the specificity of disciplines, the priority sandbox for archivists and librarians could be sharing authority data, giving to persons, agents, organizations, dates, places, and activities more knowledge facets. Despite the uncertainties, the road of data integration seems to be drawn. The approach driven by RDA and IFLA-LRM (Riva et al. 2017), jointly with the future, stable version of RiC-CM, could be the starting pillars to base on the collaboration. Several pathways to reach this goal could be followed. The first, maybe more manageable, is enabling the quoted conceptual models to talk, i.e. converging on the same concepts (entities) and defining the possible relations.

To open the work to be done, the Table 1 is a starting, tentative of matching the core entities of IFLA-LRM and RiC-CM 0.2. The RIC-E01 *Thing* is not that far from the *Res* entity of IFLA-LRM, considering their relations on the one hand with *Record Resource*, *Agent* , *Event*, and *Date*, on the other with *Work/Item* (considering the substantial unicity of records), *Time-span*, *Place* and *Agent* (Person, Collective Agent). The LRM conception of *Nomen* as an appellation of *Res* could be an interesting question to be addressed in the stable version of RiC, considering the complexity of appellations in archival description: original, derived, normalized, synthesized.

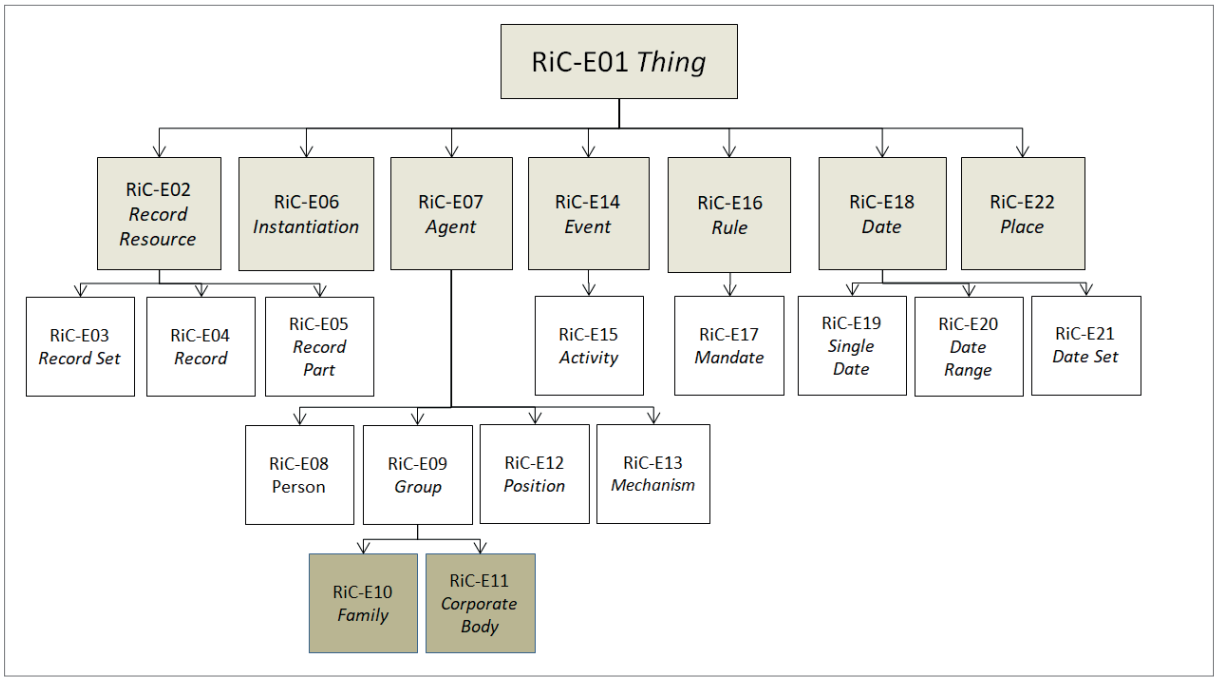

Fig. 2. RiC-O *diagram of entities* (Feliciati 2021, 99)

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

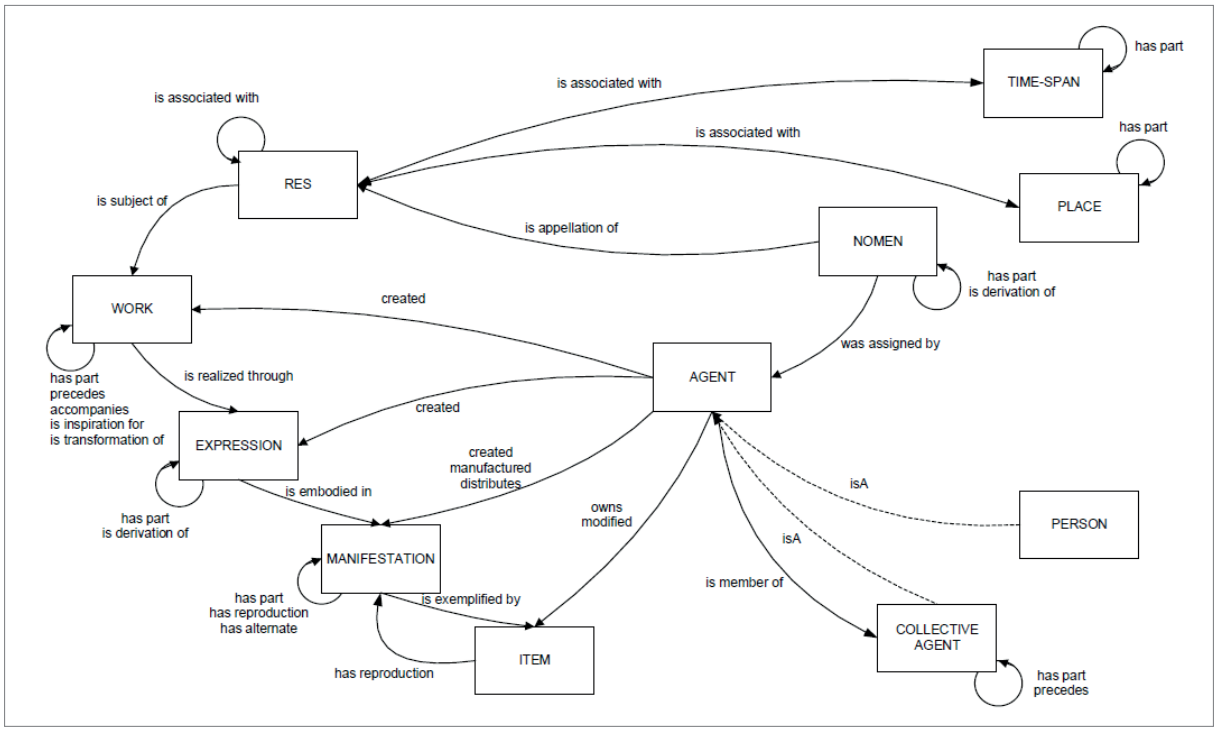

Fig. 3. IFLA-LRM table 5-6, *final overview diagram*

IFLA-LRM	**RiC-O**
Res + Nomen	Thing
Time-span	Date (Single Date, Date Range, Date Set)
Place	Place
Agent (Person, Collective Agent)	Agent (Person, Group, Position, Mechanism)
Work/Item	Record Resource

Table 1. Tentative correspondence between IFLA-LRM and RiC-O core entities

The second path to be followed is cooperating actively on a meta platform, a shared data playground, like Wikidata.

Wikidata (https://www.wikidata.org) is a project developed starting from its mother project, Wikipedia (https://www.wikipedia.org), both free and open repositories accessible over the web. Unlike Wikipedia, Wikidata stores information as structured data in a database. While the primary mission of Wikidata was to serve as a central repository for Wikipedia and other Wikimedia projects, it plays now the role of an independent, open, collaborative, and versatile platform. It could be used for «many different services and applications, from reusing identifiers to facilitate data integration, providing labels for multilingual maps and services, to intelligent agents answering queries and using background knowledge» (Vrandecic, 2013, p. 90). Wikidata uses Linked Open Data to store facts about items as nodes linked by properties as vertices; thus the project is often

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

referred to as a linked open data repository of facts, available under an open CC 0 license. Tim Berners-Lee argued that his Semantic Web vision was hard to be realized because the ontologies must be developed, managed, and endorsed by (missing) practice communities. With Wikidata successfully serving as a LOD repository of facts, the Semantic Web's vision idea seems feasible. If we regard archives and library metadata as (functional) statements of facts that facilitate access of knowledge materials, Wikidata can be adopted as an ideal tool to make these facts accessible and discernable to machines and intelligent algorithms. In fact, «Wikidata can be used to make these facts accessible and discernable to machines and intelligent algorithms to realize the vision of the Semantic Web. For instance, it is quite conceivable to imagine that library patrons in the future may no longer use library catalogues and depend on intelligent devices and algorithms to search and access library holdings over the web» (Tharani, 2021, 2).

Many working groups of librarians are active on Wikidata managing and enrichment, defining a metadata structure for libraries and uploading and sharing local metadata globally (Bergamin, Bacchi 2018)[6]. Some archivists launched recently the *Wikidata:WikiProject Archival Description*, with the aim «to create the world's most comprehensive high quality database of archival fonds and heritage collections, to represent archival structures within Wikidata where this is deemed useful and to ensure the interlinking between archival finding aids and Wikidata»[7]. The project, connected with the *Wikidata:WikiProject Archives Linked Data Interest Group*, is leaded by French archivists and is considering the elaboration of ICA descriptive standards before RiC. In Italy, since 2020, is active the *Wikidata:Gruppo Wikidata per Musei, Archivi e Biblioteche* (GWMAB)[8], inspired by the Wikidata Affinity Group[9], launched mainly by librarians but open to the potentialities of Wikidata for Museums, Archives and Libraries. The purpose of this group to support culture professionals is going to produce some results in adding and correcting metadata related to museum and archives.

In order to figure out the shared work to be done on Wikidata, it could be useful the presentation of a case of possible trans-disciplinary integration: Umberto Eco. Umberto Eco (1932 –2016) was an Italian medievalist, philosopher, semiotician, cultural critic, political and social commentator, and novelist. After his death, his library is presently going to be split into two collections: the ancient books sold to Biblioteca Braidense (Milan) and his modern books and archival records, donated to the University of Bologna. The "Eco, Umberto" authority records in ISNI (0000 0001 2283 9390), VIAF (108299403), and other sources like the Italian SBN (CFIV006213) refer just to his being an author of works. Nevertheless, he was a library collector and owner, an archives creator, a subject of books and essays, of art portraits, photos. Besides the authority record and the Wikidata entity of interest concerning him, the places related to his life and work, the institutions holding his personal library and archives, his political activity, his family, his relationship with many other people should be semantically represented by letting different professionals working on the same information units. The Wikidata element referred to Umberto Eco (Q12807) [10], relatively poor at the time of the

[6] See https://www.wikidata.org/wiki/Wikidata:WikiProject_Libraries, accessed November 21, 2021.
[7] See https://www.wikidata.org/wiki/Wikidata:WikiProject_Archival_Description, accessed November 21, 2021.
[8] See https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche, accessed November 21, 2021.
[9] See https://wiki.lyrasis.org/display/LD4P2/LD4-Wikidata+Affinity+Group, accessed November 21, 2021.
[10] See https://www.wikidata.org/wiki/Q12807, accessed November 21, 2021.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

Bibliographic Control Conference, was enriched in the subsequent weeks. It attributes properties about his personal and professional life, his *notable work* (P800), *awards received* (P166), and his being the *owner of* (P1830) a personal library. The collaboration of archivists to enrich this element could add more properties, like his being a *creator* (Q59275219), *collection creator* (P6241), enrich the element *Umberto Eco's library* (Q35029860) and create the element referred to his archive.
The third pathway to build shared authority control between archivists and librarians could be the convergence towards a brand new foundational Conceptual Model.
The focus of this line of work could be the selection of shared classes, entities and properties, such as agents (persons, corporate bodies, families), their roles/functions in different contexts, geographic names (even historical), chronological data (exact dates or data range), actions/events, qualifying their multiple relationships. To develop this needful reference model and trans-ontology could facilitate and ebnable the integration of authority records in the form of RDF assertions. Collecting, connecting, enriching and controlling high-quality semantic information provided from different data sources will increase the potential of online services, making them richer and more useful for final users.

Conclusions

A shared approach to authority control would be even more valid considering the final users' perspectives. At present, as users, we are often forced to jump from one online source to another, even produced by the same institution, to compare and choose different forms of names and attributes referred to the same entities. Our time is not saved. The quality of use for documentary environments needs to be increased through an integrated approach to authority control and the adoption of updated metadata technologies. This strategy could represent a virtuous opening to the wisdom of crowds, by systematically sharing rich LODs, allowing users' annotations, using UX mining and collaborating with a global multilingual knowledge graph like Wikidata.
Interoperability should be possible with other cultural semantic sets of LODs, mostly produced by cultural heritage institutions different from archives and libraries. The goal could be the extension and enrichment of contexts and relations, representing the actual complexity of human activities in times, without reducing the semantic richness of descriptrive data. This perspective marks a step ahead compared with web portals, harvesting simplified metadata sets from data providers' repositories and necessarily affected by the issues of overwhelming search results. In this sense, the CIDOC-CRM model paved the way for semantic models in the cultural heritage sector. Any interoperability perspective can not help but compare with its classes and properties. The challenge posed by the semantic web forces the culture professionals to take a step forward in representing human activities. We have to break down disciplinary walls, enlarge the concept of provenance (Lemieux 2016) and respect the complexity, heterogeneity, discontinuity and transversality of contexts.
Some issues could slow down this process: organizational, the availability of models for standardization, the disciplinary edges. Some organizations, better if international, should take the initiative to launch this ambitious project by calling on experts from different sectors, archivists, librarians and cultural heritage experts to action. We have just to be ready to answer.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

References

ANAI-ICAR. 2017. "Records in Contexts. A conceptual model for archival description (draft v0.1, September 2016). Il contributo italiano", *Quaderni del Mondo degli Archivi*, 2 (luglio 2017), http://www.ilmondodegliarchivi.org/images/Quaderni/MdA_Quaderni_n2.pdf, Accessed April 6, 2021.

Bergamin, Giovanni; Bacchi, Cristian. 2018. "New ways of creating and sharing bibliographic information: an experiment of using the Wikibase Data Model for UNIMARC data". *JLIS.it*, v. 9, n. 3, p. 35-74, sep. 2018. http://dx.doi.org/10.4403/jlis.it-12458. Accessed April 13, 2021.

Bunn, Jenny. 2016. *Results of the ARA SAT consultation on Records in Contexts*, https://www.archives.org.uk/about/community/groups/viewbulletin/59-results-of-the-ara-sat-consultation-on-records-in-contexts.html?groupid=21. Accessed April 6, 2021.

Chapman, J., C.. 2010. "Observing Users: an Empirical Analysis of User Interaction with Online Finding Aids". *Journal of Archival Organization*, 8, 4-30 (2010), https://doi.org/10.1080/15332748.2010.484361. Accessed April 11, 2021.

CIDOC CRM Special Interest Group. 2021. *Definition of the CIDOC Conceptual Reference Model.* Version 7.1, March 2021, http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.1%20%5B8%20March%202021%5D.pdf. Accessed April 6, 2021.

Clavaud, Florence, 2018. *Semantizing and visualising archival metadata: the PIAAF French prototype online.* May 4, https://www.ica.org/en/semantizing-and-visualising-archival-metadata-the-piaaf-french-prototype-online. Accessed April 6, 2021.

Dubois, Alain, Nef, Andreas. 2017. *The Matterhorn RDF Data Model: Implemeting OAIS and RiC in the context of semantic technologies.* Presentation, http://www.alaarchivos.org/wp-content/uploads/2017/12/3.-Alain-Dubois-Andreas-Nef.pdf. Accessed April 7, 2021.

Duff, Wendy, Stoyanova, Penka.1998. "Transforming the Crazy Quilt: Archival Displays from user's point of view". *Archivaria*, 45, 44-79 (1998), https://archivaria.ca/index.php/archivaria/article/view/12224. Accessed April 13, 2021.

Duranti, Luciana. 1992. "Origin and Development of the Concept of Archival Description". *Archivaria* 35 (January), 47-54, https://archivaria.ca/index.php/archivaria/article/view/11884. Accessed April 6, 2021.

Duranti, Luciana (compiler). 2016. Comments on "Records in Context". InterPARES Trust, https://interparestrustblog.files.wordpress.com/2016/12/interparestrust_commentsonric_final2.pdf, Accessed April 6, 2021.

Feliciati, Pierluigi. 2021. "Archives in a Graph. The Records in Contexts Ontology within the framework of standards and practices of Archival Description". *JLIS.it*, Vol. 12, No. 1 (2021), http://dx.doi.org/10.4403/jlis.it-12675. Accessed April 13, 2021.

ICA (International Council on Archives) – EGAD (Experts Group on Archival Description). 2016, *Records in Contexts. A conceptual model for archival description. Consultation Draft v0.1*, September, https://www.ica.org/sites/default/files/RiC-CM-0.1.pdf. Accessed April 6, 2021.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

ICA – EGAD. 2019a, *Records in Contexts. A conceptual model for archival description. Consultation Draft v0.2* (preview), December, https://www.ica.org/sites/default/files/ric-cm-0.2_preview.pdf. Accessed April 6, 2021.

ICA – EGAD. 2019b, *Records in Contexts Ontology (ICA RiC-O) version 0.2*, 2019-12-12, https://github.com/ICA-EGAD/RiC-O/blob/master/ontology/previous-versions/RiC-O_v0-1_release/RiC-O_v0-1.rdf. Accessed April 6, 2021.

ICA – EGAD. 2021a, *Records in Contexts Ontology (ICA RiC-O) version 0.2*, 2021-02-12, https://www.ica.org/standards/RiC/RiC-O_v0-2.html. Accessed April 6, 2021.

ICA – EGAD. 2021b, *Records in Contexts. A conceptual model for archival description. Consultation Draft v0.2*, July, https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf. Accessed August 6, 2021.

ICA - Committee on Descriptive Standards. 2000. *ISAD(G): General International Standard for Archival Description*, *Second Edition*. Ottawa, https://www.ica.org/sites/default/files/CBPS_2000_Guidelines_ISAD%28G%29_Second-edition_EN.pdf. Accessed April 6, 2021.

ICA - Committee on Descriptive Standards. 2003. *ISAAR (CPF): International Standard Archival Authority Record For Corporate Bodies, Persons and Families. Second Edition*, https://www.ica.org/en/isaar-cpf-international-standard-archival-authority-record-corporate-bodies-persons-and-families-2nd. Accessed April 8, 2021.

IFLA – International Federation of Library Associations and Institutions, Cataloguing Section and Meetings of Experts on an International Cataloguing Code. 2017. *Statement of International Cataloguing Principles (ICP).* https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf. Accessed April 10, 2021.

Lemieux, Victoria (ed.). 2016. *Building Trust in Information. Perspectives on the Frontiers of Provenance*. Springer International Publishing. Senza luoigo?

Llanes-Padrón Dunia, Pastor-Sánchez Juan-Antonio. 2017. "Records in contexts: the road of archives to semantic interoperability", *Program*, 51:4, 387-405, https://doi.org/10.1108/PROG-03-2017-0021. Accessed April 6, 2021.

Library of Congress - PREMIS Editorial Committee. 2018. *PREMIS 3 Ontology*. https://id.loc.gov/ontologies/premis-3-0-0.html. Accessed April 6, 2021.

Yakel, E.. 2003. "Impact of Internet-Based Discovery Tools on Use and Users of Archives". *Comma*, 191-200 (2003).

Mazzini, Silvia and Ricci, Francesca. 2011. "EAC-CPF Ontology and Linked Archival Data", *Proceedings of the 1st International Workshop on Semantic Digital Archives*, September 29, 72-81, http://ceur-ws.org/Vol-801/paper6.pdf. Accessed April 6, 2021.

Michetti, Giovanni. 2020. "'Il mondo come puzzle': i beni culturali nel web". *Digitalia*, Anno XV, Numero 1 - Giugno 2020, 26-42, http://digitalia.sbn.it/article/view/2485. Accessed April 6, 2021.

Riva Pat, Le Boeuf Patrick, Žumer Maja. 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf. Accessed April 6, 2021.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12733

SAA (Society of American Archivists) - Council Conference Call. 2018. *Annual Report: Standards Committee and Technical Subcommittees, Appendix D*, 30-44, https://www2.archivists.org/sites/all/files/0118-CC-V-F-Standards.pdf. Accessed April 6, 2021.

Tharani Karim. 2021. "Much more than a mere technology: A systematic review of Wikidata in libraries". *The Journal of Academic Librarianship*, Volume 47, Issue 2, March 2021, 102326, https://doi.org/10.1016/j.acalib.2021.102326. Accessed April 13, 2021.

Vrandecic, Denny. 2013. "The rise of Wikidata". *IEEE Intelligent Systems*, 28(4), 90–95. https://dl.acm.org/doi/abs/10.1109/MIS.2013.119. Accessed April 13, 2021.

W3C. 2013. *PROV-O: The PROV Ontology*, *Recommendation*, April 30, http://www.w3.org/TR/prov-o/, Accessed April 6, 2021.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

JLIS.it

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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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".

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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]

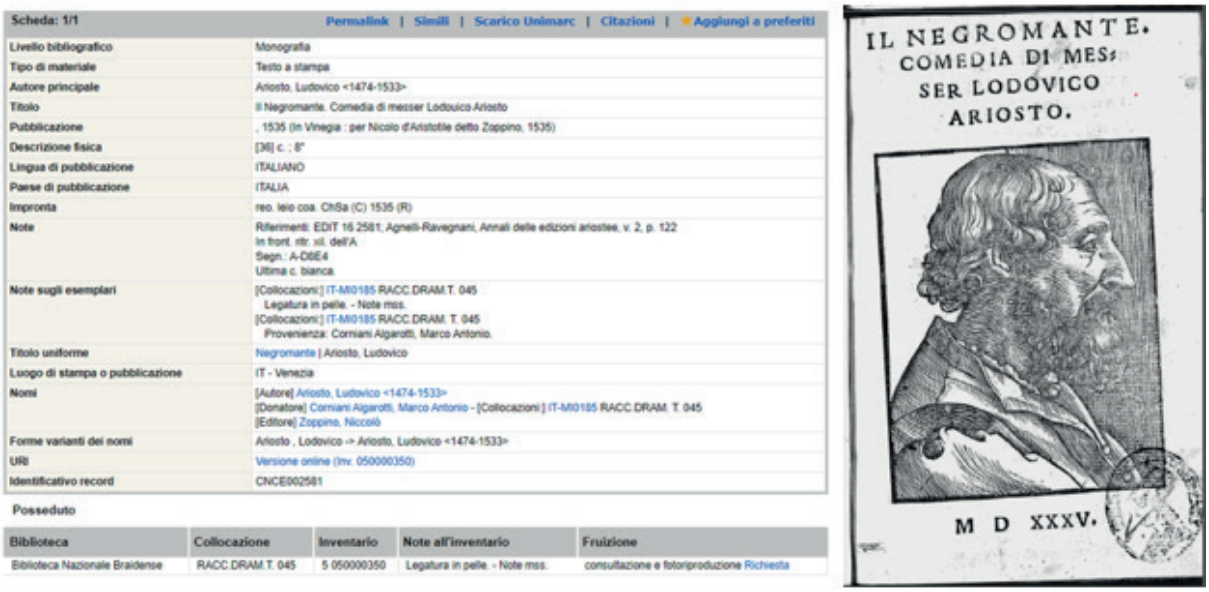

Fig. 1. Link to the digital reproduction of a copy from the record of the paper edition (Source: Opac of the Biblioteca nazionale Braidense, Milano, Italy)

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)

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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.

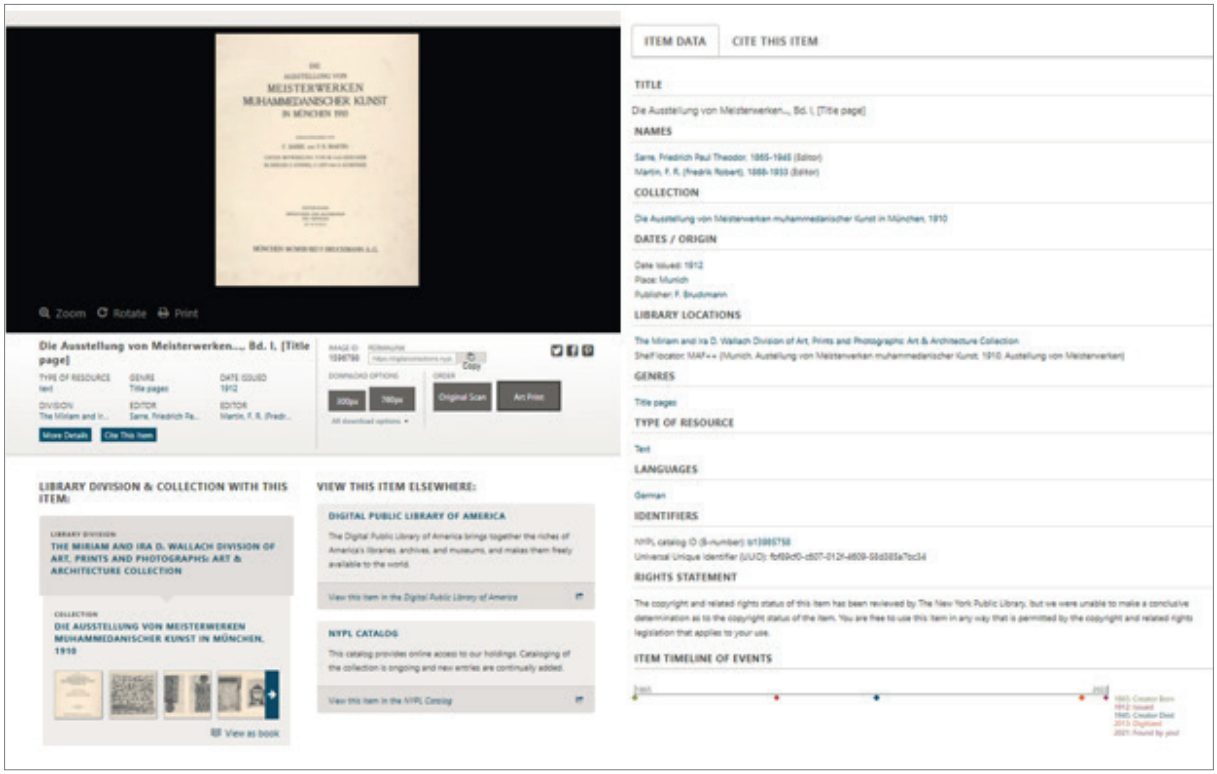

Fig. 3. The Miriam and Ira D. Wallach Division of Art, Prints and Photographs: Art & Architecture Collection, The New York Public Library. "Die Ausstellung von Meisterwerken..." New York Public Library Digital Collections. Accessed April 11, 2021. https://digitalcollections.nypl.org/items/510d47e3-84c9-a3d9-e040-e00a18064a99. 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)

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

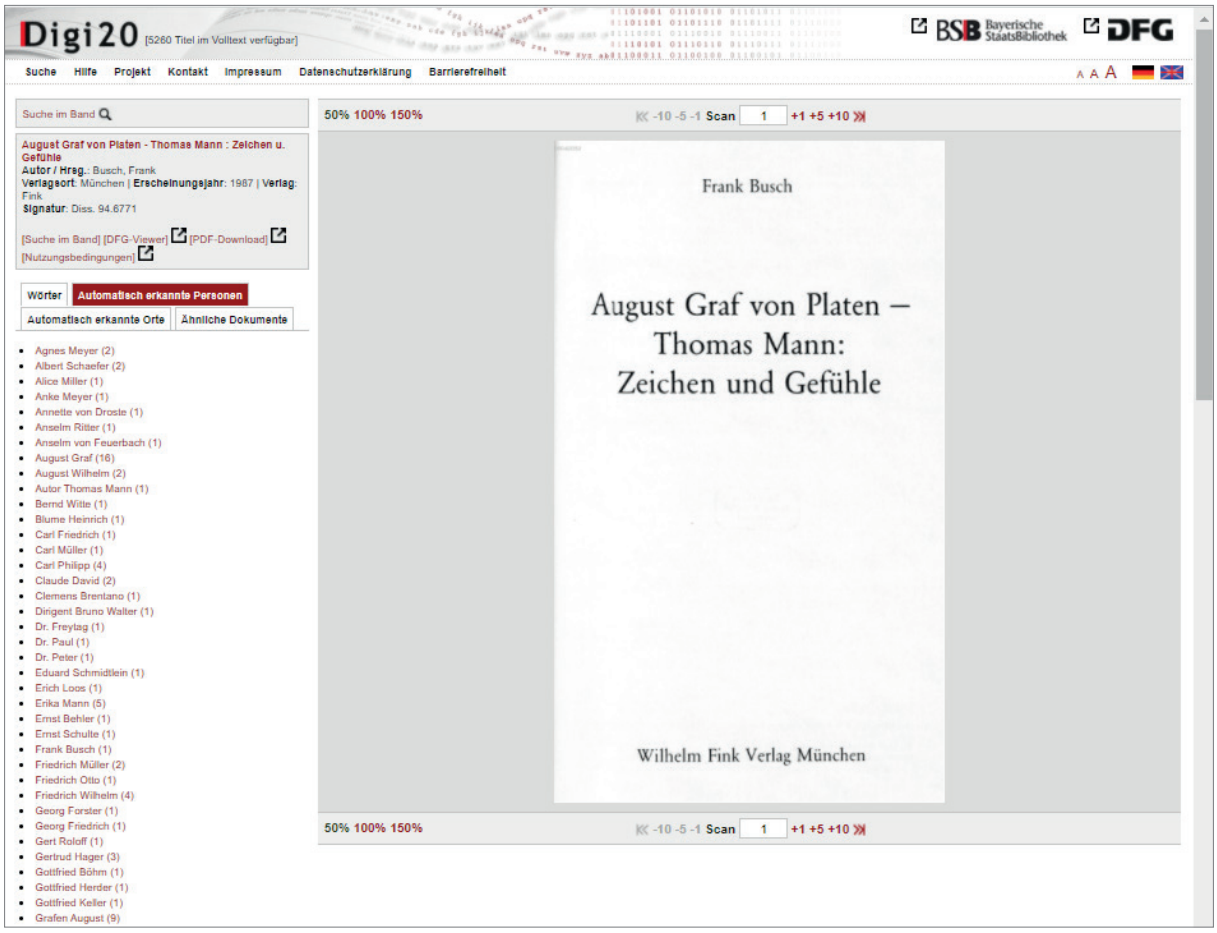

Fig. 4. Busch, Frank. August Graf von Platen-Thomas Mann: Zeichen u. Gefühle. München: Fink, 1987. The digital reproduction of the volume, carried out within the Digi20 ("Digitalisierung der DFG-Sondersammelgebiete") Project can be accessed at the URL: < https://digi20.digitale-sammlungen.de/de/fs1/object/display/bsb00042052_00001.html?leftTab=PER_ent>. 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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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: a⁸ b⁶ A-4L⁶ a-h⁶ i⁴; 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

☐ FI0098	CFICF	Biblioteca nazionale centrale - Firenze - FI - [consistenza] 2 esemplari - [tipo di digitalizzazione] parziale - copia digitalizzata
☐ 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

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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.

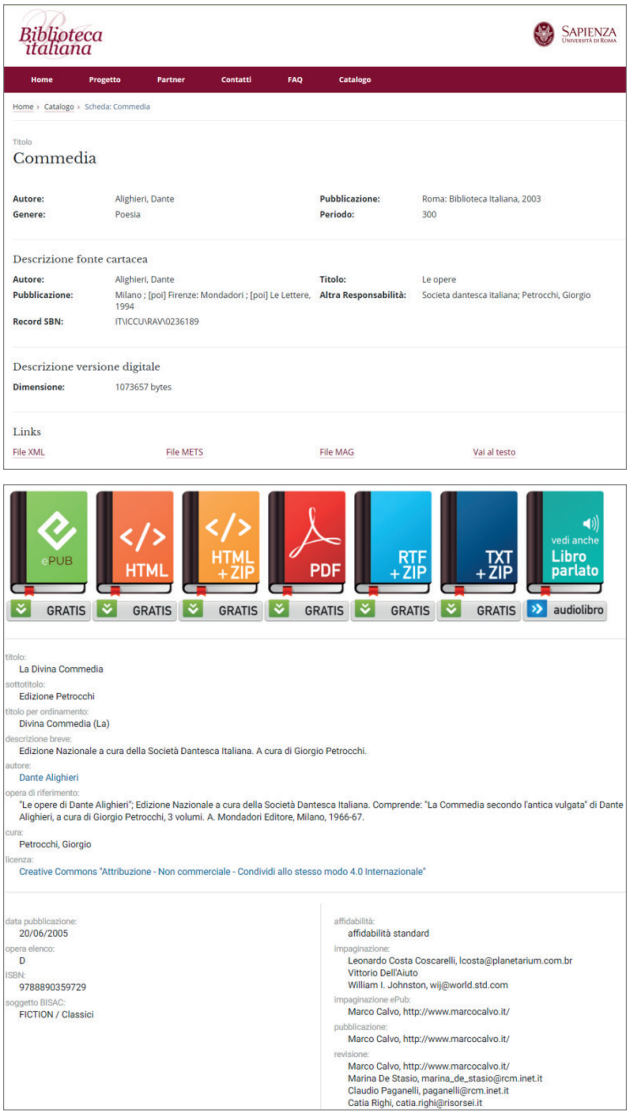

Fig. 6. Two digital editions of Dante Alighieri's Commedia. The first (top) is taken from Biblioteca Italiana, a project aimed at the publication of texts for study purposes; the second (bottom) is taken from LiberLiber, a project aimed at the creation of a public library, which fact explains the variety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

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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 *url*s. 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>
      <titleStmt>
        <title>Commedia</title>
        <author>Dante Alighieri</author>
      </titleStmt>
      <extent>711 Kb in UTF-8</extent>
      <publicationStmt>
        <publisher>Biblioteca Italiana</publisher>
        <pubPlace>Roma</pubPlace>
        <date>2003</date>
        <idno>bibit000019</idno>
        <availability>
          <p>Questa risorsa digitale è liberamente accessibile per uso personale o scientifico. Ogni uso commerciale è vietato</p>
        </availability>
      </publicationStmt>
      <seriesStmt>
        <title>Collezione BibIt</title>
      </seriesStmt>
      <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>
```


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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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.

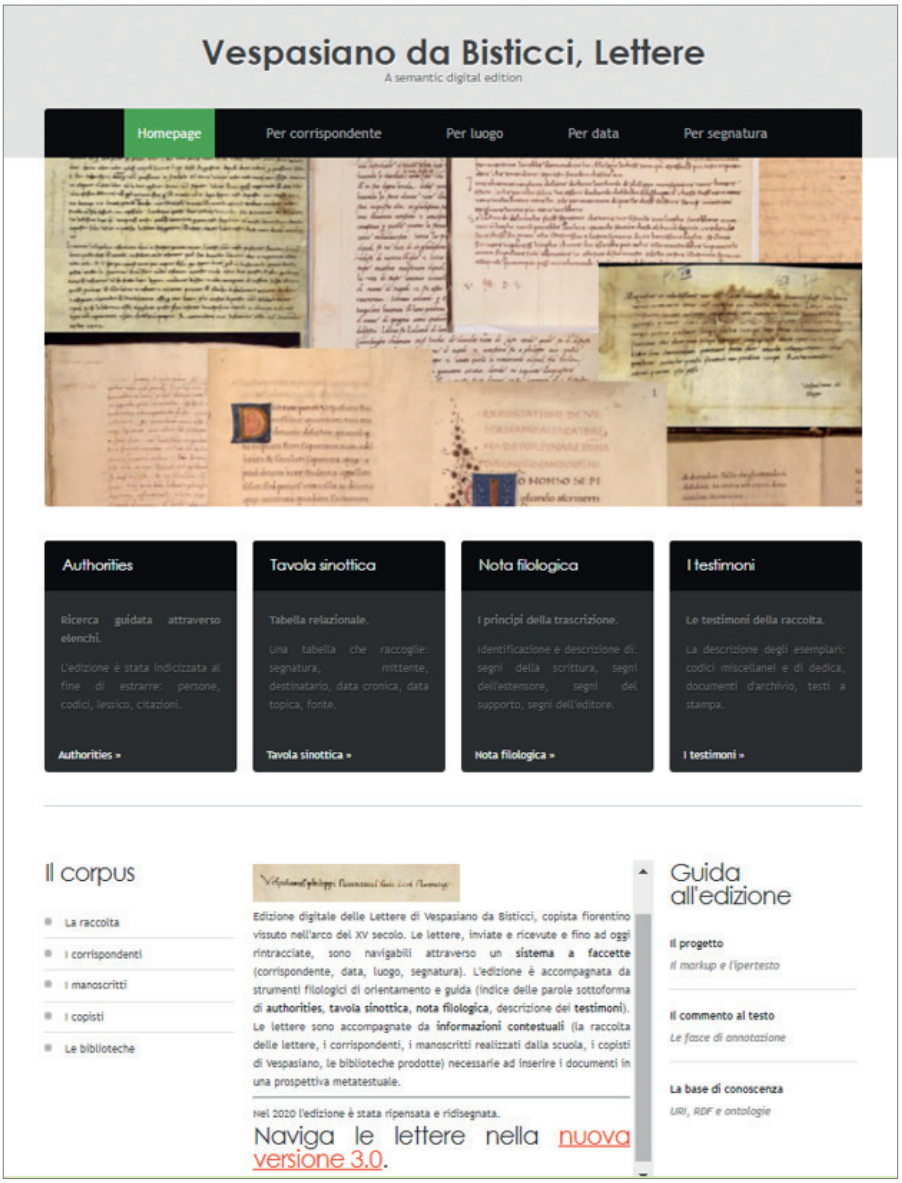

Fig. 8. The digital edition of a work, carried out as part of a Digital Humanities project, represents, for the purposes of the study, even with all the specific characteristics of the digital application, the equivalent of one or more critical essays and as such should be treated in the context of cataloguing 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/#)

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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.

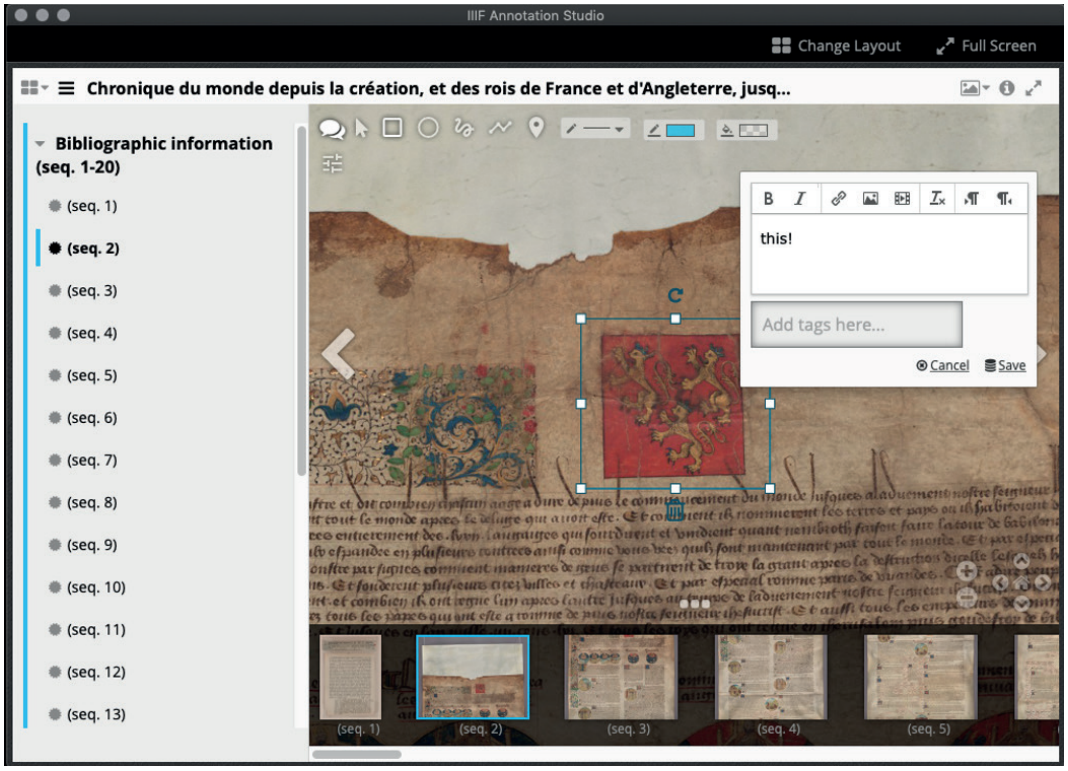

Fig. 9. Users can comment on, transcribe, and draw on image-based resources

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

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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 Lankes 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 FRSAD), 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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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 metadatazione, 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).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

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. Australia[8] and Germany[9] 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".

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12743

References[11]

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 metadatazione. 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.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

JLIS.it

The National Library of Norway – policies and services

Oddrun Pauline Ohren[(a)]

a) National Library of Norway

Contact: Oddrun Pauline Ohren, oddrun.ohren@nb.no
Received: 16 April 2021; **Accepted:** 29 May 2021; **First Published:** 15 January 2022

ABSTRACT

The operation of National Library of Norway (NLN) is governed by the Legal Deposit Act of 1989, latest amendment in 2015. By this law, all documents of any type made publicly available in Norway, must be provided to the National Library for registration, preservation and dissemination. According to an added regulation in 2018, NLN may also require the digital version of printed documents, as well as core metadata.

Another important policy document is issued by The Ministry of Culture and The Ministry of Education and Research, outlining a library strategy for the period 2020-2023. While including all types of libraries, the strategy has a strong focus on NLN as a driving force and service provider for the rest of the Norwegian library sector, in mandating NLN to support other libraries in a number of ways, - financially through funding development projects, structurally by way of providing crucial infrastructure and developmentally by conducting our own innovation activities.

The national bibliography forms the backbone for many of the infrastructure services, like the *future Metadata Well*, constituting one single authorized source of metadata for Norwegian libraries, various authority *files*, as well as several *thematic bibliographies*. It also lies at the heart of enabling end users to access the vast collections of digitized material, even much of the IPR-restricted material, obtained through deals with rightsholder associations.

KEYWORDS

National libraries; Governance; Library services; National bibliographies.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

Introduction

A key role of most national libraries is to collect, describe and preserve everything that is published in a particular country. The exact interpretation of "published" and "everything" may vary among countries, – IFLA National Libraries Section expresses its overall goal as "supporting the vital role of national libraries in society as custodians of the worlds' intellectual heritage, providing organisation, preservation of and access to the national imprint in all its forms" (IFLA National Libraries Section 2015). However, during the recent years, openness – both in terms of access to collections and physical space, is seen as an increasingly important value in academic libraries (Anderson et al. 2017, 14, Larsen 2017, 52). This trend is thoroughly embraced by the National Library of Norway (NLN), and is also clearly demanded by the Ministry of Culture. NLN's work on openness first and foremost applies to the collections, and the major premise for that is the mass digitization activities since 2006. Nonetheless, and in spite of a somewhat austere-looking (listed) building, there has also been put strong focus on welcoming students and researchers as well as the general public into the library building, be it for studying, socialising with colleagues and friends or participating in some event.
The following is an account of the NLNs approach to fulfilling its mission.

Governance

The operation of National Library of Norway (NLN) is governed by several policy documents, each with their separate time horizon, from the long term Legal Deposit Act (Norway. Ministry of Culture 2015), via a medium term national Library Strategy (Norway. Ministry of Culture and Norway. Ministry of Education and Research 2019) to the yearly Letter of allocation. The regulatory documents and their influence on NLN's internal strategies and operations are described in more detail below.

The Legal Deposit Act

The Legal Deposit Act of 1989 represents the very "raison dêtre" for NLN, as it defines a stable, very long term basis upon which to construct the national library organisation and its operations. By this law, all publishers, producers or importers of documents made publicly available in Norway, are responsible for providing those documents to the National Library for preservation and dissemination.
The Legal Deposit Act applies to 'any' kind of documents, both physical and digital, on any media – and has done so since 1989. However, through an amendment in 2015 together with an added regulation in 2018 two important changes were stipulated.
Firstly, NLN now may also require the digital files from which the published documents are produced (e.g. pdfs used for printing books), and secondly we may require some core metadata with the deposit. The national library for its part is responsible for creating/enriching the metadata to a level befitting a national bibliography, as well as managing the catalogue and catalogue products.
The Norwegian national bibliography covers several types of materials: Monographs/books, pe-

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

riodicals, recorded music published on physical carriers, sheet music, articles in periodicals and resources related to the Sami population in Norway.
These two updates to the law represent great opportunities for streamlining the material flow, and thereby getting the content out to the public faster.

The National library strategy 2020-2023

Another important policy document is *the National strategy for libraries 2020-2023* issued by The Ministry of Culture and The Ministry of Education and Research, outlining a four-year strategy comprising all types of libraries, but with strong focus on NLN as a driving force and service provider for the rest of the library sector. Commonly referred to as *The Library Strategy* for short, and with its emphasis on active dissemination, it mandates NLN to support the other types of libraries in their endeavours, – financially through funding development projects, as well as structurally by way of providing crucial infrastructure to the libraries.
In the words of the strategy document, libraries should develop into "a space for democracy and self-cultivation" (Norway. Ministry of Culture and Norway. Ministry of Education and Research 2019, 3). Moreover, it is points out very strongly that the national library is the government's main instrument and driving force to achieve this, listing a number of duties and tasks for the national library.
The "infrastructure services" to be provided by NLN to other libraries, may be subdivided into 3 groups.

1. Content: The strategy's strong focus on dissemination naturally implies a requirement to provide content, as much as possible digitally, but also physically.
2. Metadata: To administer the content, bibliographical data as well as authority data of good quality are needed.
3. Library tools and guidelines: Part of NLN's duties is also to function as national competence and resource centre for other institutions in the library and cultural sector. This involves keeping up with the development within library science in general and in the bibliographic domain in particular, and provide useful standards, tools and guidance for the whole sector.

Both (digital) content, metadata and authority data should be adapted for machines and humans alike, and the access mechanism must be easy to understand, well documented and readily available for libraries as well as third parties.
Lastly, the guiding principle is that infrastructure services (content, metadata, standards, etc) that NLN provides to the library sector are to be free of charge for end users as well as libraries, – or as cheap as possible.
Summing up, the main message to NLN from the Library strategy is the responsibility to actively disseminate and expand its own collections in various ways on many platforms, to users inside and outside the libraries. Equally important is NLN's obligation to enable other libraries to offer high quality services to their local patrons. To achieve this, NLN shall develop shared, national infrastructure, to make sure that local, often thinly staffed libraries can spend their time and effort to serve their local patrons, not on work that could just as easily be performed centrally at a national level.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

The letter of allocation

This document, issued by the Government after its yearly process of negotiating the national budget, defines NLN's total budget for the year in question, along with any specific areas to focus on, sometimes accompanied by dedicated funding.

The Public Library Act

This law (Norway. Ministry of Culture 2013) stipulates that all municipalities in Norway must offer a public library to its citizens. NLN's role is to enforce the law, in particular the paragraph about competence, instructing each municipality to hire a library manager educated in librarianship.

Content creation and dissemination

The material acquired through legal deposit forms the core of the library's collection, although other material is purchased, in particular documents published abroad which is relevant to Norwegian affairs (Norvegica Extranea). Handling deposited material efficiently is the task which forms the foundation for everything else, and the task that is our sole responsibility. Making the content accessible to users, also involves documenting it in terms of structured metadata, which ultimately forms the national bibliography. Hence, the topic of bibliographical control lies at the heart of the whole process of receiving and processing legal deposit.

Through an extensive digitization project since 2006 – at present about 600 000 books are digitized, practically all the books in our national bibliography. Also, about 60 % of NLN's historical newspaper collection is digitised, and all current newspapers are deposited digitally as well as in paper.

In addition to handling deposited material, NLN also creates content itself, mainly in some way based on the collections. Among these are digital productions like podcasts and streamed events, as well as research-based publications, theme-based bibliographies and re-publications of older literature.

As already mentioned, the national Library strategy focuses very strongly on dissemination: *"… The goal is for libraries to introduce new users to literature and reading, facilitate knowledge dissemination and expand digital collections. The government will implement strategic measures that support libraries and librarians in attracting more users, including those who do not visit libraries.*" (Norway. Ministry of Culture and Norway. Ministry of Education and Research 2019, 3)

A large proportion of our physical collection is digitized and therefore – technically – available anywhere through NLN's digital library nb.no[1] and through the main catalogue discovery service Oria. An overall goal is that as much as possible of NLN's content can be accessed throughout Norway. Since much of the material is constrained by copyright, there are legal and financial challenges to be overcome. Consequently, an important part of NLN's dissemi-

[1] https://www.nb.no/en/the-national-library-of-norway/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

nation strategy is to negotiate agreements with IPR holders about exposing digitized material still under copyright. At the time of writing, digital books published before 2001 may be read on any device with a Norwegian IP address. The clear message from the Ministry of Culture is that NLN should try to overcome more IPR obstacles, so that books newer that 2001 can be accessed anywhere in the country. Hence new negotiations with the publishers and their organisations will be opened soon.

While great resources are put into maximising the digital content that may be accessed directly by end users, direct availability for end users is not possible for everything. Another agreement with publishers enables patrons of local libraries to access all deposited material from within the walls of their local library, provided it is used for documentation or research. This includes all digitized newspapers and books – also those newer that 2001. Through a special agreement with about 70 running newspapers, any library visitor may read them from 2 weeks after publication onwards.

To provide the same democratic access to physical material for all inhabitants of Norway is no small challenge, Norway's geographical and demographical conditions being as they are. Norway is the 3rd least densely populated country in Europe[2], and the distance from south to north almost spans the whole continental Europe. At the same time, all municipalities are mandated by law [publ act] to offer a public library, however small and however few people live there. Hence, interlibrary lending is by necessity an important element of the library services. To support this, NLN has provided a service called *Biblioteksøk*[3] ('Library search'), a joint discovery&lending service covering the holdings of all Norwegian libraries. Through Biblioteksøk users may reserve books held by any library in the country, and pick it up at their local library.

The NLN Depot library, containing almost a complete set of Norwegian printed monographs, journals, and newspaper microfilms is by far the largest supplier to interlibrary lending. It is built up from legal deposit 'leftovers', transferred material from other libraries and some purchase. The actual lending process is handled by an efficient automatic storage facility. Thus, the Depot library greatly decreases other libraries' burden that interlibrary lending usually represents.

An important part of the Depot Library is the Multilingual Library Collection. Many of the small libraries in Norway, struggle to be able to offer books to their immigrant population. *The multilingual library*[4] is a service for public libraries designed to remedy this to some extent. Its collection comprises literature acquired from a multitude of countries in equally many languages. While being available for all through ordinary interlibrary loan, its main purpose is to support libraries in providing services to their multilingual and multicultural population. Any public or school library may borrow 'mini-collections' or 'book-cases' composed according to their own requirements to be used as their own holdings for a period of up to 6 months.

[2] https://www.worldometers.info/population/countries-in-europe-by-population/
[3] https://bibsok.no/
[4] https://dfb.nb.no/multilingual-library

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

Research data – The language bank

In order to develop high-quality language technology for Norwegian, big datasets with Norwegian speech and text are needed. Since Norwegian is a very small language in terms of speakers, we cannot rely on others providing such resources. Consequently, NLN Language Bank has as its primary task to provide and organize such data sets. Our resources are aimed at researchers and students, as well as commercial companies developing language technology software. The resource collection comprise among other things, lexical resources like wordnet and dictionaries, corpora of written and spoken language, i.e. large collections of text and speech in machine-readable format. The language bank is NLN's main contribution to the shared European research infrastructure called Clarin[5].

All language resources are available online via the National resource catalogue[6], which also includes resources from other Clarin centres in Norway. The metadata here follows the Clarin-defined framework, in which profiles and subcomponents can be defined, understood and reused across Clarin centres.

Closely associated with the Language bank is Digital Humanities Laboratory, a service supplying scholars, students, and library users with digital tools and methods in their studies, as well as assistance in their use.

The Norwegian national bibliography and other bibliographical services

The NLN is responsible for developing and maintaining an online national bibliography, holding the view that a national bibliography is not merely a tool for information retrieval but is in itself a rich source of insight into a nation's cultural heritage and intellectual production, and as such constitutes valuable research data within many fields of study. While defined by Parent (2008, 10) as "a *current*, *timely*, *comprehensive* and *authoritative* list of all titles published in a country", NLN also includes titles published abroad by Norwegian agents or about Norwegian affairs.

Some parts of the Norwegian national bibliography are still managed in separate legacy databases, but its main portion – along with other, thematic bibliographies – reside as virtual subsets in the main catalogue shared with about 80 other academic libraries.

The whole national bibliography can be accessed through a separate instance of the discovery interface, the discovery service of our main catalogue. The bibliographical data are also freely available via OAI-PMH in MARC 21 format and Dublin Core, to be used for research purposes as well as anything else.

Below is the access page for the Norwegian national bibliography at the search & discovery interface to the catalogue. As shown in the figure, it is subdivided into the subsets of books, serials, musical recordings, sheet music, Sami publications and articles in Norwegian and Nordic publications. The parts that reside in a legacy system are Norwegian registry of serials, Index to articles in Norwegian and Nordic periodicals as well as older parts of the Registry of Norwegian printed sheet music.

[5] Common Language Resources and Technology Infrastructure: https://www.clarin.eu/
[6] https://www.nb.no/sprakbanken/en/resource-catalogue/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

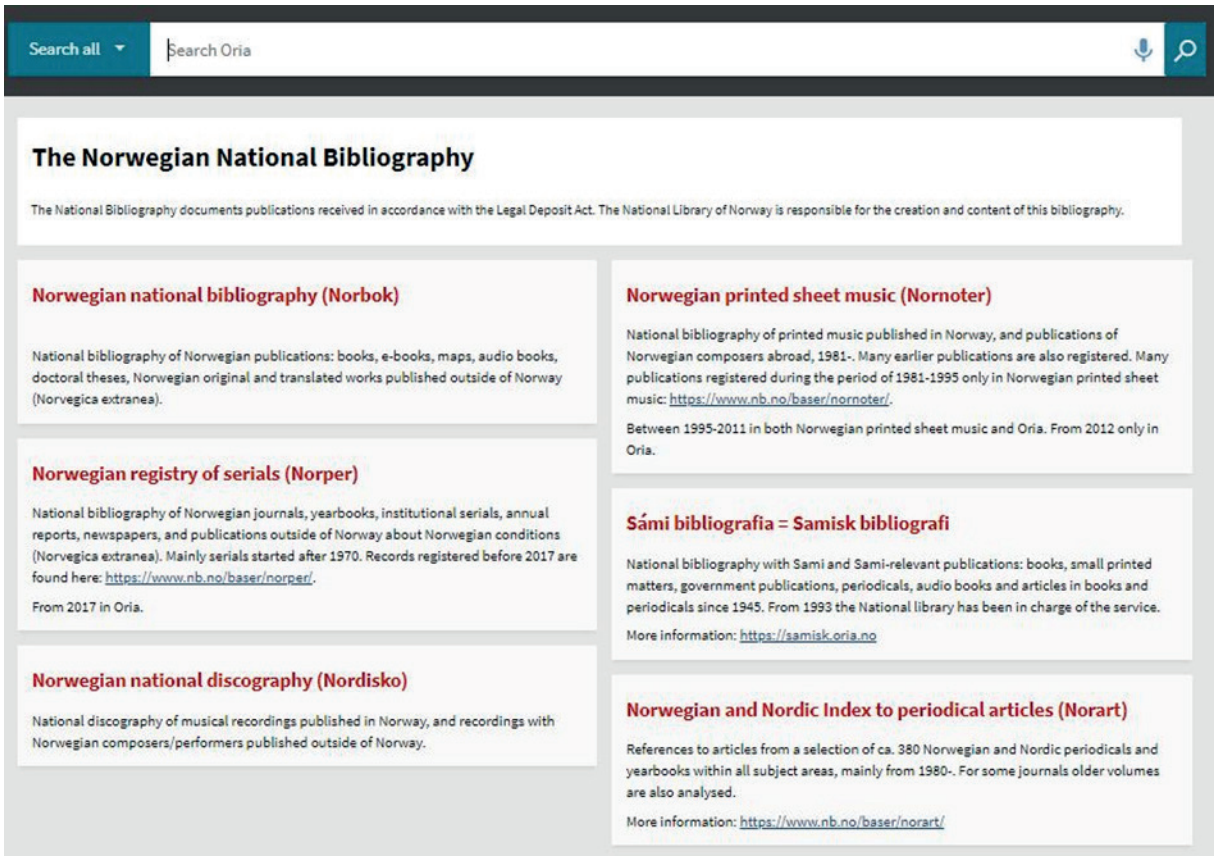

Fig. 1. End user access page for the Norwegian national bibliography

Authorities

Connecting bibliographical data to authorities is an important measure to maintain a certain degree of consistence and reliability in the metadata. So far we have a fairly large authority file for agents (persons and corporations) as well as up-to-date Dewey via WebDewey, a genre form thesaurus published as linked data. A work authority file is in progress, partly through NLN's participation in the library-driven collaboration project Share-VDE[7].

While machine-readable access to agent authorities is provided through a REST API as well as harvesting and download facilities, human users may browse the same 2 million authorities in a dedicated search interface . The set includes all agents referred to by NLN holdings, as well as the holdings of the other academic and special libraries sharing the same catalogue. An increasing number of authorities contain references to other registries, in particular to VIAF[8] and ISNI[9].

[7] https://www.share-vde.org/
[8] Virtual International Authority File: http://viaf.org/
[9] ISNI (International Standard Name Identifier): https://isni.org/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

Metadata delivery to the library sector

The current Library strategy mandates NLN to supply free metadata to all libraries. Because prepublication metadata are important sources of acquisition for the libraries, it is important that these metadata be produced as soon as information about a planned publication is available.
NLN's internal processes can at present not guarantee such promptness, hence since 2017, metadata about Norwegian publications (printed book, audiobooks, ebooks and language cources) have been procured from commercial metadata vendors, making sure they follow NLN's cataloguing practice for the national bibliography, and include certain elements particularly wanted by the public libraries, such as subject headings from a certain vocabulary. An important aspect of this is to always authorize agents mentioned in the metadata. Using the API, it is possible to integrate Authority lookup into their own cataloguing system.
This approach will be continued and NLN will assess the feasibility and necessity of expanding to free metadata for other kinds of material such as film and music. Also, libraries have requested access to metadata for foreign material. During the strategy period, the National Library will attempt to find good solutions for including this in its deliveries.

Service overview

Table 1 presents an overview of the services NLN renders as mandated through the Act of legal deposit and the National Library strategy 2020-23, some of which are described in more detail above.
The services are grouped vertically according to type, horizontally according to target group (end user, libraries, third party).
The data services are in principle open for all, there is no difference between any person and a library when it comes to accessing and using our bibliographical data, nor other research data. Content is a different thing, for which copyright is a deciding factor for availability. Negotiating with rightsholders is typically best handled at the national level. So is maintaining cataloguing rules and guidelines, providing core tools like classification system, and providing a website (bibliotekutvikling.no)[10] supporting collaboration and information exchange, as well as serve as a knowledge resource for libraries.

[10] https://bibliotekutvikling.no/

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

Type of service \ Target group	End user	Libraries	3rd party
Outreach (Competence, governance)	Negotiating agreements with publishers about access	- Metadata standards, guidelines and tools for libraries - Funding development projects - Provide collaboration platform (bibliotekutvikling.no) - Manage Act on public libraries - Negotiating agreements with publishers about access in libraries	- Collaboration forum with system vendors - Interoperability requirements
Content	- The digital library, nb.no - The main catalogue (oria.no) - Events: Digital and on site - Podcasts, social media - Physical material	- Extended access to restricted material for patrons in local libraries - The Multilingual library - The Depot library and Library search	
Data	- Bibliographical data - Authorities: Persons and corporations, Dewey, Genre/form and other vocabularies - Research data: Language resources	- Bibliographical data - Authorities: Persons and corporations, Dewey, Genre/form and other vocabularies - Research data: Language resources	- Bibliographical data - Authorities: Persons and corporations, Dewey, Genre/form and other vocabularies - Research data: Language resources

Table 1. Overview of services from NLN according to type and target groups

As evident from Table 1, producing the national bibliography is only a part of the goals and tasks of the national library, yet bibliographic work forms the basis of most of the other activities and services. For example, many of the events, podcasts and other dissemination activities are directly based on objects in the library's collections. Finding and selecting the right objects in each case requires rich and reliable bibliographic data, describing the objects according to several criteria, like chronology, provenance, topical coverage and physical attributes, among other things.

Challenges and future work

Along with the growing emphasis on dissemination and 'opening up' the library to the general public, comes decreased willingness to spend human resources on cataloguing and related activities, and also stricter demands to justify the usefulness of the particular data elements that are produced. This challenges us to find ways to produce metadata more efficiently with fewer staff, yet maintaining the quality and richness. NLN approaches this from various angles:
Firstly, streamlining the processes handling legal deposit has high priority. The relatively new legislative basis for requiring simultaneous deposit of printing file, printed book and (some) metada-

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

ta, offers great opportunities for streamlining the deposit workflow, not least because the need for manual handling of the printed books is greatly reduced. Realising the benefits of this is ongoing work, and is expected to be ready for trial in a few months.
Another potential gain is the possibility for automatic or machine-supported creation of descriptive, based on text analysis of the printing files. Although no concrete action is taken, this will be looked into further down the line.
Machine learning is a type of technology that is starting to gain popularity also in the library universe, especially for subject indexing and named entity recognition, as exemplified in (Suominen 2019). Currently, NLN is in the early stages of experimenting with automatic classification, using the pre-trained BERT model (Horev 2018) for Norwegian.

The Metadata Well vision

One of the most demanding tasks assigned to NLN and its system partner UNIT[11] by the Library strategy is perhaps to establish a so-called 'metadata well': "A further step towards the goal of 'one book, one catalogue entry' will be to create a single authorized source for the metadata – a metadata vault" (Norway. Ministry of Culture and Norway. Ministry of Education and Research 2019, 32). At the time of writing, an RFI document[12] for the Metadata Well is being prepared. Although still very much on the conceptualization and planning stage, the Metadata Well may be thought of as 'an authority file for bibliographical descriptions', in which all metadata produced for the Norwegian National Bibliography are included. So will metadata contributed by other authorized contributors. In its ultimate state of completion, the Metadata Well should contain bibliographic data describing the union of collections in all Norwegian libraries. It is not perceived as a union catalogue, hence no holding data will be included. See also Figure 2 for visualization of the system.

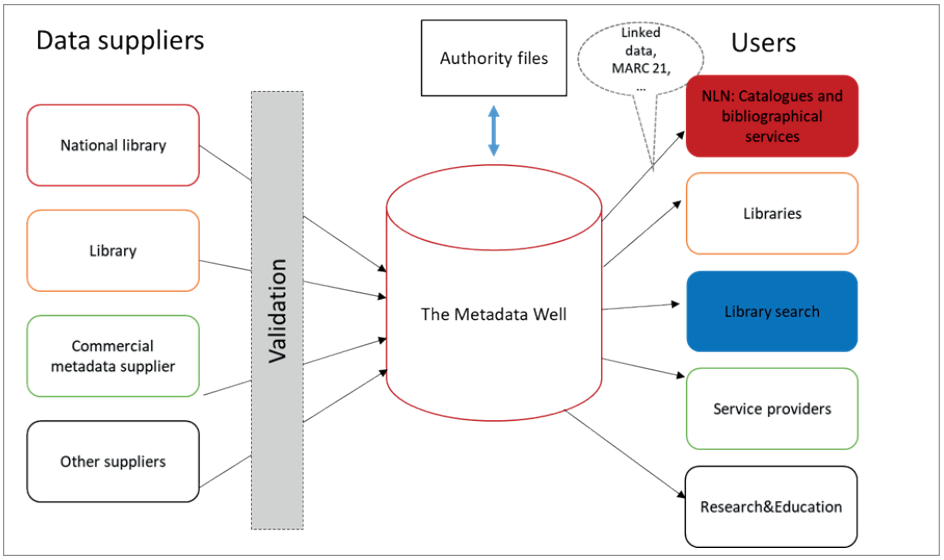

Fig. 2. The Metadata Well in context

[11] UNIT Directorate for ICT and joint services in higher education and research (https://www.unit.no/en).
[12] Request for information.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

With the Metadata Well in place, libraries can obtain bibliographical descriptions for their local catalogues, either by referring to it or by copying it to their own catalog, all the while retaining its globally unique identifier in their local data. In their catalogue, they only need to add local information.

Hopefully, this resource will constitute a hub for reuse of metadata between libraries, and as such function as a major source for resource sharing among all types of libraries in Norway, including public, school and academic libraries.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12741

References

Anderson, Astrid, Cicilie Fagerlid, Håkon Larsen, and Ingerid S. Straume. 2017. "Åpne forskningsbibliotek. Innledende betraktninger." In *Det åpne bibliotek: Forskningsbibliotek i endring*, edited by Astrid Anderson, Cicilie Fagerlid, Håkon Larsen and Ingerid S. Straume. Oslo: Cappelen Damm Akademisk.

Horev, Rani. 2018. "BERT Explained: State of the art language model for NLP." accessed April 15. https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270.

IFLA National Libraries Section. 2015. Strategic Plan 2015–2017.

Larsen, Håkon. 2017. "Aktivering av nasjonens hukommelse: Nasjonalbiblioteket i offentligheten." In *Det åpne bibliotek: Forskningsbibliotek i endring*, edited by Astrid Anderson, Cicilie Fagerlid, Håkon Larsen and Ingerid S. Straume. Oslo: Cappelen Damm Akademisk.

Norway. Ministry of Culture. 2015. *Lov om avleveringsplikt for allment tilgjengelege dokument (pliktavleveringslova)*. Oslo: National Library of Norway.

Norway. Ministry of Culture. 2013. Lov om folkebibliotek (folkebiblioteksloven). Oslo.

Norway. Ministry of Culture, and Norway. Ministry of Education and Research. 2019. A space for democracy and self-cultivation. National strategy for libraries 2020–2023. Oslo.

Parent, Ingrid. 2008. "The Importance of National Bibliographies in the Digital Age." *International cataloguing and bibliographic control : quarterly bulletin of the IFLA UBCIM Programme* 37 (1):9-12.

Suominen, Osma. 2019. "Annif: DIY automated subject indexing using multiple algorithms." *LIBER quarterly* 29 (1):1-25. doi: 10.18352/lq.10285.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

JLIS.it

The Italian National Bibliography today

Paolo Wos Bellini[(a)]

a) Biblioteca nazionale centrale di Firenze, http://orcid.org/0000-0002-5439-1364

Contact: Paolo Wos Bellini, paolo.wosbellini@beniculturali.it
Received: 7 April 2021; **Accepted:** 7 September 2021; **First Published:** 15 January 2022

ABSTRACT

The statistics on the records produced for the Italian National Bibliography (BNI) in the last decade evidence a stable development with a growing trend. In the face of that, there has been a decrease of human resources never seen before in the history of the National Central Library of Florence (BNCF), that has drastically reduced the editorial staff of BNI to few units. The institutional tasks of BNCF, provided by law, have not changed though. Among these is the archival function for the Italian bibliographic production and its representation through adequate cataloguing and bibliographic instruments.

Therefore, in order to either maintain constant or increase BNI from a quantity point of view, by preserving its quality, some variations of a technical and organizational-managerial nature have been recently implemented, pursuing the following:

1. Rapidity of cataloguing; 2. Full implementation of the recommendations of the Central Institute for the General Catalogue of the Italian Libraries and for the Bibliographic Information (ICCU) as regards, for instance, cataloguing regulations, use of the codes of bibliographic qualification, creation and management of the authority files regarding personal names and uniform title; 3. Constant attention to the role of BNCF in the cooperation within the National Library Service (SBN).

There are plans to intervene in some critical difficulties that still remain. The emergency due to Covid-19 pandemic has imposed a de facto meaningful and unpredictable reorganization of the work management (from the legal deposit to BNI and to the BNCF catalogue) through methods that could certainly be maintained even in future.

KEYWORDS

National bibliographies; Italy; National Central Library of Florence.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

1. Introduction and context

Due to an unprecedented reduction in resources, for years, Italian State libraries have been affected by a deep crisis. This also applies to the National Central Library of Florence in its many branches, including of course the sector where the Italian National Bibliography (BNI) is developed. The group of cataloguers that process BNI is by now reduced to a few staff units. These people, while being indeed few, are very competent and skilled, thanks to a very long professional traineeship carried out in close cooperation with those who made the history of the library science in our Country, such as famous librarians, who, even if retired, often continue to offer, either directly or indirectly, their collaboration. The knowledge of this little group of highly motivated persons draws, day after day, a new vital sap from the job itself, by comparing them in a challenging way with living and always changing reality of the publishing production that these people are called to represent so as to benefit the national and international community of the users.
Indeed, despite the undoubtedly dramatic and even incredible staffing situation the institutional tasks of BNCF, determined by law, have not changed. Such are the tasks that characterize the national libraries and, among them, is thus the establishment and the maintenance of the national bibliographic production's archive. An additional task is also the representation of the national bibliographic production through appropriate cataloguing and bibliographic tools, something which no country wants to renounce or to independently administer through its own national bibliographic agency, that is supported everywhere with adequate human and financial resources.

2. Timeliness

The aspect to which we pay more attention is the timeliness of the cataloguing of the books that the publishers send to the legal deposit office and that are selected, in order to be included into BNI. The promptness of the cataloguing plays a decisive role on the users' fulfilment (whichever group they belong to). This is why it has an absolute priority; yet, among the reasons why it is object of particular interest, is also the need not to create too many bottlenecks and, above all, arrears, within the circulation of the books processed in the library.
Cataloguing at BNCF, and, therefore, also at BNI, is directly implemented through the net of the National Library Service (SBN) by using the software SBNweb (https://opac.sbn.it/opacsbn/opac/iccu/free.jsp). BNCF is one of the 6590 libraries that participate in the SBN net, and so the records processed for the catalogue of both BNCF and BNI also feed the catalogue of SBN and are immediately visible to other libraries of the net as well as, after a short time, in the OPAC of SBN. The records processed for BNI are visible within 24 hours even on OPAC of BNCF, which, in its new version, has been issued only a few months ago (http://opac.bncf.firenze.sbn.it/bncf-prod/).
The selection of the books for BNI in the premises of the legal deposit is carried out once or twice a week. After the selection, the volumes are moved from the legal deposit to the offices of BNI in order for them to be catalogued; normally this happens in a very short time. The availability on the online catalogue of the selected items for BNI is therefore almost immediate. It's worth noting that the bibliographic items created *ex-novo* by BNCF within the net SBN are numerous, which witnesses how quick the book cataloguing on behalf of BNCF is. Just to give an idea, consider that the records of new creation on behalf of BNCF are currently about 784,000 out of 18,320,000

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

present in SBN. These items are not represented by the bibliographic records of the documents stored in both SBN and BNCF's libraries, which are obviously many more (since BNCF receives everything that is published), but only by the items that BNCF has created first, compared to other libraries of the net.
The important thing is that within about ten days the books selected for BNI are catalogued and available on OPAC BNCF and on OPAC SBN. There are no arrears.
Yet, in order that the books described by BNI, already visible on OPAC, may show even the semantic contents they were assigned by using the Decimal Dewey Classification and the *Nuovo soggettario*, more time is needed. Indeed, as is well known, the subject indexing is a particularly onerous activity, even if no international bibliographic agency renounces it.
The delay, with which the information of semantic genre of the books catalogued in BNI is shown, is currently equal to a few months, but it is planned to eliminate it shortly with a specific project.

3. Coverage

The other aspect that has a huge importance for every national bibliography and, of course, for the users is thoroughness, that is to say, the quantity of books included every year in the bibliography in relation to the publishing production of the Country.
Books published in the current year and those published up to two years earlier are catalogued within BNI, so in the BNI year 2020 books published in 2020, 2019 and 2018 are included.
Books, which, due to their characteristics, should be indicated in BNI but are not sent by publishers to the legal deposit within said terms, are catalogued by the cataloguing office of BNCF. That means that a special series of BNI files for the related publications does not exist.
In the last five years the number of bibliographic items issued for BNI, referring only to monographs, has been in average about 12.000 per year. This do not include dissertations, periodicals, printed music and other material.
The basic question is how to assess the production of BNI in terms of completeness with respect to the Italian book production. In other words, how representative BNI is as for that. Yet, the concept of bibliographic coverage is therefore not so strict and has to be carefully assessed, in relation to the different typologies of publishing production.
A way to assess these data is to compare them with the statistical data on the Italian publishing production. According to the ISTAT data on the book works published in Italy, in 2018 the Italian publishing houses have published 75,758 titles (https://www.istat.it/it/dati-analisi-e-prodotti/banche-dati/statbase).
This is the number of books to refer to, but, actually, the publications excluded from BNI have always been numerous. In fact, BNI is a selective bibliography. There are several kinds of publications that normally are not reported in BNI. Such publications are of course catalogued by BNCF but only in the catalogue of BNCF (and they are therefore present even on SBN) or they are handled by sets.
The complete list of such publications, not included in BNI, is the following:

1. Official publications of public administrations and international organizations, unless they have an autonomous monographic character.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

2. Non-official publications of laws, decrees, regulations, work contracts, etc., without any comment or addressed to particular categories of readers. The sole exceptions are certain collections of specialized publishers.
3. Publications of parties, unions, chambers of commerce, cultural and religious associations, etc. not of general interest;
4. Pastoral letters and other official documents of religious authorities;
5. Minor religious publications;
6. Consumer literature and reissues of romance novels;
7. Publications not addressed to commerce, but disseminated outside the normal channels of sale, in the form of subscriptions, enrolments, etc.
8. Reissues (unless the publication of reference has never been described), pre-prints, *specimina* and the likes.
9. Complimentary books or gifts, if reissues of previous publications, even if presented in different packages;
10. Manuals and texts for nursery schools, primary schools of first and second level;
11. Biographical scripts for limited use, of either occasional or godly character;
12. Almanacs and the likes of limited interest;
13. Patents;
14. Excerptions, even if presented in only one series;
15. Catalogues of trade fairs and shows prevailingly of commercial interest, catalogues of private galleries, house programs, tourist material;
16. Editorial catalogues and antique catalogues for non-historical or scientific purposes;
17. Publications of promotional and commercial nature, unless they represent the sole or main source of information in special fields, such as complete catalogues of stamps, coins, art objects;
18. Printed music, described in the two half-yearly dedicated files;
19. Texts of lessons and similar materials, in case it clearly appears that they are not addressed to the external dissemination;
20. Speeches and interventions connected to particular events, separately published;
21. Cartographical material.

The material not included in BNI due to choices of publishing policies is thus a lot.

Still in relation to 2018, it is possible to select some typologies of publications, and, by referring to the ISTAT data, it is also possible to verify the relevant quantity, just to give a concrete idea of the numbers we are talking about, without going into the details, which, in this case would be too many.

Schoolbooks, for instance, are 9,786 and children's books are 6,440. Thus, the remaining 59,332 publications, which, aside from reissues and reprints (that are however almost always included in the BNI), decrease to 46,718 publications.

In addition, there are multiple publications, like for instance, text books for primary schools (243), cookery books and recipes (396), books classified as 'entertainment', games and sport (1,004), tourist guides (422), adventure books and detective stories (4,328), comics (713) and others of non-specified genre (1,672) for a total of 8,778 publications. We fall to 37,940, a number still much higher than what remains once selected the afore-said listed items.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

Such list is useful only to give an idea of what numbers we are talking about, but many books belonging to such ISTAT categories are actually included in the BNI (for instance, but not *in toto*, cookery books, tourist guides, detective stories, adventure books and many others).

All this to say that, once eliminated all these typologies of materials, both, in part, reissues and reprints, the remaining books still to be catalogued for BNI are actually those that are catalogued, and that the coverage of BNI is good and representative. Of course, according to the historically adopted criteria, which, although selective, are also fully similar to those adopted by the operators of the same sector, working for both national and foreign libraries. Said operators, while making selections of qualitative type on the material to be reported, do not catalogue at the same level than BNI and are not part of the SBN net. They often resort, in a large percentage of cases, to editorial announcements, which is fine, though for different purposes. In any case, it must be declared and quantified otherwise we shall not supply the users with a quality service.
To end up with this important matter, it must be noticed that many other typologies of books, like all those excluded according to the above-mentioned criteria, not recorded within BNI, are catalogued in BNCF. Just to provide an example, in 2019 the office for the cataloguing at BNCF has catalogued about 24,500 monographs that, summed up to almost 13,000 of BNI, bring the total number of catalogued monographs in 2019 to 37,500, which is a number of all respect if compared to the data provided by ISTAT on the annual publishing production in Italy.

4. The human resources

In order to face the more and more serious and chronic lack of staff, some measures have been taken in these last years:

a) Procurement contracts for the cataloguing to external bodies.
The procurement contracts for the cataloguing to external bodies is a largely known and practical solution. This has both positive and negative aspects. On the one side, such solution guarantees both flexibility and the possibility to 'close up the leak' immediately. On the other side, there is the demand of the quality control that involves BNI at a higher level, even if it obviously involves all; it especially involves BNI because of its role as an Italian bibliographic agency and for the fact that all the books indicated in BNI are catalogued in the SBN net at a "super" level (i.e. level 95), including items marked by the higher authority code. Such records cannot be modified but by the central Institute for the General Catalogue (ICCU) of the Ministry of the cultural Heritage. Whereas mistakes are always possible, said records should not contain any: in other words, they have to meet the requirements of greatest authoritativeness (and here we come to the third key term that must characterize a national bibliography in addition to timeliness and thoroughness/coverage). Both enterprises and cooperatives operating in this field ensure a good level of cataloguing, yet, not always the very high level of specialization required for cataloguing in BNI and that implies heavy investment in staff training.
This is why we must pay a special attention to the control of the correctness of the cata-

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

loguing in BNI and, above all, to the alignment of the whole staff. While neglecting additional details, I would like instead to underline that it is not sufficient that the institution may control the plan and the total structure of the service if assigned to external collaborators. On such subject I would like to underline how, due to the reduced permanent staff at BNCF, the quality control subtracts resources in an unsustainable measure and poses a first insuperable limit to the quantity of work that is possible to outsource.
The other limit arises from the maximum fee to be provided for, as established by the Procurement Code.
All this basically means that each tender cannot last more than one year and that, after this term, a new procurement procedure must be performed; in addition, it means that the external staff, in part or completely, changes and that the training work for the external collaborators is each time fully frustrated.
The Covid-19 pandemic has added a further difficulty, by decreasing the number of persons who can work in co-presence in the same premises and thus obliging to temporarily suspend the external collaborations and to seek difficult management solutions.
In spite of the recent important re-adaptations of the rooms to be addressed to the storages, logistic issues that affect BNCF, as well as the scarcity of space in which catalogued books are to be stored, are a further obstacle to the research of viable organization solutions.

b) Cooperation with specialized libraries
Still to cope with the lack of personnel, another important initiative recently taken in BNI was to seek the collaboration with other libraries, somehow similar to BNCF, for the cataloguing of the books reported in BNI. At the distance of a few years, since the beginning of this collaboration, it has been possible to draw up a first balance. It is about a very positive and rewarding experience, which is fully a part of the cooperation and collaboration spirit that characterizes SBN itself. The contribution of the institutes that participate in this activity is however changeable and, in its whole, in a very low percentage. In addition, even in this case, it deals with a very demanding coordination and verification job.
To end with the topic concerning the human resources, every effort has been made to adopt measures of both procedural and organizational engineering to face the above-mentioned dramatic shortage of personnel.

5. Relationships with SBN and library cooperation

The collaboration with SBN is a fact of huge and positive importance for both BNCF itself and the other libraries of the net, because the institutions take mutually advantage on the common job by allowing a considerable saving of working time.
Since the cataloguing in BNCF occurs solely through 'books in hand' and in BNI at the highest level of authority provided by SBN, it can scarcely be said how BNCF does a considerable maintenance job for the catalogue of the National Library Service through the revision of the records processed by other libraries and through the interventions on all the components of the network that gives rise to the card itself.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

One of the decisions recently made was to progressively eliminate every discrepancy out of the application of the rules and of the cataloguers' usage between BNI and ICCU, not only to minimize the needs of intervention on the captured records but also in consideration of the positive value which the uniformity of the choices has in a shared catalogue.
The interventions that are carried out on the records 'captured' by SBN for BNI are very many and I do not fear to exaggerate by assuming that on about 80% of the capture cards it is necessary to intervene more or less significantly.
Listed below are the most frequent and meaningful interventions that are sufficient for me to mention:

- *Ex-novo* publication or updating of the authority file of the personal names with reference to all the books catalogued for BNI which are thus many more than the produced cards.
- In the authority file the codes of the Countries and languages are always enhanced (the code of languages is unfortunately absent also in great part of the cards at a level 97) and the fields "Dating", "Information note", "Sources" and "Cataloguer note" are always compiled through all the proper researches, according to what provided by ICCU guidelines, as for the drawing up of the authority files;
- All data reported in the authority files are monitored on the bibliographic directories normally in usage, such as – for instance – the national bibliographies of various Countries, the catalogues of great libraries, biographical dictionaries, encyclopedia, both national and international authority files, etc., and the connection is executed in all provided and applicable cases;
- If available, the code ISNI is added. In this regard, I remark that in ISNI numerous duplicates are present. When two ISNI numbers are attributed to the same entity, it is not simple to decide which one has to be included in the registration of the name that is being processed. Therefore, comparisons on VIAF shall be carried out, and whenever more incidences of the same name, complemented with an ISNI number, are found even in VIAF, the incidence to which more libraries are connected, and/or the most appropriate one, will be chosen;
- A very remarkable chapter is that of the link to the uniform title, which also constitutes a recent entry for BNI as well as the BNCF catalogue. Until last year, such link was created only for the translations into Italian of foreign works or for ancient works. At present, in full compliance with the FRBR scheme, we follow the SBN regulations, which include the creation of a uniform title in any case. The decision to connect a uniform title for all catalogued publications (and subsequently trans-codified into BNI), in compliance with what provided by ICCU and following the Italian Regulations on the cataloguing for authors (Reicat) for the choice, has definitely caused a remarkable burden of work and an extension of the cataloguing times. It is furthermore important to notice that the application of the rule in SBN is often carried out in a mechanical way, so that the uniform title adopted for the Italian publications, not in translations, always replicates the title itself, by often reporting even the complement of the title, something apparently not always correct. Recently, it has not been rare that uniform titles of foreign works translated into Italian were the Italian title itself. This also implies the need to intervene often in the corrections, fusions, additions of links to variable uniform titles, and so on. It would definitely be preferable that

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

ICCU organized other courses especially addressed to the registrations of the authorities. Indeed, the application of the guidelines is not certainly something to be executed though automatisms. Suffice it to think of the complexity of the qualifications and other elements used to distinguish identical titles.

- As per the authority files, BNI participates in the ICCU working groups for the authority files of the names and on the uniform titles. Such interventions are executed by single users who intervene on the authority entries, not through the contextual updating of the Hub database, but rather directly operating on the collective catalogue by means of centralized working methods, that is to say within the so-called "direct interface";
- Proceeding with the reporting of the significant interventions that need be done more frequently on the records 'captured' by the SBN net, the link to an institute is created, still in case of anonymous books, or, in case the books have no main liability, if they are present in the title page, in the cover page or in another relevant position. Such is one of the most frequent interventions on the card 'captured' by SBN (because the link is almost never present), and, due to the controls they bring, said interventions are also very onerous;
- The link to an institute is created even in the case of a series with a generic title just published by that institute;
- Unfortunately, in SBN the cases of records without any element of controlled access are more and more frequent, also when created by libraries of primary importance, even when it is absolutely clear that there are either major or secondary liabilities;
- as already mentioned before, if necessary, the notes are added and, in case there are any, both their adequacy and homogeneousness are controlled;
- the codes of bibliographical qualification are indicated in all the events provided by the regulations on SBN cataloguing;
- relator codes for all the names linked to a new entry are always developed;
- as per the genre of the resource, whose indication is optional, in BNI the codes J "Biographies", S "Exhibitions" and Z "Conference proceedings" are indicated;
- The application of the code "Typology of literary text" has been included.
- The form of both the content and the kind of mediation is obviously indicated (which is indeed mandatory);
- the codes of designation of the type of support are indicated, by using the list of the codes MARC21, managed by the Library of Congress, also used in RDA;
- The link to names, even for not previously considered secondary liabilities, is carried out: such as preface authors, postface authors, and so on;
- Lastly thanks to an important innovation, the indices of the BNI subjects, starting from 2015, have been activating links to the terms of both Thesaurus and New Subject Indexing.

According to the above list, the number and quality of the interventions carried out in SBN are actually high and bring to a very meaningful increase of the granularity of the catalogue, as well as to an enhancement of research opportunities.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

6. COVID-19 emergency

The Covid-19 emergency has imposed a radical reorganization of the work. At first, without any prior experience in the sectors of both BNI and cataloguing, the method of remote working has become mandatory for everyone. The access to the necessary devices was made possible in record time by the IT staff of BNCF within 24 hours. The main problem has been the transportation of the books from BNCF to the cataloguers' homes and vice-versa. This required the willingness on behalf of both the workers and their families. Additional difficulties have been, and still are due to the inadequacy of the equipment available for the employees, as well as the inadequacy of the net, the uncomfortable spaces within their homes, the lack of ergonomic equipment and furniture at their disposal.
The lack of all the needed controlling devices (not everything is available in the web), that, at times, makes it necessary to complete the cataloguing on-site. Both bureaucracy and the restrictive interpretation of the regulation are ever enemies: perhaps, it might be unavoidable, but the adopted registration of the withdrawn and returned books takes precious time away.
Certainly, the emergency has compelled the administration to implement this new working method, something they had been talking about for years, without succeeding in going beyond the stages of the mere projects, more or less created *ad personam.* Now we are able to assess accurately its impact on the library organization and on the real lives of the employees, which has been a huge step forward in only a few months.
Among the so many considerations that can be made, I would like to point out how the Covid emergency has imposed verification processes of the very stringent workflows that makes it possible to assess usefully the daily-executed work.

7. Future perspectives and conclusions

There are also criticism and, above all, objectives that we should set, if we had the necessary human resources to reach them.
Among these, I underline the need to optimize the fruition of the BNI product, too hidden within the site of BNCF, and, therefore, complicated to consult. Yet, what is worth it for BNI is also worth it for other national libraries, that is to say the tendency to incorporate and merge with the catalogues of the national libraries that produce them. Although, this is really a primary target and something on which other national libraries work really very well.
Again, BNI has been making its own data available in PDF format since 2012 and, in XML and UNIMARC, since 2015. The availability of BNI even in RDF format is so far a goal to be reached.
We should work in order to give again more space, to enhance the contents and even to change the separate series, besides those of monographs: the periodicals, the printed music and the doctoral thesis.
As already said, a goal to reach as soon as possible is to increase the timeliness with which even the information of semantic type is made available, together with the descriptive record of the books.
In conclusion, I think that, from what above described, it seems clear that through the adoption of the multiple measures of organization and procedural character, a very small and more and more exiguous, yet skilled, expert and motivated group of people was able to contrast, in my opinion amazingly, a completely adverse situation.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12722

They made it possible to increase the grade of thoroughness, timeliness, authority of the Italian national bibliography, at the level that has always characterized it and according to absolute adequate standards, as provided by the National Library Service.
Let's not ignore that the situation is by now extremely critical and that what has been possible so far to do will however not be so for much longer.
The critical mass needed to transform the ideas and the encouragements into concrete projects is going to fail, also because the burden of the daily work has become too overwhelming and the situation is too difficult for everybody. Not only for us at BNI but also for the other sectors of BNCF, as well as for the other organizations with which we should collaborate, starting from the other libraries of the net and of the central Institute of the Ministry itself.
The risk is that a huge and unique heritage made of knowledge, experience, skills, inheritance of generations of librarians, will be lost, without the possibility that it be conveyed to those who will follow. It would be an unrecoverable damage for the world of libraries, for culture, for our Country.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12775

JLIS.it

Artificial intelligence, machine learning and bibliographic control. DDC Short Numbers – Towards machine-based classifying

Elisabeth Mödden[(a)]

a) Deutsche Nationalbibliothek, http://orcid.org/0000-0001-6809-3926

Contact: Elisabeth Mödden, e.moedden@dnb.de
Received: 25 August 2021; **Accepted:** 20 September 2021; **First Published:** 15 January 2022

ABSTRACT

Digital publications now account for the majority of new accessions at the German National Library each year. Due to this growing number, it has become quite challenging to collect and catalogue these items properly. At the same time, these changes allow for new ways, in which it can use the collections. For a number of years, the DNB has been addressing the question of how subject cataloguing processes can be automated so that bibliographic records can be enriched with metadata as comprehensively and uniformly as possible. In the course of introducing automated subject cataloguing procedures, work is also being done on the automated assignment of Dewey Decimal Classification numbers. For this purpose, a set of abridged DDC numbers based on is being developed. The article sheds light on how artificial intelligence is used in this process. Furthermore, the challenges posed by the development of DDC short numbers and machine-based classification for different scientific subjects will be addressed. Also, it discusses how the DNB deals with the issues of data provenance, data delivery and quality management.

KEYWORDS

Dewey Decimal Classification; DDC Short numbers; Artificial intelligence; Machine-based classification.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

Introduction

In the German National Library (Deutsche Nationalbibliothek / DNB), both verbal and classificatory subject cataloguing are used for subject indexing. In the course of introducing automated subject cataloguing procedures, work is also being done on the automated assignment of Dewey Decimal Classification numbers. For this purpose, a set of abridged DDC numbers based on, but not limited to, the DDC Abridged Edition 15 and hereafter referred to as DDC Short Numbers, is being developed.

First experiences in the automatic assignment of abridged numbers were gained in the field of medicine (DDC 610). Since 2005, medical dissertations have been classified using a set of 140 DDC Short Numbers. Since 2015, these Short Numbers have been assigned automatically by utilizing artificial intelligence. Short Number sets for other DDC areas are currently being developed. It is planned to extend the automatic assignment of Short Numbers to all subjects and to constantly review the process and its results.

Initial situation

Digital publications now account for the majority of new accessions at the German National Library each year, and the number is rising (see figure 1). In 2020, the collections grew by approx. 1 million online publications like e-books and electronic journal articles. Due to this growing number, it has become quite challenging to collect and catalogue these items properly. At the same time, these changes allow for new ways, in which we can use our collections; for example, it is possible to search for and retrieve individual articles.

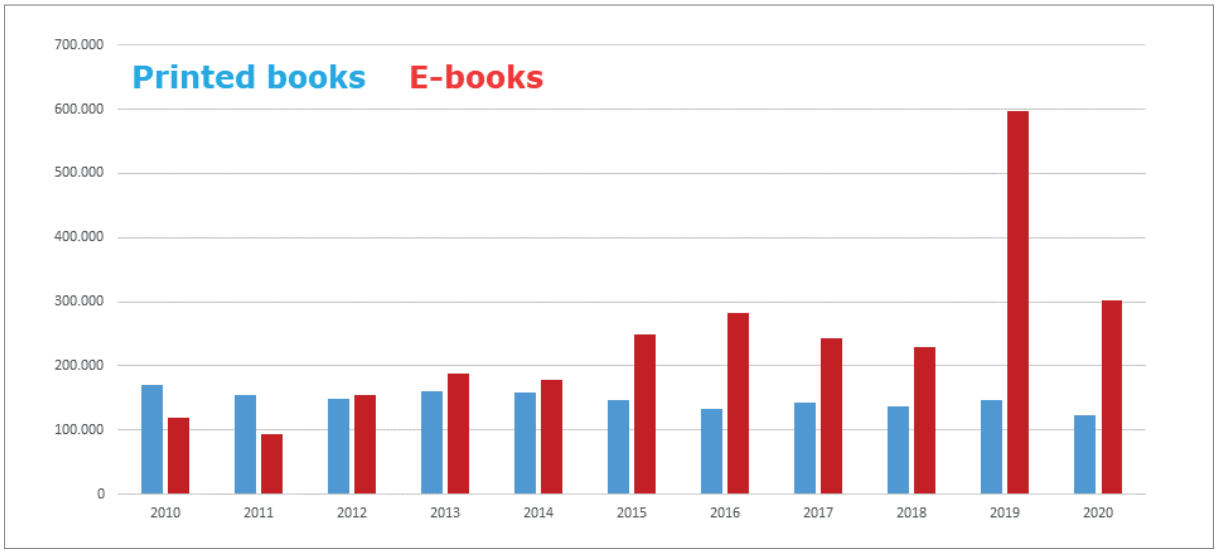

Fig. 1. Increasing amount of publications to be catalogued (e. g. monographs)

Subject cataloguing makes it possible to structure the library's large collections thematically and thereby facilitate the retrieval of publications in these collections. For a number of years, the

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

DNB has been addressing the question of how subject cataloguing processes can be automated so that bibliographic records can be enriched with metadata as comprehensively and uniformly as possible – despite new media formats and ever-increasing number of units. Other advantages of automated processes, e.g. the possibility of cataloguing component part of works such as the above-mentioned journal articles both by classification and by assigning subject headings, should be exploited consistently.

Since 2010, the DNB has increasingly been classifying and indexing digital publications using automated procedures rather than intellectual processes (Gömpel, Junger, and Niggemann 2010). In September 2017, the use of machine-based cataloguing procedures was extended to physical publications (Junger and Schwens 2017) ("Cataloguing Media Works" n.d.). In the DNB's Strategic Compass 2025 (Deutsche Nationalbibliothek 2016a) and Strategic Priorities (Deutsche Nationalbibliothek 2016b), the reorganisation of subject cataloguing is addressed as a significant area of activity that will continue to be important during the years to come. This article sheds light on how artificial intelligence is used in this process. Furthermore, the challenges posed by the development of DDC short numbers and machine-based classification for different scientific subjects will be addressed. Also, it discusses how the DNB deals with the issues of data provenance, data delivery and quality management.

Cataloguing methods

Subject cataloguing at the DNB is based on the Series of the Deutsche Nationalbibliografie (German National Bibliography). Every publication catalogued since the bibliographic year 2004 is assigned to one of roughly one hundred subject categories, which are organised in accordance with the Dewey Decimal Classification (DDC) system ("Dewey Decimal Classification (DDC)" n.d.). Beyond that, the publications from the publishers´ book trade provided in Series A are processed intellectually using built numbers from the DDC and subject headings from the Integrated Authority File, the Gemeinsame Normdatei ("Gemeinsame Normdatei (GND)" n.d.).

The development of software applications for subject cataloguing purposes started with the PETRUS project (Schöning-Walter 2010). Machine-based subject category assignment began in 2012, while the automated assignment of subject headings got under way in 2014. Medical publications were first automatically assigned DDC Short Numbers in 2015. At present, work is under way to develop DDC Short Numbers for all subjects.

The DNB employs a support-vector machine for the use in machine-learning processes to facilitate automated classification using DDC Subject Categories and DDC Short Numbers (Mödden and Tomanek 2012). The characteristics of selected text parts and existing metadata are analysed by means of linguistic and statistical methods. During the training phase, the system analyses publications with intellectually assigned Subject Categories and Short Numbers to generate a reference model for all classes of DDC short numbers. When creating this model, it is essential that each class contain sufficient numbers of appropriate learning examples. During the cataloguing process, the system then calculates a statistical measure to determine how closely the content of a new publication matches the patterns learned. As the result of topical classification, the best-matching Subject Categories and Short Numbers are assigned to the publication (see figure 2).

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

The cataloguing software was created in cooperation with the Freiburg-based company Averbis and is integrated into the DNB's system infrastructure. Machine-based classification has been implemented for texts in German and English.

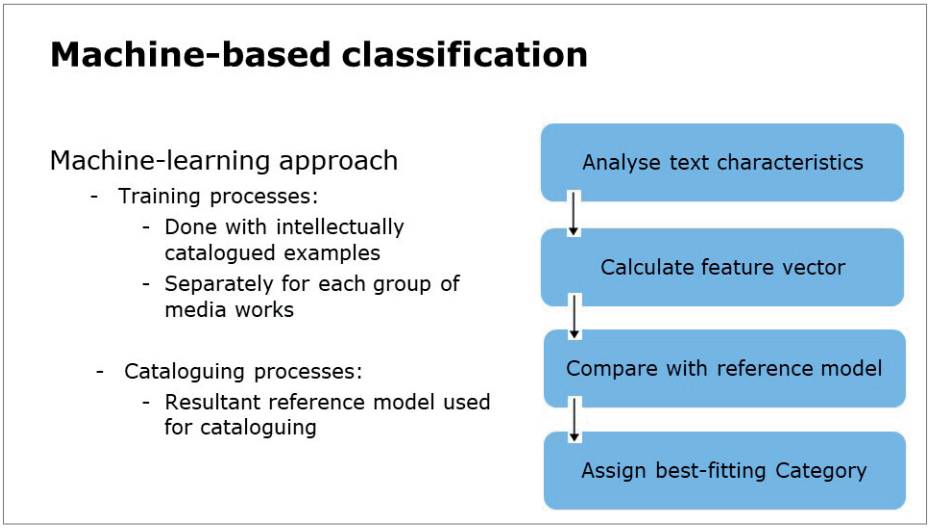

Fig. 2. Processes used in machine-based classification

Workflow

In productive operations, the machine-based cataloguing process (see figure 3) begins automatically at a fixed time every day by sending a list of publications that require first-time processing [1] to a web service. This service retrieves the existing metadata [2] from the cataloguing database (CBS) and the digital full text files or tables of contents [3] from the repository. Before being transmitted to the cataloguing software [4], the storage formats are converted into simple text files and the main language of the publication is determined. Once they have been processed in this way, the results of the analytical process [5] are added to the publication's bibliographic record [6]. Anomalies found during processing are recorded in system files.

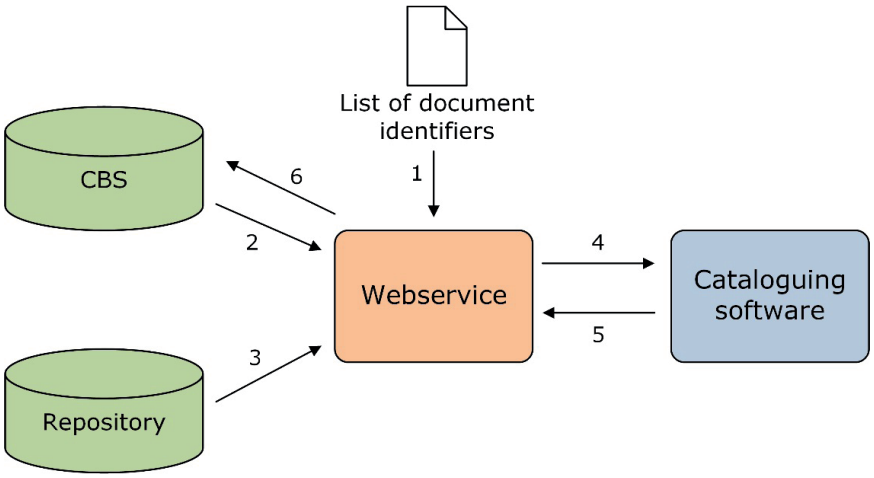

Fig. 3. Technical process used at the DNB for automated cataloguing (productive operation)

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

The cataloguing software has various configuration options enabling different types of publications to be processed in different ways. These configurations consist of parameter settings, which have been optimised during test runs. They facilitate the identification of the classification model, for example when assigning Subject Categories. Depending on the features of the publication a certain configuration is set: for instance, digital monographs are processed differently from journal articles, German-language texts are processed differently from English ones, full text files are processed differently from digitised tables of contents.
The software, training corpora and added GND vocabulary undergo regular maintenance and development to ensure that the system as a whole improves constantly. At certain times, digital publications catalogued intellectually are added into the machine-learning processes as new examples. This raises the question whether automated cataloguing processes should be repeated when significant progress is made. In the future, we want to introduce cyclical repetition in order to improve the quality of our machine-generated metadata. We also want to include publication formats that previously were omitted.

Milestones

At the beginning of 2017, the automated cataloguing processes were extended to include journal articles in digital format. The import procedure for e-journals started at the beginning of 2016. Around 675,000 journal articles were integrated into the DNB's collections in 2016 alone. Beginning with journals published by Springer Publishing, the DNB is now enriching individual articles with subject cataloguing metadata. In view of the great number, periodical online publications can only be catalogued economically at this extent by using automated methods.
Another strategic milestone was reached in September 2017 when the automated cataloguing processes were extended to printed monographs in the Deutsche Nationalbibliografie's H Series ("Deutsche Nationalbibliografie" 2019)[1]. Since September 2017 the DNB no longer applies full DDC numbers for this Series. These are to be gradually replaced by DDC Short Numbers. Publications from the publishers' book trade (Series A) will continue to be catalogued intellectually.
In due time, all existing digital resources, for example parallel online editions, tables of contents, abstracts, blurbs and cover texts, will be used for machine-based subject cataloguing of physical media works. At present, publications are catalogued on the basis of digitised tables of contents and the bibliographic metadata that has been supplied. However, since there is less text and a lack of substantial information in some tables of contents the conditions for text analysis are frequently more unfavourable than in the case of online publications. Therefore, the automatically assigned Subject Categories for Series H are all reviewed intellectually.

[1] In Series H are university publications: Dissertations and postdoctoral theses from German universities and German-language dissertations and postdoctoral theses from abroad.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

How DDC Short Numbers are selected

Until they are ready for productive use in automatic classificatory indexing, DDC Short Numbers have to pass a multi-stage workflow. Dewey numbers are selected per subject, using the DDC Subject Categories as a guide. This process is accompanied, if necessary, by a comparison with Dewey numbers of DDC's Abridged Edition 15. The next step is to analyse the frequency of occurrence of DDC numbers on the basis of the literature published over the last ten years. Building on this, suitable numbers are selected while numbers with low literature warrant are discarded. This data set is then used for initial technical tests to see how well the selected numbers are working for automatic assignment in the respective subject. Mismatches are analysed and Short Numbers are adjusted in an iterative process. Finally, if the results are convincing, the Short Numbers are put into productive operation and the selection process begins for the next Subject Category. The experts at the Department for Subject Indexing are closely monitoring this iterative process.

Provenance data

The decision to apply automatic processes goes along with the decision to assign the machine-generated metadata to the bibliographic record, to display it in the DNB portal, to use it for retrieval purposes, and to deliver it via the data services. In addition to this, metadata for journal articles is now available in the DNB catalogue and can be obtained through the data services. The DNB's database structure was modified to supply information on the provenance and reliability of the machine-generated metadata. In our database, the machine-generated metadata is recorded together with the date, the configuration name and the confidence value, which is an estimate of the data quality. The machine-generated DDC Short Numbers and subject headings are indicated as such when displayed in the DNB portal (see figure 4).

Link	http://d-nb.info/1211853292
Titel	Keine Auswirkungen des Antibiotikums Norfloxacin auf die Hämodynamik und Rho-Kinase-Expression bei portaler Hypertension im Tiermodell
Person(s)	Bücher-Ollig, Doris Claudia Kristin (Verfasser)
Theses	Dissertation, Rheinische Friedrich-Wilhelms-Universität, 2020
Subject headings	**Norfloxacin* ; Tiermodell* ; Pfortaderhypertonie* ; Leberzirrhose* ; Hypertonie*** (*machine generated)
DDC Number	**616.1*** (*machine generated DDC Short Number)
Subject Category	**610 Medizin, Gesundheit*** (*machine generated)

Fig. 4. Title of an automatically catalogued Series O publication displayed in the DNB catalogue with subject headings, DDC Short Number and DDC Subject Category

The data exchange format MARC 21 has also been modified so that standardised information on the provenance of the metadata can be distributed as well.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

Quality and monitoring

Along with daily controls of the process operation, technical checks are carried out by means of sampling. Here, a selection of the publications submitted for automated cataloguing is also classified and assigned subject headings on an intellectual basis. All metadata generated during the cataloguing processes is recorded in the bibliographic database. For display and use in the portal and data services, preference is given to metadata assigned intellectually if available.

For quality management purposes, the quality of the machine-generated classifications is evaluated statistically by comparing automatically and intellectually assigned metadata. Existing metadata for parallel editions is also used for this purpose if applicable. Over the last five years, the DNB has reviewed approximately 18% of the automatically classified online publications in Series O ("Deutsche Nationalbibliografie" 2019). The machine-generated Subject Categories agreed with the intellectually assigned Subject Categories in 76% of cases. This average was actually clearly exceeded in some subject areas, e.g. in law (92% consistency) and medicine (87% consistency). However, machine-based classification does not yet function satisfactorily particularly in the case of subjects on which there is little literary warrant, because there is not enough of the training material required for the learning processes. One such subject for example is the history of South America (DDC Subject Category 980).

There are several issues with machine-based classification of DDC Subject Categories. The main problem is that the machine-assigned DDC Subject Category determines the Short Number. If the Subject Category is wrong, the Short Number will be wrong. Another challenge is posed by the fact that the DDC is continuously updated; even if changes on the broader hierarchy levels do not occur frequently, both changes in the meaning of the class (e.g. change of caption, added or removed major topics) and notational changes such as new or deleted numbers can have an impact on the correct assignment of a Short Number and thus must be taken into account in the process.

Combining machine-based and intellectual cataloguing

Automatic cataloguing procedures are not free from error. Along with imprecise or incorrect assignments, they also generate a bulk of metadata that is not useful for our patrons. The task of quality management is to critically evaluate the error ratio and its effects on the metadata stock in order to adjust the cataloguing processes if necessary. The goal is to achieve a high degree of reliability for the cataloguing data, irrespective of whether it was generated intellectually or automatically. The intellectual and machine-based processes are to be linked more closely in the future. Quality management serves to control and determine which publication forms can be catalogued automatically and which cataloguing services have to be performed intellectually.

Outlook

In 2018, the company Averbis announced a stop to further software developing for machine-based cataloguing. The existing software will be supported only for the next 5 years. Thus, the development of a new machine-based cataloguing system is under way. The target is a new software with a modular structure. This will make it easier to replace individual tools in the future. For this

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

purpose, the project "Erschließungsmaschine" – EMa was started. By this, the Averbis software is scheduled to be replaced with a new modular software system by 2022.

Major requirements for the new system are individual modules for text extraction, language recognition, classification, subject indexing, management of text corpora, of terminologies and of notations, etc. After a detailed market study, the Annif toolkit was selected. The National Library of Finland has developed Annif as a tool for machine indexing. The open-source toolkit "uses a combination of existing natural language processing and machine learning tools including Maui, Omikuji, fastText and Gensim. It is multilingual and can support any subject vocabulary (in SKOS or a simple TSV format). It provides a command-line interface, a simple Web UI and a microservice-style REST API." ("Annif – Tool for Automated Subject Indexing" n.d.). For more details, see the very interesting paper by Osma Suominen (Suominen 2019) and the Documentation on GitHub ("GitHub – NatLibFi/Annif:." n.d.). The DNB is very much looking forward to working with Annif, since it is a very promising new tool and is firmly believing that it will pose new opportunities for machine-based classifying and indexing.

In addition, a new, innovative AI (artificial intelligence) project is being launched at DNB. The DNB wants to develop new methods for processing and analysing content and metadata. The new approach should improve the quality of machine-based content indexing in a significant way. Potential AI developments, which are suitable for cataloguing text-based publications, will be investigated, selected, combined and adapted. Research will be conducted to determine which AI methods can be used for machine processing and analysis of natural language texts in order to obtain the most complete and accurate indexing data. The DNB aims for flexibly reusable tools (open-source tools), so that other libraries or institutions with comparable tasks can use these developments as well.

A good database, based on high-quality intellectual indexing by subject experts, is an indispensable prerequisite for the AI project. Therefore, the Department for Subject Indexing will be intensively involved in the development of new procedures. Furthermore, the rules for subject cataloguing should be adapted in such a way as to benefit the combination of both approaches – intellectual and machine-based subject indexing. In the end, the DNB is convinced that high-quality indexing can be achieved by combining intellectual and machine generated classifying and indexing.

JLIS.it vol. 13, no. 1 (January 2022)
ISSN: 2038-1026 online

DOI: 10.4403/jlis.it-12775

References

"Annif – Tool for Automated Subject Indexing". n.d. Accessed 30 July 2021. http://annif.org/.

"Cataloguing Media Works". n.d. Accessed 29 July 2021. https://www.dnb.de/EN/Professionell/Erschliessen/erschliessen_node.html.

"Deutsche Nationalbibliografie". 2019. https://www.dnb.de/EN/Professionell/Metadatendienste/Metadaten/Nationalbibliografie/nationalbibliografie.html.

Deutsche Nationalbibliothek. 2016a. *2025: Strategic Compass*. Leipzig, Frankfurt, M: Deutsche Nationalbibliothek. https://d-nb.info/1112299556/34.

Deutsche Nationalbibliothek. 2016b. *Strategic Priorities 2017–2020*. Leipzig, Frankfurt, M: Deutsche Nationalbibliothek. https://d-nb.info/1126595101/34.

"Dewey Decimal Classification (DDC)". n.d. December. Accessed 30 July 2021. https://www.dnb.de/EN/Professionell/DDC-Deutsch/ddc-deutsch_node.html.

"DNB_Strategic-Compass-2025_lesesprache_englisch.Pdf". n.d.

"Gemeinsame Normdatei (GND)". n.d. Deutsche Nationalbibliothek. Accessed 30 July 2021. https://www.dnb.de/DE/Professionell/Standardisierung/GND/gnd_node.html.

"GitHub - NatLibFi/Annif: Annif Is a Multi-Algorithm Automated Subject Indexing Tool for Libraries, Archives and Museums. This Repository Is Used for Developing a Production Version of the System, Based on Ideas from the Initial Prototype." n.d. GitHub. Accessed 30 July 2021. https://github.com/NatLibFi/Annif.

Gömpel, Renate, Ulrike Junger, and Elisabeth Niggemann. 2010. "Veränderungen Im Erschließungskonzept Der Deutschen Nationalbibliothek". *Dialog Mit Bibliotheken* 22 (1): 20–22.

Junger, Ulrike, and Ute Schwens. 2017. "Die Inhaltliche Erschließung Des Schriftlichen Kulturellen Erbes Auf Dem Weg in Die Zukunft". *Dialog Mit Bibliotheken* 29 (2): 4–7.

Mödden, Elisabeth, and Katrin Tomanek. 2012. "Maschinelle Sachgruppenvergabe Für Netzpublikationen". *Dialog Mit Bibliotheken* 24 (1): 17–24.

Schöning-Walter, Christa. 2010. "PETRUS – Prozessunterstützende Software Für Die Digitale Deutsche Nationalbibliothek". *Dialog Mit Bibliotheken* 22 (1): 15–19.

Suominen, Osma. 2019. "DIY Automated Subject Indexing Using Multiple Algorithms". *LIBER Quarterly* 29 (1): 1–25. https://doi.org/10.18352/lq.10285.

"The Integrated Authority File (GND)". n.d. December. Accessed 29 July 2021. https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html.