

J LIS.it 11, 2 (May 2020)
ISSN: 2038-1026 online
Open access article licensed under CC-BY
DOI: 10.4403/jlis.it-12623

J LIS.it

When Linked Data is (not) enough. Cataloguing Tools between Obsolescence and Innovation

Alessandra Moi^(a)

a) @CULT, <http://orcid.org/0000-0003-0104-4999>

Contact: Alessandra Moi, alessandra.moi@atcult.it

Received: 24 January 2020; Accepted: 12 February 2020; First Published: 15 May 2020

ABSTRACT

The irruption of the informative space of Web, as new manner to disseminate and share information content, has imposed in the last years an important rethinking of international bibliographic production in its theoretical and methodological basis. In this paper we try to understand the evolutive way of the cataloguing tools through the identification and definition of several change factors (technical and functional): which could be in the next future the new challenges related to the creation and treatment of information?

KEYWORDS

Bibliographic control; Opac; Linked data; Information retrieval.

CITATION

Moi, A. "When Linked Data is (not) enough. Cataloguing Tools between Obsolescence and Innovation." *J LIS.it* 11, 2 (May 2020): 1–19. DOI: [10.4403/jlis.it-12623](https://doi.org/10.4403/jlis.it-12623).

Introduction

When the Library of Congress published its “On the record” (Working Group on the Future of Bibliographic Control 2008) in January 2008, it set a new starting point for the international speculation on cataloguing. Adopting and anticipating, at the same time, the rapid changes that affected the global structure of libraries, the famous report proposed a broad assessment of the future role that bibliographic production would play in the information space of the WWW.

The central element of the document is the combination between the bibliographic production (meaning the process of creating metadata for the entities of the bibliographic world) and bibliographic control, considered an essential step to “facilitate discovery, management, identification, and access” of resources.

With respect to the availability of resources of various nature, bibliographic control confirms itself to be a real “distributed activity”, primarily focused on *managing relationships*, with relationships meant as information networks among resources from different parties: the commitment of each involved actor is to ensure high data quality and to affirm its authority over the produced data. Obviously, the construction of such a context is strictly based on technological changes and on the ability of libraries to incorporate and align with them, also facing strong retraining of the library professionals. The desired technological evolution – which, at least at the theoretical level, should lead to a “release” of bibliographic data outside the restricted world of catalogues – is linked to some, essential, factors: the abandonment of the MARC format; the sharing of vocabularies and controlled terminology in the library field with the world of the Web; the reference to unique models and cataloguing standards for greater international data consistency.

This ability of the catalogues to open up to external information systems should, consequently, lead to a better dialogue with users, who should make the best use of the cataloguing tools even in their most complex functions: with reference to meaningful data, the report highlights that the majority of end-users “have low knowledge of how to use the library catalog”. The solution of the gap between users and libraries must envisage an overall reorganisation of the information retrieval methods, more oriented towards the discovery and facilitation of the interaction processes between users and the catalogue.

This paper aims at offering some thoughts on the current tools for the creation, research and retrieval of bibliographic information and on how much they are actually aligned with the “new” horizon envisaged, more than ten years ago, by the Library of Congress. The second point of investigation focuses instead to the possible technological evolution of these tools and in which unprecedented scenarios they will be placed.

Integration, discovery, sociality

Opac 2.0, SOPAC, ILS, Discovery tool. By performing a selection of some articles or essays available in the vast professional literature, one often comes across the use of multiple terms, indicating both a particular evolutionary form of the catalogue and specific auxiliary tools that enhance some of its functions. In the technical differences characterising these information recovery systems, however,

there is a common intent: that to respond, in a more or less adequate way, to those expectations, to those “concerns” raised by the appearance of the digital information world.

On the one hand, there are the needs of the end-user, who wants simple, intuitive and rapid searches, whose results can provide the desired information or resources; in addition to that, users have the consolidated habit to move from one content (or context) to another to improve the results of the initial search.¹ On the other hand, though, librarians demand increasingly integrated systems, where the cataloguing component is included in a wider management system that deals with the full workflow.

Despite their different purposes, both the needs expressed by the users and the requests made by librarians seem to converge on specific aspects, which condition and unite the functioning of the aforementioned tools. The mention is specifically about integration, discovery, sociality.

Integration

“I think the largest barrier we face in implementing the ideas of ‘Library 2.0’ is that libraries have never really solved the fundamental problem from the days of ‘Library 1.0’ — namely, integration” (Tennant 2007).

“A next generation catalog should be inclusive and integrated” (Marchitelli 2009).

The primary aspect to highlight is integration, with a three-fold meaning: management integration, research integration, integration with the external context.

The integration of the management systems goes back to the technical and organizational practice where each sector of the library work (cataloguing, administration, circulation, electronic resources) had its own tool, separate, not communicating with the others, and the flows had rigidly separated working spaces.

If from a purely technical point of view, that issue has been resolved with the adoption of new generation integrated management systems such as the ILS (Integrated Library System), the resolution of the organizational structure of the institutions is far more problematic; in fact, they should aim at overcoming the traditional and rigid separation between sectors in favor of a more fluid operating environment.

To this extent, the clear separation between paper and electronic resources belonging to the same “collection” is a classic example; in fact, their treatment is often entrusted to distinct professionalism and workflows. How could these types of resources ensure greater integration? Sharing during the metadata phases, of authority data might be an example: the paper and electronic resources,

¹ The investigations conducted by OCLC and the University of California Libraries regarding the needs for information retrieval of end users and the ability of the cataloguing tools to effectively meet these needs continue to remain valid, although dating back to the early 2000s (OCLC Online Catalogs 2009; The University of California libraries 2005). The survey conducted on the methods of use and research strategies by university users, in application to the new generation OPAC of the Musicology Library of the University of Pavia, is more recent and related to the Italian context. (Bianchini 2017, to compare with: Fast 2005).

fragmented and distinct due to their physical being, would recompose in a unitary virtual collection on the basis of common aggregating data.

The unity of the systems should therefore accompany the overall unity of the library as a world in which the individual components organically contribute to the performance of final functions and objectives.

The management integration goes together with the integration of research, seen as the strengthening of services for end users.

Traditionally linked only to the physical collection, the OPAC² perfectly adheres to the cataloguing treatment of records: the creation, modification or elimination of any cataloguing data finds immediate and correct representation in the OPAC.

The exclusion from the search of all those data not processed by the cataloguing management system is an important limitation to the described mechanism: therefore, we speak of databases, electronic resources, subscriptions and external accesses, institutional repositories. Progressively replacing what is configured as the “old” Opac, the new tools offer a unified search interface that allows to query, recover and access the multiple resources available: these are known as discovery tools, not exclusively oriented to the recovery of known resources (the FRBR “to find” function, now IFLA LRM) but instead, to the newborn function of discovery, of exploration of what is unknown or not yet known to want.

With the aim of being more familiar and easy to use to the user, the discovery systems have a simple interface similar to well-known search engines like Google, of which they take over and often enhance some functions: e.g. search suggestions, terms’ self-completion, results’ organization according to the criterion of relevance and rationalisation based on pre-established filters or “facets”,³ although the application of these research ways has raised not a few doubts for their “commercial” provenance (La Barre 2007, 85).

The third and complementary element is integration with the “outside” and, more specifically, with the Web.

Despite being a widely debated topic, one wonders what the expression “integration with the Web” means and how it can be achieved. If compared to a more traditional context, the integration of the catalogue with the vast resources of the Web can be ensured through the structuring of a network of connections and links towards web pages with informative content: examples here are the relationships existing among OPACs 2.0 or discovery tools and publishers’ websites, external databases of articles and ebooks, social networks and international encyclopaedic projects, Wikipedia above all.

Although advantageous in terms of enriching and expanding the contents of the catalogue, the most recent integration concept does not imply that much the inclusion of external resources and sources

² With the term Opac (Online Public Access Catalogue), used to indicate the way to search online for bibliographic records, more and diverse technological tools are included, historically determined from the evolutionary steps that brought the OPAC to evolve from first generation in the early 70’s to the extended OPAC in the 90’s. For an historical review of the OPACs, see Marchitelli and Frigimelica 2012. More thoughts on the role of the OPAC in the information space of web 2.0 are in Coyle 2007a; Coyle 2007b.

³ Similar functionalities were already available in the most modern OPAC systems; as Giovanni Bergamin noted, these functionalities include: 1) dynamic grouping of the results; 2) autocomplete and suggestion of the search terms; 3) order of the results based on their relevance (Bergamin 2008); on the same topic, see also Biagetti 2010.

in the catalog, but rather the integration of the catalogue with the external world of the Web: the linked data technology currently appears to be the best way to achieve this difficult goal.

In addition to requiring significant technical innovation, libraries will be called upon to face a radical paradigm shift in the concrete application of linked data: abandon the record in favour of an atomized data structure; predispose to the use of common systems, models and languages that favour interoperability; accept the free re-use of their data, also for purposes and contexts other than the ones they were originally created for; develop large relational networks with authoritative global datasets.⁴

Discovery

“Discovery would be a rich, satisfying experience that would leverage the potentially powerful combination of library-generated metadata, user tagging and other user interactivity, and full-text Web discovery” (Schneider 2006).

According to Paul Weston, the major limitation of the traditional catalogue is related to the enrichment, or to the inability to offer users an enriched information system. The content enrichment gives value to the catalogue by structuring the connections with external sources that can enhance the information content (Weston 2006, 60).

Nonetheless, such a cataloguing tool should propose new and renewed functions, aimed at simplifying the search experience, ensuring full use of data and information to users and consistently highlighting all available resources.

Currently, the tools that are more responding to this approach are the aforementioned discovery tools. Although often defined as “new generation Opac”, the discovery tools present themselves as radically different from the OPACs rather than their natural evolution.⁵

The first and significant distinction characterizing the discovery tool is its technological infrastructure; the internal management of cataloguing data and the display and use of such data by end users is totally separate. As already mentioned, while the OPAC is presented as an integrated tool in the management tool, the discovery is a technical superstructure in which the data are transferred and adapted to the new proposed functions.⁶

The abandonment of the OPAC identification-localization-obtaining process in favor of a more complex path oriented mainly towards the discovery is substantial. Anticipating the times, already in 1987 Charles R. Hildreth argued that the emergence of a new generation of online catalogues should

⁴ “Technical interoperability; Semantic interoperability; Human-resources interoperability; Organisational interoperability” (Iacono 2014, 97–98).

⁵ Lorcan Dempsey talked about a real “extrapolation” and “reincorporation” of bibliographic data from ILS to tools more discover-oriented: “discovery of the catalogued collection will be increasingly disembedded, or lifted out, from the ILS system, and re-embedded in a variety of other contexts” (Dempsey 2006).

⁶ The complex technology behind the discovery tools and their difference (and distance) from the organisation and search paths in the OPAC is well described in the NISO White Paper that mentions “a set of products within the genre of index-based discovery services, often marketed as «web-scale discovery services» which rely on a large central index populated by metadata, full text, or other representations of the content items in a library’s collection” (Breeding 2015, 2).

go “beyond boolean” to release the processes of information query and retrieval from the rigid and absolute matching of exact terms or keywords (Hildreth 1987, 647–667). The discovery tools provide results presented not as reliable answers but as a range of possibilities that the user can reduce and limit to certain aspects or expand, on the basis of his needs and of multiple exploratory paths.⁷

Sociability

The aspect of sociability referred to the tools for retrieving bibliographic information can be brought back to the broader idea of the “participated library”; the generic concept of users is opposed to an active participation in the choices, spaces and activities of the library as an institution.

Despite being the primary intermediary between end-users and the library collections, the catalogue follows to preset languages and mechanisms established by the librarian. This way, the guarantee of data quality goes along with the gradual estrangement of users from the cataloguing tool and from the knowledge of its operation. The idea of a social catalogue, or the phenomenon of social cataloguing, aims at defining the spaces of participation within the current and future cataloguing tool, so as to ensure greater involvement of users who are now no longer passive users. The approach proposed by the “social catalogue” has numerous variations and potential applications that can be linked to two fundamental aspects: creation / sharing and learning.

As for the aspect of creation, both in the discovery tools and in the so-called Opac 2.0, users can produce additional content such as reviews, summaries, observations. In addition to this simple but potentially effective interaction method, there is the possibility of creating terms that are functional to semantic indexing according to the folksonomy procedure (Mathes 2004; Macgregor and McCulloch 2006) – that is the content categorization through the use of keywords. Although this has been partly perceived as a disqualification of the semantic indexing related processes and activities that are based on controlled vocabularies and established procedures, the autonomous practice of defining the concepts could “bring the existing or potential users further closer, transforming them into collaborators: allowing everyone to keep track of their own readings, to judge them, to organize them according to the subject and above all, according to their own mental scheme and, again, to become creator and user of a network, parallel to the solid and formal one ensured by the OPAC (Marchitelli and Piazzini 2008, 5–6)”. This way, the paradigm of the SOPAC, or Social Opac where social means “social network” becomes true. The management of the spaces meant for user interaction, in fact, is largely inspired to sharing environments such as Facebook or Twitter; there, the availability of socialisation places accompanies the creation of new content, and users are the promoters of a shared virtual space. Taking WorldCat as an example, Weston also talks about the spread of the concept of the Opac as a portal in which a large network of libraries and “any potential reader who accesses the portal’s services from any part of the world” benefit from the possibility of creating and enriching shared contributions (Weston 2011, 8).

An additional “social” aspect of the OPAC is that of the catalogue as a learning space. The unstoppable propagation of the Web and, above all, the information retrieval potential of the search

⁷ “This step transforms certainty into probability, equality into similarity, precision into ranking, retrieval of know information into discovery of yet unknown resources” (Marchitelli 2014, 13).

engines marginalised the search and localization functions of the catalogue, favouring on the other side the nature of the catalogue as a bibliographic repertoire and as a “supplier” of enriched information (Galeffi and Marchitelli 2006).

Unlike the phenomenon of social cataloguing where the enrichment processes are connected to users’ activities within the catalogue, the enhancement of the cognitive aspect of the catalogue lies in different practices. Among them, the presence of additional data relating to authors / contributors / editors / etc., to the works (according to IFLA LRM) and to the editorial history of a publication are to mention. Such data are often the result of the underestimated authority control work, and are easy to link to external sources such as Wikipedia, thus defining a large learning space that starting from the limits of the catalogue, reaches the information contents of the Web: a path of oriented and guided discovery would emerge in this case, based on the paths already traced by the librarian.

Towards the infinity and beyond: libraries in the interconnected world

With his well known provocative statement, back in 2014 Roy Tennant stated “The OPAC is dead” (Tennant 2014). This announced death of the OPAC – even though not yet fulfilled – accompanied the simultaneous birth of tools which have oriented the information retrieval processes towards new functions and methods, as presented in the previous paragraph. However, this was limited to superimposing auxiliary tools on the primary and anachronistic structure of the OPAC, without this entailing a truly renewed approach neither to the services nor to the processes of creation and management of information. If Antonella Iacono speaks of the “loss of identity” of the cataloging tool, based totally on rigid algorithms and not on a real information organization (Iacono 2013, 88), Karen Calhoun highlights the monolithicity of the systems (Calhoun 2006, 41), above all with reference to the management part (ILS).

Despite the proclamations in terms of interoperability and discovery, how do current bibliographic tools and IT systems demonstrate their weakness compared to an ever-changing universe of information?

The first aspect to consider is that of the “overlap”, which happened at the expense of the replacement. An evaluation of the most modern ILS currently on the market and supported by valid IT companies at an international level shows that the major innovation these systems introduced almost exclusively concerned the rethinking of the workflows defining an environment common to all operational practices. However, reflecting on the specific aspect of resource metadata and data management, the proposed structure is the traditional one of the creation and manipulation of bibliographic and authority records. The possibility to import data from external systems appears to be limited exclusively to bibliographic records, with a widespread devaluation of authority and with a consequent failure to apply the information organization proposed by IFLA LRM.⁸ Although behind user-friendly graphics and functions, the cataloguing practices continue in the usual binaries of the card record that is a legacy of the analogical past, with little possibility of manipulation of single data, reuse, insertion of values and controlled vocabularies.

⁸ Although referred to a Marc oriented context, the management of data as entity and its effect in the catalogue organisation and cataloguing practices were already showed in one of the first FRBRisation of WorldCat. See: Bennet 2003.

A second aspect is that of the lack of integration, an apparent contradiction: not integration of resources and services, but integration among different systems. Products from competing companies, ILS and discovery tools do not allow the inclusion of external modules or components within their information systems; the migration process from one system to another often requires high costs in both economic and human terms, and always requires data adjustments.

All what is described above happens in a transition phase, in which the new tools have attempted to propose renewed approaches to the bibliographic information treatment, without however reaching a real replacement of anachronistic paradigms and approaches: in other words, it is a coexistence between obsolescence and innovation, destined to end in favour of a renewed vision.

Future scenarios

Although designed to specifically provide a practical version on RDA application to the processing of bibliographic and authority data, the three “database implementation scenarios” designed by Tom Delsey provide alternatives on possible data management developments (Delsey 2009).

Of the three scenarios, the first actually outlines an implementation of object-oriented databases, based on the de-structured treatment of individual data, grouped together to build an entity. The aggregation of bibliographic data is functional to the conventional creation of records which, however, will be specific to each bibliographic entity formulated by the IFLA LRM model: Work, Expression, Manifestation and Item. These primary entities are flanked by the processing of authority data, aiming to realize controlled access points through specific records. Overall, what Tom Delsey offers is a largely relational context, in which the primary entities are connected to the Agents and the semantic data, that constitute their access points, albeit with some elements still highly traditional. Despite trespassing the “single” bibliographic record from which it’s possible to derive the access points of each entity (Opera, Expression, Agents, etc.), a vision still remains focused on the treatment of the records, rather than the entities seen as a grouping of data. The usual, and so far outdated, approach that places bibliographic entities in a prominent role, relegating the entities linked to authority control to the simple role of access points, is outdated.

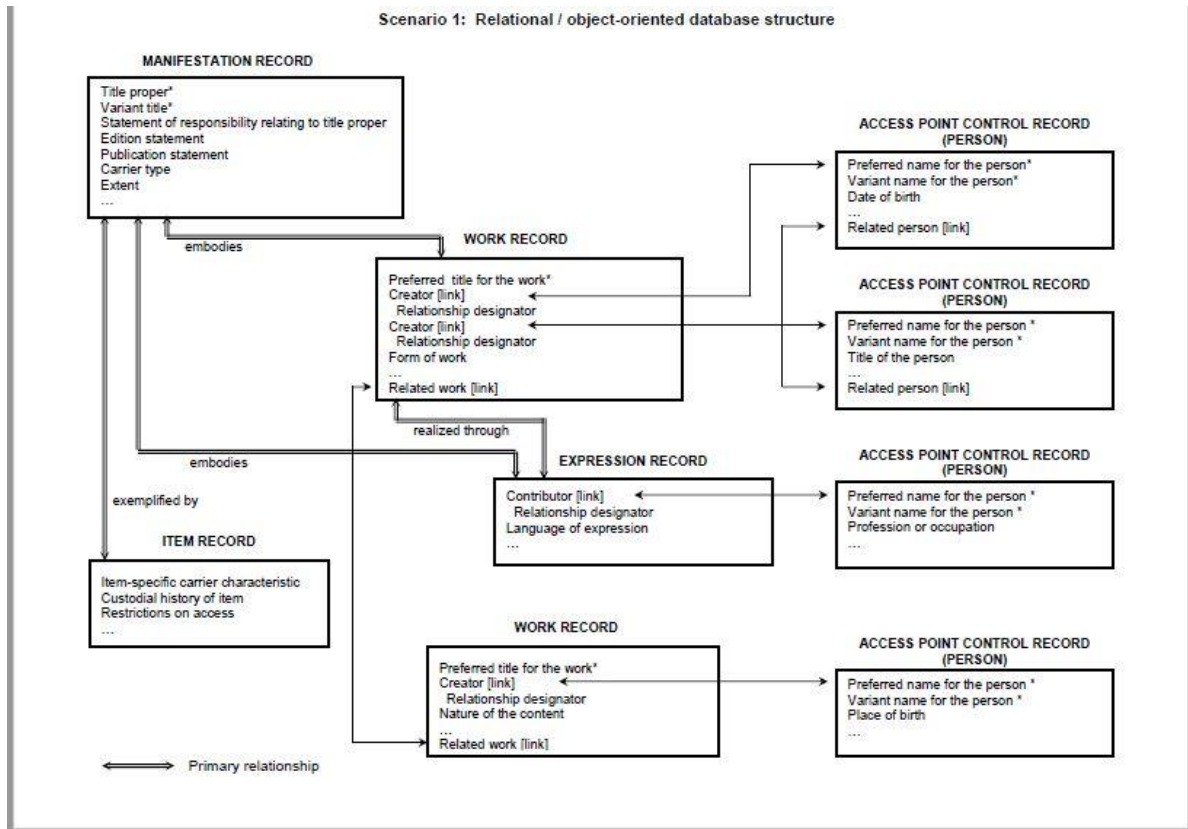


Figure 1. Relational/object-oriented database structure

An interesting vision focused on the disintegration of the record is that provided by Andrea Marchitelli who, taking up a term coined by Roy Tennant, speaks of catalinking (Marchitelli 2014, 9). The unitary vision of the record is replaced by the complete fragmentation of the data (“data mosaic”), in which the entities are reconstructed through the relationship of individual attributes, which assume meaning exclusively on the basis of the value given to the relationship. In this scenario, the metadata activity is reduced to its minimum elements, i.e. the formulation of controlled data, often obtained from external vocabularies and datasets, and the structuring of a complex and potentially unlimited relationship network.

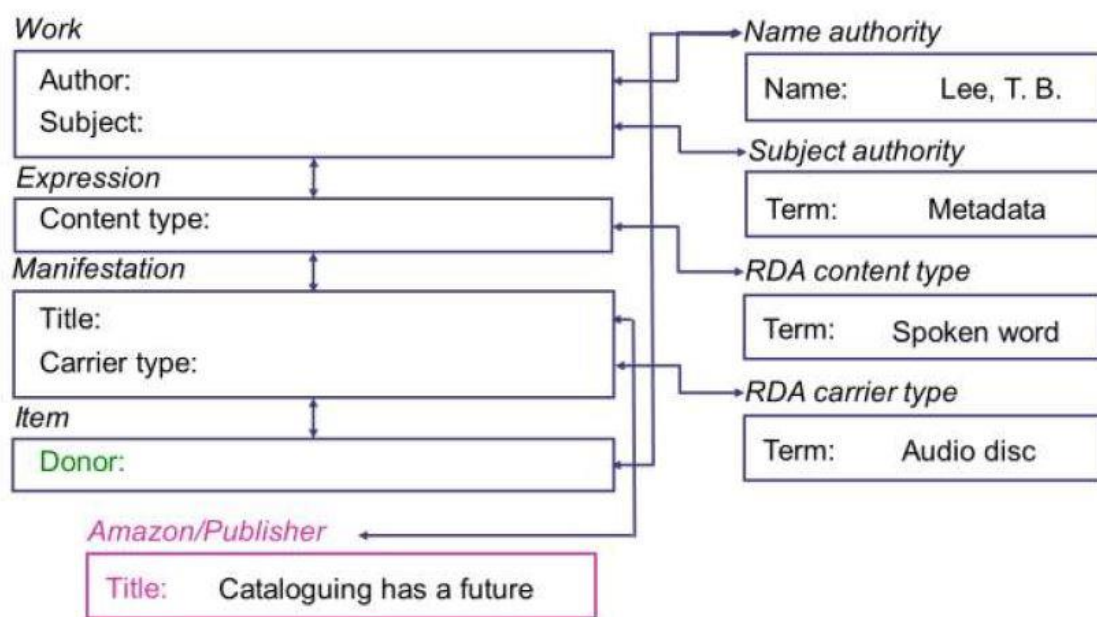


Figure 2. Catalinking

Despite the differences, the two models offer a general approach in the data creation, management and organization processes which is strongly oriented towards the semantic web and, in this case, the possible definition of an infrastructure that allows the direct formulation of bibliographic data in linked open data.⁹

In other words, linked data are perceived as that great possibility to proceed towards a real integration within the information context of the Web, to intercept ever larger portions of users and to proceed towards a radical rearrangement of workflows, although as highlighted by Karen Coyle, the biggest challenge will consist not so much in converting bibliographic data into LOD, as in “creating a new system for access and use of bibliographic data that is compatible and works within the web (Coyle 2013, 57)”.

⁹ The representation of FR family entities through RDF syntax and their possible application to the Semantic web have been the subject of following and important debates, including: Dunsire *et al.* 2011; Dunsire 2012. The assertion of Tim Berners-Lee is fundamental in this regard: “If you are used to the “ER” modelling system for data, then the RDF model is basically an opening of the ER model to work on the Web” (Berners-Lee 1998).

What cannot be given up. Sharing, grouping, connecting

In thinking about the long history that led from the first generation OPACs to the most recent discovery tools and, above all, the attempt to delineate a future evolutionary framework, it was deemed appropriate to highlight “what we will not be able to (or will have to!) give up” to build management and information recovery systems that truly meet the long-standing needs of interconnection, integration and flexibility. As already seen, technological change will be indispensable with the evolution towards LOD and the abandonment of MARC formats,¹⁰ but the “ontological” change in data processing is equally important, with the imposition of new practices in the activities of metadata: libraries will design flows of control and data management no longer in the limited national or local context, but in a truly global perspective.

We chose to propose three aspects now considered essential: sharing, grouping, connection, in the idea of an intimate connection between an innovative theoretical approach and technological evolution.

Sharing

The issue of sharing has long been debated within the international library community: starting from the aforementioned *On the record*, the need to proceed towards ways of exchanging data emerged with the dual aim of achieving greater quality control and significant economic and human effort savings in the cataloguing activity. This aspect is even more valid for authority control, in which the complexity would require participatory tools for the control of titles, names, semantic indexing terms, etc.

A major obstacle to data sharing on a global level is the difficulty of information exchange deriving from MARC formats and the little presence of international authority files, to draw from and also to actively participate to, in a reciprocal exchange relationship.¹¹

However, with the progressive spread of linked open data technology and of the languages constituting its foundation (RDF syntax, ontologies and shared vocabularies), datasets of global size and participation, concerning multiple entities from the bibliographic world and not have been created.

Projects such as ISNI¹² and VIAF¹³ for Agents entities, Geonames¹⁴ for Places, Wikidata¹⁵ and DBpedia¹⁶ for any entity of the human knowledge, although with significant differences regarding their governance and their maintenance, are based on common assumptions:

- **Consistently structured datasets.** The creation of a reliable dataset involves several factors: if on the one hand the central role played by the qualitative aspect of the data and the relevance of their

¹⁰ In addition to his famous speeches about the abandonment of MARC (Tennant 2002a, Tennant 2002b), Roy Tennant already in 2004 underlined that “we do not need a bibliographic record format. We need a bibliographic metadata infrastructure that has a number of components, each of which may have multiple variations” (Tennant 2004).

¹¹ For a quick examination of the advantages connected to the shared authority work see Fons 2014.

¹² <http://www.isni.org/>.

¹³ <https://viaf.org/>.

¹⁴ <https://www.geonames.org/>.

¹⁵ https://www.wikidata.org/wiki/Wikidata:Main_Page.

¹⁶ <https://wiki.dbpedia.org/>.

provenance¹⁷ is undeniable, on the other hand, the validity of the dataset is entrusted by the “good practices” of the LOD. These datasets are therefore based to a large extent on the reuse of already published ontologies or ontological languages approved by the W3C, and provide multiple serialization formats for viewing and any download of data by external users.

- **Persistent identification of entities.** Persistence¹⁸ is one of the great challenges of the semantic web, where the ability of machines to understand and create knowledge is based on a concatenation of URIs. However, the structuring of stable URIs is not always sufficient to guarantee the persistence, especially in large datasets containing millions of data and information and subject to errors because of their dimensions. Using the very large VIAF dataset as an example, the Agents and Works entities are created from data coming from international bibliographic agencies; errors are frequent given the amount of information to be treated and, in case of periodic data “cleaning” activities, entities considered no longer valid can be deactivated. A technical solution partially compensating the issue of cleaning duplicate entries is that of the Redirect, or the redirection from one entity to another. The persistence of URIs is therefore entrusted both to stability policies and to the quality of the data itself, that determines the validity of the entities to a large extent.
- **Free import and / or query of data.** Aligning with the directives of the five stars,¹⁹ the cited datasets (and many others) expect entirely open licenses and support the use of the data included in the dataset through specific tools. The downloading of the entire dataset combines with APIs modules for the on-fly acquisition of data and appropriate SPARQL endpoints for the formulation of query queries internal to the dataset.
- **Shared data.** Up to now, the discussion was about opening, querying and acquiring data, giving more space to external sharing, or to the ability of these datasets to be predisposed to disseminating their data. Sharing goes in the opposite direction, too: albeit to a different extent and according to different mechanisms, data from VIAF, ISNI, Wikidata and others still derive from “community” projects and activities, inside a collaborative flow of processing, structuring and availability of data. This concept is consistent with the new configuration of the production cycle of the data in the semantic web, where “individuals and organizations play at the same time the roles of information producers, gatekeepers, and consumers of information in an ever-reconfiguring ecosystem (Barbera 2013, 95)”.

In a future evolution of bibliographic management systems, the development of systems for acquiring and integrating data from these authoritative datasets by means of direct interrogation mode (single call to datasets) or downloading and inserting entire authority files into the management system will be needed; as mentioned by Simona Turbanti about the authority work “we could continue to have

¹⁷ The concept of provenance bounds in the practice to the “identification of the responsible entity” (Salarelli 2014, 287), providing the metadata triple with information on the origin of the data: from a triple to a quadruple then, according to the data organisation subject-predicate-object-context (Carroll, Bizer, Hayes, and Stickler 2004; Harth, Polleres and Decker 2007).

¹⁸ Persistence indicates the specific property of the URIs to constantly (or indefinitely) refer to the resource they are associated with. Persistence is listed among the Best practices for publishing LOD on the Web: W3C Working Group 2014.

¹⁹ <https://5stardata.info/en/>.

our own in-house authority files in almost all the libraries if we wish, but probably it's time for a faster, smarter and richer authority control" (Turbanti 2014, 56).

Gathering

Managing the "flat" information from the Web of documents as an entity, has opened up a series of questions which, as already seen, are pertinent to multiple technical and non-technical aspects, such as the correct and unambiguous identification of the entities, the reliability of the data associated with them, the possibility of structuring increasingly shared practices and languages.

One of the most debated aspects is undoubtedly the "proliferation" of entities, or the presence of multiple objects that actually refer to the same resource.

It would be difficult, or counterproductive, to aim for a unique representation of the entities: each dataset, if it is authoritative, contributes to increase the knowledge about a given entity, publishing and disseminating data otherwise not owned by other institutes or producers. The entity Alessandro Manzoni, "object" in the semantic web, is represented in projects such as Wikidata, data.bnf, VIAF, ISNI: each project obviously identifies the same entity (the person named Alessandro Manzoni) with different data, exploiting the information potential. If that constitutes a precious opportunity in terms of increasing knowledge, on the other hand it poses the problem of gathering entities that are only apparently different from each other, as they are not identified by the same URI. The term "grouping" used here must be understood as a way of aggregating a specific entity from multiple representations into a single container: technically this container, the result of the agglomeration of identical objects, is called a cluster. Unlike the practice of interlinking, which will be seen below, the clustering processes are oriented to the creation of a unitary group that includes all the objects representing the same entity within itself: the technical definition proposed by Willer and Dunsire, "a cluster is a set of statements about the same thing, with every triple having the same subject URI" (Willer and Dunsire 2013).

Clustering is entrusted to automatic processes through complex algorithms that reconcile these objects through the recognition of common elements. Since the mechanisms are based on the machines processing capacity, clustering is subject to errors and inaccuracies: different entities could coexist within the same cluster or, on the contrary, multiple clusters could be created for the same entity.²⁰ This is obviously related to the essential issue on data quality, and consequently to the enhancement of provenance: if the data are qualitatively valid and correct, the algorithms will have more chances to read and reconcile the entities correctly.

Connection

We have repeatedly mentioned the idea of connection in the broadest sense of the term, referring to the possibility of connecting the single "library system" with the outside universe: connection between library institutions; connection with other interlocutors, such as memory institutes (museums,

²⁰ The crush algorithms used by VIAF for Names and Titles aggregations are exemplifying, as shown by: Manzotti 2010, 363–368.

archives) or commercial interlocutors (publishers, data producers);²¹ connection with the WWW; connection with the end user.

The connection is above all between data in the technological system of LODs: the same conceptual foundation of linked open data is based on the possibility of structuring links between data, through the formulation of RDF triples that link the resources contained in a dataset with similar resources contained in further datasets. The procedure mentioned here takes the name of interlinking, that have the advantage in the possibility of acquiring external data in order to enhance own datasets.

As this is a technical mechanism, interlinking presupposes the correct structuring of the RDF triples which, for the purpose of building such an interconnection network between datasets, can take the form of:

- Identity links. Taking up what has already been said for the clustering functions, the objects that identify the same entity in the semantic web can be innumerable. However, not all authoritative projects related to the dissemination of bibliographic LODs are oriented towards the structuring of those large “containers” which are precisely clusters, whose construction requires refined automatic processes for the reconciliation of entities. A solution is offered precisely by interlinking and, specifically, by the structuring of triples which clarify the similarity between entities that belong to different datasets.

Here is a concrete example, taken from the data.bnf project:

```
<owl:sameAs rdf:resource = “http://wikidata.org/entity/Q1064” />  
<owl:sameAs rdf:resource = “http://viaf.org/viaf/14356” />  
<owl:sameAs rdf:resource = “http://www.idref.fr/027006956/id” />
```

The triples have a similar structure, typical of identity links:

- Subject: the Alessandro Manzoni entity of the data set data.bnf, in this case implied;
 - Preached: the identity relationship is produced by the sameAs²² property, deriving from the OWL ontology, or from similar properties, although this is the most common.
 - Object: the URI alias, or the other URIs identifying the same entity Alessandro Manzoni in different datasets.
-
- Relationship link. As indicated by the name, the relationship links are used to connect not a single entity, but to link to each other different resources contained in multiple datasets, through which you want to make explicit the membership of a domain or their relevance to a specific topic. These links may relate to the logical connections between the animal and its habitat, between a sporting event and the participants, between a branch of knowledge and the works that have determined its maximum diffusion. The interconnection between datasets is well represented by the famous

²¹ Carlo Bianchini refers to modularity as “actual use and immediately application of data, also partial, produced by other agencies and, conversely, the direct use from other agencies of high-quality bibliographic data produced in library environment” (Bianchini 2012, 314).

²² “owl:sameAs is used to state that two URI references refer to the same individual” (W3C 2012). The problem of the identity link among entities is still discussed, especially for what is called the utilisation abuse of sameAs property and for the so-called “identity crisis of linked data” (Halpin, Herman e Hayes 2009).

Linked Open Data Cloud²³ which, with a real explosion of relationships, shows how the possibility of creating links from resource to resource and, consequently, from dataset in dataset, is practically inexhaustible.

As indicated for the sharing aspect, future resource metadata systems will have to base their operations on these functions of connecting and recovering data from external sources, in a constant expansion of the cross-domain information wealth, or in an intersecting manner between different areas of knowledge. This will obviously be based both on the interoperability capabilities of the data and systems, repeatedly stressed, and on the creation of datasets with totally open licenses, in the acceptance that one's data can also be reused in contexts and for purposes totally different from the original ones. As mentioned by Mauro Guerrini and Tiziana Possemato, this "increases the credibility and authoritativeness of the dataset and triggers a virtuous circle of sharing and enriching data" (Guerrini and Possemato 2015).

Conclusions

The path outlined so far had the objective of tracing the broad evolution of the cataloging tools. The first element that emerged is that this is a transition period. The perception of the changes resulting from the information processing of the Web required a great effort to libraries to rethink the standards and principles underlying international cataloguing practices, and above all, to rework their information strategies to identify and respond to the users' renewed information needs. From the first management systems and the first generation OPACs we got to tools presenting complex functionalities (ILS and discovery tools) but still substantially linked to a traditional vision of information, limited to the bibliographic area and with little possibility of integration with the "external" sources; above all, a vision still bound to the MARC record where information is self-contained in. The beginning of the semantic web and the possibilities offered by its concrete application opened a phase of test: multiple projects for the conversion and publication of bibliographic data as linked open data (data.bnf, SHARE-VDE, etc.) shown remarkable potential in terms of information discovery, enhancement and dissemination of this kind of information in the large space of the Web.

An additional element is the problematic transition from traditional management and information retrieval systems towards new ways of processing and making data available, because of two crucial issues: technological backwardness and lack of awareness.

In an article by Gillian Byrne and Lisa Goddard from 2010, these two aspects are considered complementary one to each other (Byrne and Goddard 2010). The development and use of LODs appears limited and there is still a long way to go to reach full permeation between the semantic web and the web of the documents; nonetheless, using such technology in the library context implies further challenges related to the processing and conversion of millions of data and the adaptation to these "foreign" languages and practices. However, the biggest obstacle does not come from the technological limitations that are undeniable and still possible to improve, but it comes instead from the lack of awareness or poor knowledge of the international library community about the application

²³ <https://lod-cloud.net/>.

potential of LODs. Despite numerous studies and projects, in fact, a widespread culture on the implications of linked open data and of the wider context of the semantic web does not exist yet, as well as a reconsideration of methods and workflows now considered permanent. Such a situation encourages the use of traditional standards and processes; it also adds to the substantial inability of the libraries to contract with the “vendors” to obtain internal management systems and infrastructures for the use of data that offer innovative solutions.²⁴

How this transition phase will end is not clear yet: to a large extent, this will be determined by the ability of libraries to rethink their role within the information space of the Web, without this leading to their loss of identity. The intrinsic possibilities of the technological change can be truly effective only if accompanied by a prospective change; moving from the limited local context, the change should lead to the creation of an interconnected, open and shared global bibliographic community network.

References

(Last consultation of the websites: 3rd May 2020).

Barbera, Michele. 2013. “Linked (open) data at web scale. Research, social and engineering challenges in the digital humanities.” *JLIS.it* 4, 1:91–101. DOI: [10.4403/jlis.it-6333](https://doi.org/10.4403/jlis.it-6333).

Bennett, Rick, Brian F. Lavoie, and Edward T. O’Neill. 2003. “The concept of a work in WorldCat: an application of FRBR.” *Library Collections, Acquisitions, and Technical Services* 27, 1:45–59. DOI: [10.1080/14649055.2003.10765895](https://doi.org/10.1080/14649055.2003.10765895).

Bergamin, Giovanni. 2008. “OPAC. Migliorare l’esperienza degli utenti.” *Bibliotime* XI, 1. <http://www.aib.it/aib/sezioni/emr/bibtime/num-xi-1/bergamin.htm>.

Berners-Lee, Tim. 1998. “What the Semantic Web can represent.” Last modified: September 17, 1998. <https://www.w3.org/DesignIssues/RDFnot.html>.

Biagetti, Maria Teresa. 2010. “Nuove funzionalità degli OPAC e relevance ranking.” *Bollettino AIB* 50, 4:339–356. <https://bollettino.aib.it/article/view/5340/5103>.

Bianchini, Carlo. 2012. “Dagli OPAC ai library linked data. Come cambiano le risposte ai bisogni degli utenti.” *AIB studi* 52, 3:303–323. DOI: [10.2426/aibstudi-8597](https://doi.org/10.2426/aibstudi-8597).

Bianchini, Carlo. 2017. “«Funziona come Google, vero?»». Prima indagine sull’interazione utente catalogo nella biblioteca del Dipartimento di musicologia e beni culturali (Cremona) dell’Università di Pavia.” *AIB studi* 57, 1:23–49. DOI: [10.2426/aibstudi-11557](https://doi.org/10.2426/aibstudi-11557).

Breeding, Marshall. 2015. *The future of library resource discovery*. Baltimore: NISO. https://groups.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf.

²⁴ “We need application software to enable retrievals and displays such as topic maps and other cluster display capabilities, utilizing the FRBR relationships. [...] We hope to see new information retrieval systems, corporate integrated systems, that move towards XML-based data packages.” (Tillett 2005, 204).

- Byrne, Gillian, and Lisa Goddard. 2010. "The Strongest Link. Libraries and Linked Data." *D-Lib Magazine* 16, 11/12. DOI: [10.1045/november2010-byrne](https://doi.org/10.1045/november2010-byrne).
- Calhoun, Karen. 2006. "The Changing Nature of the Catalog and its Integration with Other Discovery Tools", prepared for the Library of Congress. Final Report 17 March 2006. <https://www.loc.gov/catdir/calhoun-report-final.pdf>.
- Carroll, Jeremy, Christian Bizer, Patrick Hayes, and Patrick Stickler. 2004. "Named Graphs, Provenance and Trust." In *Proceedings of the 14th international conference on World Wide Web WWW*, 613–622. DOI: [10.1145/1060745.1060835](https://doi.org/10.1145/1060745.1060835).
- Coyle, Karen. 2007a. "The Library Catalog. Some Possible Futures." *The Journal of Academic Librarianship* 33, 3:414–416.
- Coyle, Karen. 2007b. "The Library Catalog in a 2.0 World." *Journal of Academic Librarianship* 33, 2:289–291.
- Coyle, Karen. 2012. *Linked Data Tools. Connecting on the Web*. Chicago: ALA TechSource.
- Coyle, Karen. 2013. "Library linked data. An evolution." *JLIS.it* 4, 1:53–61. DOI: [10.4403/jlis.it-5443](https://doi.org/10.4403/jlis.it-5443).
- Delsey, Tom. 2009. "RDA Implementation Scenarios." <http://www.rda-jsc.org/archivedsite/docs/5editor2rev.pdf>.
- Dempsey, Lorcan. 2006. "Lifting out the catalog discovery experience." May 14. <http://orweblog.oclc.org/Lifting-out-the-catalog-discovery-experience/>.
- Dunsire, Gordon. 2012. "Representing the FR family in the Semantic Web." *Cataloging & Classification Quarterly* 50, 5-7:724–741. DOI: <https://doi.org/10.1080/01639374.2012.679881>.
- Dunsire, Gordon, Diane Hillmann, Jon Phipps, and Karen Coyle. 2011. "A Reconsideration of Mapping in a Semantic World." In *International Conference on Dublin Core and Metadata Applications*, 26–36. <https://dcpapers.dublincore.org/pubs/article/view/3622>.
- Fast, Karl, and D. Grant Campbell. 2005. "‘I still like Google’: University student perceptions of searching OPACs and the Web." In *Proceedings of the American Society for Information Science and Technology*, 41, 138–146. DOI: [10.1002/meet.1450410116](https://doi.org/10.1002/meet.1450410116).
- Fons, Theodore. 2014. "Authorities, Entities & Communities." In *IFLA WLIC 2014*, Lyon, 16-22 August 2014. <http://library.ifla.org/1034/>.
- Galeffi, Agnese, and Andrea Marchitelli. 2006. "Il catalogo come learning place. Nuove competenze del bibliotecario." In *Bibliotecari al tempo di Google. Profili, competenze, formazione*, Milano, 17-18 March 2016. <http://eprints.rclis.org/29965/>.
- Guerrini, Mauro, and Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei. Perché l'informazione sia del web e non solo nel web*. Milano: Bibliografica.
- Halpin, Harry, Ivan Herman, and Patrick J. Hayes. 2009. "When owl:sameAs isn't the same. An analysis of identity links on the Semantic Web." <https://www.w3.org/2009/12/rdf-ws/papers/ws21>.

- Harth, Andreas, Axel Polleres, and Stefan Decker. 2007. "Towards a social provenance model for the Web." In *Workshop on Principles of Provenance (PrOPr)*. <https://aran.library.nuigalway.ie/handle/10379/527>.
- Hildreth, Charles R. 1987. "Beyond Boolean. Designing the Next Generation of Online Catalogs." *Library Trends* 35, 4:647–667. <https://core.ac.uk/download/pdf/4816836.pdf>.
- Iacono, Antonella. 2013. "Verso un nuovo modello di OPAC. Dal recupero dell'informazione alla creazione di conoscenza." *JLIS.it* 4, 2:85–107. DOI: [10.4403/jlis.it-8903](https://doi.org/10.4403/jlis.it-8903).
- Iacono, Antonella. 2014. "Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC." *JLIS.it* 5, 1:77–102. DOI: [10.4403/jlis.it-9095](https://doi.org/10.4403/jlis.it-9095).
- La Barre, Kathryn. 2007. "Faceted Navigation and Browsing Features in New OPACs: Robust Support for Scholarly Information Seeking?." *Knowledge Organization* 34, 2:78–90.
- Macgregor, George, and Emma McCulloch. 2006. "Collaborative tagging as a knowledge organisation and resource discovery tool." *Library Review* 55, 5:291–300. DOI: [10.1108/00242530610667558](https://doi.org/10.1108/00242530610667558).
- Manzotti, Giulia. 2010. "Analisi e riflessioni sul VIAF, Virtual International Authority File." *JLIS.it* 1, 2:357–381. DOI: [10.4403/jlis.it-4520](https://doi.org/10.4403/jlis.it-4520).
- Marchitelli, Andrea. 2009. "Stat rosa pristina nomine? Biblioteche e cataloghi ai tempi del nuovo Web." *AIDAInformazioni* 27, 3/4:31–43. <http://eprints.rclis.org/14137/>.
- Marchitelli, Andrea. 2014. "Il catalogo connesso. Dal silos di dati al network informativo." *Biblioteche oggi* 32, 6:5–15. DOI: [10.3302/0392-8586-201406-005-1](https://doi.org/10.3302/0392-8586-201406-005-1).
- Marchitelli, Andrea, and Giovanna Frigimelica. 2012. *OPAC*. Roma: Associazione italiana biblioteche.
- Marchitelli, Andrea, and Tessa Piazzini. 2008. "OPAC, SOPAC e Social networking. Cataloghi di biblioteca 2.0?." *Biblioteche oggi* 26, 2:82–92. <http://eprints.rclis.org/10964/>.
- Mathes, Adam. 2004. "Folksonomies. Cooperative Classification and Communication Through Shared Metadata." December 2004. <http://adammathes.com/academic/computer-mediated-communication/folksonomies.html>.
- OCLC Online Catalogs. 2009. *What Users and Librarians Want. An OCLC Report*. OCLC: Dublin. <https://www.oclc.org/content/dam/oclc/reports/onlinecatalogs/fullreport.pdf>.
- Salarelli, Alberto. 2014. "Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento." *AIB studi* 54, 2/3:279–293. DOI: [10.2426/aibstudi-10128](https://doi.org/10.2426/aibstudi-10128).
- Schneider, Karen G. 2006. "How OPACs Suck, Part 3: The Big Picture." *American Library Association*, May 20. <http://www.ala.org/tools/article/ala-techsource/how-opacs-suck-part-3-big-picture>.
- Tennant, Roy. 2002a. "MARC must die." *Library Journal* 127, 17:26–28.
- Tennant, Roy. 2002b. "MARC exit strategies." *Library Journal* 127, 19:27–28.
- Tennant, Roy. 2004. "A bibliographic metadata infrastructure for the twenty-first century", [Preprint]. <http://eprints.rclis.org/5464/>.

Tennant, Roy. 2007. "Lipstick on a Pig 2.0." May 4. <http://www.thedigitalshift.com/2007/05/roy-tennant-digital-libraries/lipstick-on-a-pig-2-0/>.

Tennant, Roy. 2014. "The OPAC is dead." February 6. <http://www.thedigitalshift.com/2014/02/roy-tennant-digital-libraries/the-opac-is-dead/>.

The University of California libraries. 2005. *Rethinking How We Provide Bibliographic Services for the University of California*. Final Report. <https://libraries.universityofcalifornia.edu/groups/files/bstf/docs/Final.pdf>.

Tillett, Barbara B. 2005. "FRBR and Cataloging for the Future." *Cataloging & Classification Quarterly* 39, 3-4:197–205. DOI: [10.1300/J104v39n03_12](https://doi.org/10.1300/J104v39n03_12).

Turbanti, Simona. 2014. "Cui prodest libraries authority work." *JLIS.it* 5, 2:49–59. DOI: [10.4403/jlis.it-10062](https://doi.org/10.4403/jlis.it-10062).

W3C. 2012. "OWL 2 Web Ontology Language. Primer." Second Edition. W3C Recommendation 11 December 2012. <https://www.w3.org/TR/2012/REC-owl2-primer-20121211/>.

Weston, Paul Gabriele. 2006. "Il catalogo. Dalla tradizione ai nuovi servizi." In *Biblioteche e informazione nell'era digitale. Atti del convegno della IV Giornata delle biblioteche siciliane*, Ragusa, 26 maggio 2006, 56–82. Associazione Italiana Biblioteche: Sezione Sicilia. <http://eprints.rclis.org/19468/>.

Weston, Paul Gabriele. 2011. "Dall'OPAC tradizionale ai cataloghi di nuova generazione." In *Opac, blOpac, socialOpac. Da catalogo elettronico a strumento di cooperazione e social network*. Treviso: Regione del Veneto e Provincia di Treviso. http://www2.regione.veneto.it/cultura/cms/allegati/Biblioteche/Opac_bOpac_SocialOpac.pdf.

Willer, Mirna, and Gordon Dunsire. 2013. *Bibliographic Information Organization in the Semantic Web*. eBook. Oxford: Chandos.

Working Group on the Future of Bibliographic Control. 2008. *On the Record. Report of The Library of Congress Working Group on the Future of Bibliographic Control*. <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.