OSSERVATORIO

Old wine, new bottle? Principles and methods for a true innovation in LIS perspectives.

The view of Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer

prepared by the AIB Study group on
Cataloging, indexing, linked open data and semantic web (CILW)

Premise

This article consists of interviews with personalities of Library and information science (LIS) field, realized asking three questions to five internationally recognized scholars. Through these essays, the AIB Study group on Cataloging, indexing, linked open data and semantic web (CILW) wants to set up its activity starting by a detailed evaluation of the 'classic' and 'innovative' theories and principles, and of the views of scholars on them. It is possible, thus, to review current development of information and resources description and organization, in order to generate a fruitful debate and a necessary comparison also with disciplines neighbor to LIS, starting from theorists and professionals that are most focused on the issues of change, whose research activities are also linked to the 'paradigm shift' warned in our disciplines.

So, questions developed by the group are intended to call attention to some fundamental questions. First, it would help clarify ideas about the role a catalog may still have, in libraries and for the society: if it will continue to be a central tool for research activities, or whether its role will continue to decline until it is aligned or overshadowed compared to the other current tools for information research and discovery. The catalog, thus, is in front of the semantic Web, the linked data methodology, and the need to open up to other communities, in order to be the preferred tool for users.

AIB CILW STUDY GROUP (Cataloging, indexing, linked open data and semantic web), composed by Roberto Raieli, Marinella Cisternino, Elena Corradini, Enrico Francese, Antonella Iacono, Roberto Morelato, Antonella Trombone, website <www.aib.it/struttura/commissioni-e-gruppi/cnc>, email cnc@aib.it.

The lack of effectiveness of catalogs in information searching is often due to the transformations of the search objects, namely the knowledge resources: it is necessary, then, wondering about what might be a definition for the new object of the catalog, what are actually the resources of knowledge, and what tools are more suited to treat them. Resources useful to knowledge today are very different from each other and from what has so far seen as 'document', their main quality is the interest users have in them, they are placed in multiple webs of meaning and belong to multiple contexts.

Finally, regardless of the technological tools best suited to the needs of society, it is necessary clarify policies and methodologies desirable for the treatment of resources, information and data, that remain the primary object of any knowledge organization and management tool: it is increasingly important that data are correct and reliable, and this need grows as much as it increases their possibility of growing and dissemination. Description should be decentralized, being confident of the work of others, fostering integration in the semantic Web, but without losing the data credibility.

In view of the possible answers, it is always essential to maintain an appropriate balance between 'conservation' and 'revolution'. Seen from different perspectives, the basic principles, more or less classic as they are, can be reconsidered at a theoretical level, based on their strengths and critical points, in order to set up a project that can be said to be truly innovative in terms of theoretical and methodological aspects for our area of investigation. A real progress and a real innovation are obtained not only by 'overcoming' principles and practices already obsolete or just in line with outdated realities – as the wine becoming old and spoiling –, but also by developing and 'reconsidering' that stable principles, unrelated to temporal or epochal shifts, which remain valid for the maintenance and development of library and information science disciplines – as the wine aged in solid barrels.

If the underlying principles are not adequate, up-to-date and at the same time durable, it's no use proposing a new container, be it technological or conceptual, to try to present them as a real methodological innovation.

Interview

Catalog and the 'search'

Is the catalog, as we know it today, destined to lose its central role among other knowledge and search tools pertaining to libraries?

New information search and retrieval tools allow the expansion of selecting and mediating activities far beyond the classical boundaries held by the library, towards the endless network of information and resources available via the Web. But, if establishing reliability and authority of many of its contents is not always easy, the idea behind the semantic Web is, however, the creation of directories from which to capture data through controlled and reliable sources. In this new perspective, can a 'unique', shared and participated catalogue exist? Can an open source catalog exist, populated and maintained by experts with new skills, where cataloging and creation of ontologies and dictionaries will approach more and more?

A current concept of resource

In librarianship and library science the concept of 'document' was a steady element for a long time. Now the *International cataloguing principles* (ICP) restrict this definition to some archival resources. Moreover, in today's cultural scenario, it is necessary to include different kinds of objects that can be recorded, and, as a consequence, described and used as reliable information, and possible sources of knowledge. In

this perspective, which could be a viable and functional definition of the new concept of 'resource', and what could be its identity?

In short: what is the object we have to deal with? How can we look at something 'fluid' and objectify it into a record, a catalog, a reference system? What do we consider as an information unity to be described and linked? A forum thread and/or a single post? A symphony or each of its movements?

'Authoritative' data

By applying linked open data (LOD) to catalogs and other information search tools, libraries and other cultural heritage institutions will have to become accustomed to not having the 'ownership' of data with which to describe and mediate resources. This results in a true paradigm shift that renews the documentation tradition, and digital tradition on the whole, from the definition of the concept of information. In this scenario, how can the different cultural institutions maintain the role of authority in the process of production, dissemination, retention and maintenance of data, and being the gatekeepers for their quality?

Will the different institutions be able to integrate well into the semantic Web project, and develop a more accurate interpretation of their own identity?

The view of Marshall Breeding

Through linked data to resource discovery

In many respects, the catalog has already undergone a major shift from its traditional form of an online search tool focused on the print inventory of a library. Today, libraries work to provide their communities with broader discovery services that address the full breadth of their collections, spanning all types of materials. Unlike online catalogs that returned titles of books or journals, it's essential now also to provide access to individual articles, book chapters, or digital objects and to display full text or visual representations whenever possible. These discovery services must pay close attention to user experience, providing access to library content using the interface conventions and techniques to which the public has become accustomed in their daily experience of the Web. These interfaces must also be optimized for the mobile devices or tablets that represent the majority of use rather than the larger screens of laptop and desktop computers. The genre of index-based discovery services has found a great deal of acceptance in academic libraries, with large proportions adopting them in addition – or increasingly instead of – traditional online catalogs. Public libraries demand discovery services or catalogs that provide equilateral access to e-books as well as print books. Through initiatives such as ReadersFirst, public libraries have demanded and suppliers have provided dramatically simplified methods to discover, select, and download e-books to reading devices.

These new types of discovery interfaces represent only one thread of activity. Linked data and semantic Web technologies have also gained increasing interest. The universe of resource of interests to libraries exposed as linked data has grown considerably in recent years, through the majority of the body of scholarly content remains locked within proprietary publishers. Libraries have made important progress toward shifting from record-oriented bibliographic description, primarily though the various MARC formats to constructs based on linked data, especially BIBFRAME. Developers of many library management systems and discovery services have begun work to incorporate BIBFRAME, though we remain more in a time of experimentation and prototypes than operational products. I anticipate hybrid systems that will use relational databases and indexes for transaction-oriented business processes but that will increasingly tap into the universe of linked data to supplement discovery, access to content, and visualization of results.

Semantic Web technologies will also help libraries improve the discoverability of resources of interest to their communities. We must acknowledge the reality that most users do not come to the catalogs or interfaces that libraries provide, but rather rely on general search engines. The incorporation of semantic coding in the presentation of library resources, such as defined in schema.org, dramatically improves the discoverability of resources beyond library interfaces. Since patrons don't come to the library, the library must work hard to make library resources discoverable and available in the places on the Web they inhabit.

MARSHALL BREEDING, Library Technology Guides, marshall.breeding@librarytechnology.org.

This transition from record-orient catalogs to semantic Web technologies does not necessarily mean less control for librarians in the information ecosystem. Libraries will continue to curate and describe local collections, but increasingly using metadata structures based on linked data, RDF, XML, and less using library-specific records and protocols of little interest to the broader information ecosystem. While MARC has served libraries well in some ways, it has also led to some degree of isolation relative to publishers, the e-commerce sector, and other information-oriented industries. Adoption of the semantic Web may result in improving the relevance and position of libraries. It also means just as much work as ever in creating descriptions for resources that relies on authoritative sources, but increasingly expressed as links and relationships and less as self-contained records. The core roles of libraries have persisted throughout many phases of society and cycles of technology. How librarians carry out that work evolves in accordance with changes in the broader environment.

The incredible diversity of resources

Library collections today include diverse types of materials. While print documents, monographs, photographs, and manuscripts constitute a part, libraries also collect many different types of digital content. Academic libraries, for example, generally expend the vast majority of their collection funds on access to scholarly and professional articles provided through subscriptions to content packages offered by commercial and non profit providers and take advantage of the growing body of open access materials. Digital objects form another vital part of library collections. Many libraries have longstanding programs in digitizing books, newspapers, manuscripts, photographs, and other items of historic or academic interest. Cultural heritage is increasingly represented in native digital form. Professionals in libraries, archives, and museums routinely describe materials in all types of media and formats, taking advantage of a variety of applicable metadata standards and schemas. Any formally definition of what constitutes a document or set of cataloging principles must encompass the incredible diversity of materials and their corresponding metadata conventions. No single static definition of what constitutes a library resource can necessarily accommodate the continually expanding range of materials collected and curated by libraries and related cultural institutions.

Decentralization of bibliographic description

Increased involvement by libraries and related institutions in the creation of tools based on linked open data represents a positive step since it provides increased opportunities for the discovery of library-oriented resources. This transition may mean a more decentralized model of bibliographic description. The elements of bibliographic description may reside in RDF triples instead of MARC records stored in centralized databases, but the same type of intellectual effort will be required to support the management and discovery of collections. National libraries, regional consortia, and organizations such as OCLC will continue to play seminal roles following the anticipated transition to linked open data. Other players may also emerge. The concept of ownership of bibliographic data may well be further diminished. We are already seeing more freedom of exchange and less assertion of ownership as bibliographic records become more of a commodity and as expectations grow for collaborative exchange. Projects such as Europeana require contribution of bibliographic records under the most liberal Creative Commons license (CC0) regardless of whether they derive from OCLC or other sources that previously asserted some degree of ownership. The eventual transition to increased involvement in

linked open data will not necessarily fundamentally change the role of libraries in creating and describing collections, but will hopefully provide a variety of benefits. The creation of high-quality metadata will likely continue to take place in ways that can be collaboratively distributed throughout the global library community. Even as the containers that convey bibliographic description change dramatically, the operations, strategies, and values of libraries will endure.

The view of Mauro Guerrini

Better tools for users

This issue covers the topics of the 'catalog' as having a central role, the need for 'open source' catalogs and catalogers with new skills creating ontologies and dictionaries. A central role of a catalog is certainly true as the institution's tool to know what it has and where its resources are located so its users (in the largest sense of the word) can know what exists and how to get to it (find, identify, select, obtain... – user tasks of FRBR). Is an 'open source' catalog as far as we can see into the future? Isn't this an opportunity to explore better tools for users – what would be the ultimate device to connect users to all the information they need? – not just the limits of Google search engine resources, but the wealth of resources in collections and institutions willing to share their resources.

I believe that we should not confuse the traditional local OPAC (which was an evolution of the printed catalog) with the integrated search tools and social type of new generation, which is starting to be possible. What we hope for the future is instead to change from the 'autistic' catalogs of many libraries to a more cooperative environment. I believe it is necessary to have the courage to abandon old paradigms of cataloging, and I believe we need to take really the model of shared cataloging by the linked data.

I know work is going on with the semantic Web and linked data, and I think it is important to see how that works out, but it is very important that the results give us no less than we have with traditional cataloging in terms of helping users select and identify what it is that is available.

I believe, also, that the hypothesis of a catalog 'unique', or 'shared and participated catalog' is likely to be impossible, and may uncorrect. I believe, instead, that it is preferable to focus on the concept of having many catalogs, nature, type and field different, but connected and united by the use of consistent data structuring, the RDF precisely, for now.

The object of interest for users

As far as one can read, ICP never uses the term 'document', neither in its text nor in its glossary. I think this is due quite to the possible misunderstanding between document as usually intended in LIS, and document in archival science. The aim of ICP was to be much broader (and more precise) to include all possible resources – things collected not only by archives, but also libraries, museums, and all other cultural institutions. The translation of the term 'document' into different languages from English is also problematic, as the term as used in other languages has many connotations – another reason for ICP to reject the continued use of that term.

MAURO GUERRINI, Università degli studi di Firenze, Dipartimento di studi sul Medioevo e il Rinascimento, piazza Brunelleschi 3-4, 50121 Firenze, e-mail mauro.guerrini@unifi.it.
I would like to thank Barbara Tillett for her valuable suggestions; I also compared with Carlo Bianchini, Andrea Marchitelli, Dorothy McGarry, Valdo Pasqui, and Tiziana Possemato.

The preferred concept – the more universal and hopefully more unambiguously understood term is used in ICP – is 'bibliographic resource', which admittedly includes the problematic term 'bibliographic', but that also is noted to be beyond just books to be any of the resources collected by cultural institutions. Maybe the next generation can find a more acceptable more encompassing adjective that is well understood in the future.

Whether 'resource', or 'bibliographic resource', the term will be clear to all in the future. I think those terms were an attempt to use a neutral term that could be used to imply all types of materials in libraries (and in other institutions) in a clearer manner. Bibliographic resource is defined as «an entity within the realm of library and similar collections consisting of the products of intellectual or artistic endeavor» (ICP Glossary). A very important addition is that «bibliographic resources in the FRBR model are the Group 1 entities: work, expression, manifestation, and item», because this is the joining link among ICP principles and the international standard RDA, where the term resource loses the adjective 'bibliographic' and is used to identify "a work, expression, manifestation or item" (RDA Glossary). The term includes not only an individual entity, but also aggregates and components of such entities: «(e.g., three sheet maps, a single slide issued as part of a set of twenty, an article in an issue of a scholarly journal)». By this approach, and generally speaking, resource means the object of interest for the users; as far as it applies to libraries, archives, and museum and to all other cultural institutions, resource is a comprehensive term that is intended to include old and new books, documents, objects and finds of any kind.

The confidence in the work of others

This issue is about LOD and loss of 'ownership' or loss of being able to authenticate the data being provided to users, so librarians and users can trust what they are being told. At first glance this seems like the 1980's when some catalogers were more concerned with the poor quality of cataloging done by other people than thinking of the value of providing access to collections to users. The alarms of early shared databases pointed out the flaws in cataloging records when they were being shared beyond one's own catalog – especially OCLC at the beginning. The percentage of mistakes and errors was actually very small – human error, quite frankly what could be found even when one examined one's own catalog that had been created by many people over a long time. So, the solution and attitude was to 'just fix it' when it was discovered, and over time the quality of the shared database continued to improve – but the catalogs will never be perfect, so we must just strive to keep making it better. And doing that work in a shared environment is a great benefit – as more people can help. If you never trust or value the work of others, you will not ever be happy with any shared solution. Unfortunately, I still find that mind set or attitude in some catalogers.

As to whether the current technology of semantic Web and linked open data is (or can be) the answer, it is what is currently the 'best' technology for international sharing of data, reducing the individual costs to libraries and institutions through a shared endeavor. Our primary roadblock, as I see it, is the lack of an international shared system that is freely accessible and maintainable by all, perhaps that is my optimism, that such shared systems are good, but I believe that working with those in the information community beyond libraries is extremely important just now so we can all raise the quality of service to users.

In the future I believe that the identity of the libraries should move from the metadata encoded in the catalog to that of institutions that provide and promote value-added services for the environment and relevant institutions. In any case, we

have to provide the quality of the content of the bibliographic data. It doesn't help the user if incorrect access is provided. I believe in value-added services, and these must be built on accurate data.

I think we should take note of the great innovation of the RDA, *Resouce description and access*; the guideline pays great attention to the quality of the 'data' and relationships between entities, rather than emphasizing the record. RDA's focus on both description and access gives a great deal of attention to the accurate identification and description of the single entity, but also enables the cataloger to put that entity in the context of the rest of the bibliographic universe through the relationships and access points. Thus, using RDA and following the ICP principles upon which it is based should lower the risk of very bad quality of the records (long descriptions, descriptions very linked to the locally used, and written in complex languages), in favor of the quality of individual data and therefore to its reusability (interoperability) in other domains. In short, the RDA principle of «take what you see» from the resource is connected with providing an 'accurate description'. That is complemented by the RDA focus on providing 'access' through the associated or related persons, corporate bodies, families, concepts (controlled subject terms), and relationships (especially to related works). I believe that we need to concentrate our energies and attention to the creation of authorized (and preferred) access; this attention should trigger a virtuous cycle of creation of data quality, re-usable by everyone in any context and continue to emphasize the great value of cataloging (it can continue to call this or, if you prefer, call 'metadating') to enable users to connect with the resources they want and to discover related resources they may find helpful.

The view of David Weinberger

A catalog evolved for the net

Traditional catalogs have been too limited in the amount of information they can capture, and have required centralized, uniform decisions. Overall we seem to be headed away from such systems and toward decentralized webs of information.

This is part of an ongoing change from the basic strategy for dealing with metadata by treating it like a shadow object that is attached to the 'real' object that it describes. Instead, especially with linked data, we get swirling clouds of metadata molecules that do not give priority to any particular type of object as being the 'real' one. For example, if a researcher cares about a particular book, that becomes the center of a web of linked information. If she cares about works published in a particular city at a particular time, those become the points around which data gathers. If she cares about works that mention Michelangelo, then that becomes the center... until her interest shifts.

But there is room for something like a catalog in this. These linked data clouds are well served by the existence of persistent identifiers. These enable references within the cloud to be identified as talking about the same thing. Without this, the soup is a thin consommé, whereas we need a hearty, chunky broth. An authorities file can serve as a type of catalog, providing persistent identifiers and the basic information that can be anticipated as useful to many users.

With persistent and reliable identifiers, a linked data soup can be turned into a 'graph' in which some entities are indeed privileged so that relationships among them can be determined and made navigable. A library graph presumably would include entities such as works, authors, dates, places, publishers, etc. Such a graph could be considered to be a catalog evolved for the net, where the aim is not to reduce information about a work to an essential set but to enrich that information by linking it into a vast context.

Resources as part of multiple webs of meaning

Even in the physical world, a thing is a web. A book is only a book because it's part of multiple webs of meaning. That is, we couldn't understand that this thing is a book if we didn't also know that it has two-dimensional content inside of it, that there are pages that go in a sequence even if the content is non-sequential, that the content was put together by a human, that the content came before the book, that the book was manufactured, that it was published by an organization, that that organization has some physical address, that the book has weight, that it has economic value, that it will have to be put somewhere, etc. These webs of meaning intersect and eventually can be linked out to everything in the world because the world is itself the totality of these webs of meaning (this is Heidegger straight out of *Sein und*

DAVID WEINBERGER, Harvard University, Berkman Center for Internet and Society, 23 Everett Street, Cambridge MA 02138, e-mail david@weinberger.org.

Zeit). As David Lankes says[1], we could even count the conversations around these objects as catalog-able.

Catalogs have traditionally chosen what their central objects are – books or other physical items. This choice reflects an anticipation of how most users will want to navigate the collection's items. It also reflects the logistical challenges owners and collectors face, since they have to find a place for each of these items. It also reflects the way the economic transactions are structured: you can buy a book but not an author or the city in which the publisher has its headquarters. These are all good reasons for structuring a catalog around objects.

But in the computer age, we can do what software developers call 'late binding'. In this case that would mean letting users of the catalog make the choice about which objects are primary at the moment they come to the catalog with a project in mind.

A data structure like linked data represents the state of the system far better than any one crystallization of that state, whether it's a traditional catalog that starts with the items, or any users particular view into it. It's better to let the user decide than decide for her – better in that it accommodates more uses, including unanticipated ones.

Of course, constructing such an ur-catalog is not small matter. And doing so does require making choices for the user about which types of information and relationship might be useful. But it requires making fewer choices ahead of time, which increases the number of choices that can be made by the users of this new type of catalog.

Gaining credibility

In this new ecosystem, institutions can choose between controlling access or being visible. Assuming that they want to have authority that matters – that is, authority that's visible on the net – they will gain that authority not by gatekeeping but by being taken as having credibility.

Looked at from the other side, there is also the question of how those who use the information can have confidence that the data is credible. As linked data triples escape into the wild, they well may lose data about their provenance. But in a linked open data world, it should be possible for researchers to check data at reliable institutions, or simply to get data directly from those institutions.

OCLC's library knowledge graph is taking a promising approach. The atomic elements of linked data ideally consist of URLs (or URIs, to be precise) that point to some reliable, public resource. The elements coming out of the OCLC project will point generally to OCLC resources. Those URIs will have a base address of 'www.oclc.org' (approximately), so that any computer or human who looks at them will immediately know their source. Knowing the provenance of the data they find on the net will continue to be an urgent need of serious researchers.

1 R. David Lankes; Joanne Silverstein; Scott Nicholson, *Participatory networks: the library as conversation*. 2006, <http://quartz.syr.edu/rdlankes/Publications/Others/ParticiaptoryNetworks.pdf>.

The view of Paul Gabriele Weston

Measuring the catalogue's success

In recent years the expression 'catalogue 2.0' has often been used, but hardly ever, I am afraid, in the most appropriate way. From time to time the term has designated: the catalogue enriched with reproductions of the covers and part of the contents, the existence of facets (which should be more appropriately defined search filters) as devices to aggregate in various ways descriptions with similar characteristics (the presence of the same person, the use of a certain format, the nature of the bibliographic publication and so on), the possibility for the user to add comments and ratings, to share descriptions on social networks and, in case of e-books, direct access to contents. *Catalogue 2.0*[1] is also the title of a collection of contributions produced by international experts such as Lorcan Dempsey, Emmanuelle Bermès, Marshall Breeding and Karen Calhoun. The publication, edited by Sally Chambers, is introduced by the question: there will still be a library catalogue in our future and if so, what will it look like?

From the printed to the electronic catalogue, passing through the filing cabinets, the application of information technologies has offered from time to time new research opportunities to readers, at the price of an at least partial rewriting of cataloguing codes. However, as Dempsey observes, Web environment requires the catalogue to be reconfigured in such forms as to make it a non-individually identifiable element of the library service. This is undoubtedly a consequence of the transition from a physical distribution of information and documents to a digital one (workflow switch), but also reflects the users interest shifting from the local boundary ('my' library, 'my' store, 'my' city, 'my' Country) to the network as a whole (we turn to Google, Amazon, Expedia and so on) (attention switch). The two combined impulses tend to continuously push the catalogue user outwards, in other words exercising a centrifugal force which tends definitely to deprive the cataloguing tool of that visibility it has enjoyed in the past and, at the same time, to undermine its more established functions. Already in 2006, reflecting on the role of the catalogue in the information economy, Dempsey wrote: «Libraries have rich deep collections, and the aggregated library system is a major achievement. However, in our current network environment, libraries compete for scarce attention. This suggests that if the 'library long tail' is to be effectively prospected then the 'cost' of discovering and using library collections and services needs to be as low as possible»[2]. Therefore Dempsey places emphasis on the issue of logistics, a component of the services that in the complex organization of contemporary society becomes increasingly important for the need to combine supply and demand within a network of a potentially infinite number

PAUL GABRIELE WESTON, Università degli studi di Pavia, Dipartimento di studi umanistici, p.zza del Lino, 27100 Pavia, paul.weston@unipv.it.

1 *Catalogue 2.0: the future of the library catalogue,* edited by Sally Chambers. London: Facet, 2013.

2 Lorcan Dempsey, *Libraries and the long tail: some thoughts about libraries in a network age*, «D-Lib magazine», 12 (2006), n. 4, <http://www.dlib.org/dlib/april06/dempsey/04dempsey.html>.

of subjects. Link resolvers, which make possible data discovery, as well as the delivery of information and documents through a chain that does not have, on the user side, apparent interruption between the time of research and the time of use, can certainly be seen as an example of logistics applied to the world of libraries. In this perspective, the success of the catalogue is measured by the ability to coordinate and aggregate more efficiently, to the benefit of users (and of its own future), the demand and supply of services, and doing it bearing in mind the need to maintain that high quality standard of that is recognized to the cataloguing mediation (and so far awarded), even in an age in which other types of resources have definitely taken over as the primary means of searching the Web.

The strategic choice of quality

I personally think that quality is not an option, but a strategic choice. We can refer to the metaphor of the Blue ocean strategy (BOS), an economic theory illustrated in 2005 by W. Chan Kim and Renée Mauborgne[3], whose application involves identifying areas of growth and in providing services not yet covered, or not covered adequately, a scenario that is opposed to the other, the one in which «the excess of players leads to a ruthless competition that will turn the red ocean bloody». Pragmatically, this strategy implies that the most powerful players of the Web, whose strength relies on the huge mass of data they handle, should not be confronted on their own ground. On the contrary, priority should be given to criteria such as data reliability, pointing at the same time to the integration of data and services offered with complementary digital resources freely accessible on the Web.

With the introduction of the electronic catalogue, originally conceived as a tool for managing the administrative control of the library and the borrowing procedures, the library has reduced its capacity to offer pathways not only within his own documentary universe, but also towards the exterior (and I am not only referring to other catalogues), a capability that in the printed books environment was guaranteed by the availability of refernce material and, in the more advanced libraries, the skills of the librarians. The ability of the catalogue to act not just as an arrival point, but also as an intermediate node, whose added value mainly consists in the selectivity and the reliability of connected resources, responds to the generalized behaviour of the Web users, which tend to conceive search tools and information resources as part of a single hypertext, rather than as an individual device aimed at providing just answers.

That said, it should be added that to enhance the catalogue in the indicated direction is not by itself a sufficient strategy to ensure adequate visibility to the tool. To overcome the risk of opacity on the Web one needs to have a critical mass of data (hence the aggregation in bibliographic tools of increasing size in which data and resources are integrated in depth and/or the implementation of systems pursuing the highest interoperability with external resources) and enjoy at the same time of a specifically defined profile, which does not always imply the existence of a huge universe of documents, but is rewarded by the choice to scan a well-defined horizon (as Edit16 does and more generally catalogues related to highly specialized documentary fields do). The opportunity to be a system requires organizational and planning skills, as well as cockpits that in past years our Country has not proved to have, a situation which the wide dispersion of the cultural heritage on the territory

3 W. Chan Kim; Renée Mauborgne, *Blue Ocean Strategy: how to create uncontested market space and make the competition irrelevant*. Boston: Harvard Business School Press, 2005.

and among institutions of different belonging makes even more critical. This applies even more to the digital library initiatives, which, at least at this stage, cannot compete with similar initiatives by Google and many other institutions rich in historical heritage with a long activity in populating the Web of popular cultural contents. Selectivity in the choice of documents and collections to be scanned should be privileged as a rule.

Integrating the catalogue in the digital ecosystem

Returning to the issue of the relationship between the catalogue and the Web, we must reflect on the fact that, despite technological developments and new services mentioned above, the catalogue remains a tool modelled on the philosophy of Web 1.0, a scenario in which data is only moved from the centre to the periphery, with a clear division between those producing information and the ones who benefit from it. Beyond a few and nevertheless significant exceptions, catalogues still adopt a stellar architecture, as is demonstrated by the use of proprietary formats, impenetrable to search engines and unintelligible in environments different from their own. This has confined catalogues, even the largest, in the darkness of deep Web. The idea that each record can have its own story, made of catalographic choices discussed among specialists, changes, aggregations and mergers, transfers from an archive to another by electronic devices, comments and additions made by readers, and that this story becomes an integral part of the record accompanying it in its itineraries, has not been given sufficient attention so far. The cataloguing system closest to this model is currently Open Library[4], the creation of which is due to the collaboration of Karen Coyle. The solutions adopted in Open Library could somehow foreshadow the functioning of a cataloguing system based on BIBFRAME[5].

The idea that readers can enrich the content of the archive, both improving the granularity of the information and making explicit the nature of the relationships linking the entities which are described, is underlying the Linked Jazz project directed by Cristina Patuelli[6]. Crowdsourcing is used to allow lovers of jazz, a community generally highly motivated and competent, to share their memories regarding musicians (as posters, autographs, photographs, sound recordings) that can help to shed light or to better define existing relationships among musicians themselves. In this way, the project achieves at least two results: in the first place, the connections, mostly expressed by generic statements such as 'played with' or 'has been a friend of', are implemented by accurate factual information (they played together on that occasion, when a musician noticing the presence of a colleague invited him or her to play a certain piece); on the other hand, by taking roots in the jazz community, the project gains in terms of identity, authority and vitality.

Another service that could be widely enriched by integrating the catalogue by means of information produced within a different system is storytelling. The reference to one of the many areas of cooperation between libraries and Wikipedia, a cooperation

4 <https://openlibrary.org/about>.

5 Antonella Trombone, *Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicativi dei dati bibliografici*, «AIB studi», 55 (2015), n. 2, p. 215-26.

6 <https://linkedjazz.org>. Leanora Lange; Maria Cristina Pattuelli, *Linked Jazz: building with linked open data*, «EDUCAUSE review online», 30 June 2014, <http://www.educause.edu/ero/article/linked-jazz-building-linked-open-data>.

viewed with increasing interest, leads us to consider the possibility of enhancing relationships between works or authors, for example, by intertwining their respective authority records with the Wikipedia entries that widely illustrate these relationships, so that the reader does not face a number of links insufficiently explained, but can benefit from a comprehensive presentation which may produce references to the works as well as the occasional serendipitous discovery[7].

Many examples could be produced in order to demonstrate how the catalogue can evolve in reason of its more or less successful integration into the digital ecosystem. I shall reserve a final mention the strengthening of the service of digital reference. Here is the mandatory reference to data.bnf.fr, a tool kit digitally created by National Library of France by aggregating information dispersed in several catalogues and throughout the digital library Gallica[8]. The project, which won its creators the Stanford prize for innovation in research libraries (SPIRL), assumes the use of open data and the compliance to Web standards. A similar service, albeit in a completely different context, is offered by Wikipedia personensuche[9], a device for efficient use of the names of people present in the Wikipedia pages. The tool relies on VIAF clusters to identify informational resources on the Web and in library catalogues and networks partaking in VIAF, in addition to those found in Wikipedia in a relationship of some kind with the entries dedicated to that subject. Thus, for each individual, the user is provided with a list of other individuals in various ways connected (kinship, profession, partnerships etc.), and a list of resources (biographies, periodicals, collections of scores, illustrations, specialized catalogues etc.) that it is dynamically updated and offers a wealth of ideas for pursuing research by enhancing his or her knowledge on the matter.

In conclusion, I'd like to stress the importance of the contribution offered by libraries over the years. Most developments were made possible thanks to the high quality of their work and the availability of a huge amount of structured data. To carry out their function and have a meaning data must not be considered in a monadic way, deprived of their context, for they belong to that articulated and complex structure which is the bibliographic description. It is this structure, which evolved through the secular experience of librarians, the added value of catalogue information on the Web.

7 A working group (Camilla Fusetti, Agnese Galeffi and Antonella Trombone) is considering the application of storytelling to the enhancement of complex works in applying the model FRBRoo.

8 <http://data.bnf.fr>. Romain Wenz, *Linked open data for new library services: the example of data.bnf.fr*, «JLIS.it», 4 (2013), n. 1, <http://leo.cineca.it/index.php/jlis/article/view/5509>.

9 <https://tools.wmflabs.org/persondata>.

The view of Maja Žumer

Open data to other communities

Most librarians, when asked this question, would, without hesitation, say that of course the catalogue as we know it now is here to stay. This strong belief is based on centuries of the monopoly of libraries as major information providers and the catalogue as their main tool. But on the other hand we have evidence that library catalogues are perceived as unintuitive, confusing, and difficult to use by library users. The first warnings were already published in the 1980's (such as Borgman[1], and Matthews *et al.*[2]), in parallel with a broader implementation of computer catalogues. At that time the situation was not alarming yet, because the users did not have a clear alternative in terms of information sources. The Web, browsers and other tools have dramatically changed the information landscape and we now see reports that a vast majority of users start their enquiry with a search engine (and none on the library web site)[3] and avoid using the catalogue even when they know they want to borrow a book from the library.

Libraries have a tradition of producing and maintaining high quality metadata about their resources. Bibliographic and authority records created in libraries are without question the best metadata about publicly available information resources and other entities associated with them, such as agents. So the main question is why this data is not used to its full potential. Or, as my colleague Tanja Mer un said recently: «Our data is on vacation. We have to make it work harder!».

In order to fully utilise the wealth of their data, libraries will have to promote it and above all open the data to other communities. The semantic Web technology is the technical infrastructure needed, but to fully use it, clear conceptual models are essential. Conceptual models not only clarify the domain modelled, but also enable mutual understanding of different domains. And this is the window of opportunity for libraries. By implementing commonly agreed conceptual models and opening their data (currently held in domain-specific MARC formats) to other communities will foster the use and reuse of library data. Thus the libraries will regain and maintain their position as major players in the semantic Web.

The general notion of document and the context

Since collecting, selecting, organising and providing access to recorded knowledge is seen as the main mission of libraries, 'document' is the central theoretical

MAJA ŽUMER, Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva 2/V, SI-1000 Ljubljana, maja.zumer@ff.uni-lj.si.

1 Christine L. Borgman, *Why are online catalogs hard to use? Lessons learned from information retrieval studies*, «Journal of the American Society for Information Science», 37 (1986), n. 6, p. 387-400.

2 *Using online catalogs: a nationwide survey*, edited by Joseph R. Matthews, Gary S. Lawrence, Douglas K. Ferguson. New York: Neal-Schuman, 1983.

3 Online Computer Library Center, *Perceptions of libraries, 2010: context and community. A report to the OCLC membership*. Dublin: OCLC, 2011, <http://oclc.org/reports/2010perceptions.en.html>.

construct. In the most general sense, a document is a recorded representation of thoughts. While I will not discuss Susanne Briet's example of antelope as a document[4], it is clear that the notion of document is general enough to cover both the traditional formats and digital resources, even emerging ones.

Functional requirement for bibliographic records (1998)[5], the conceptual model of the bibliographic universe, defines four entity classes according to the level of abstractness (item, manifestation, expression, and work). Using a simple example of a textual resource: my copy of a particular edition of Shakespeare's *Romeo and Juliet* is an item. The set of all identical items is a manifestation. But a book is not just a physical object, we value it for its content. The signs, words and sentences in the book (Shakespeare's original text, in this example) are an expression and the ideas represented by these signs is the work. FRBR also defines agents and relationships, connecting these entities.

Over the last years several user studies (Pisanski and Žumer[6]) have confirmed the intuitiveness of FRBR, proving that this model is an appropriate mechanism for describing resources, or documents, in a way that supports exploration and enables meaningful clustering. The traditional primary focus on the carrier is replaced by a more sophisticated and flexible view. Which brings me to the famous discussion of Lubetzky and Verona regarding the relative importance of bibliographic and literary unit, as they called them then, for the user in the context of the card catalogue and its limitations. Since, depending on context, one or the other may be of interest to a particular user, it is important to support both views and present the appropriate one for each particular situation. The technology is not a hurdle anymore and the libraries should embrace FRBR as their conceptual basis for the management of resources. *International cataloguing principles* need to take into account this new conceptualization of the bibliographic universe.

Integrate into the semantic Web

There is a substantial overlap in both the user population and types of resources of the cultural heritage institutions (libraries, archives, museums), but the three domains have, for historical reasons, developed their own conceptualisations, tools and practices. The technologies of the semantic Web for the first time offer the infrastructure which overcomes the differences and creates a common view of their respective resources. With clearly defined conceptual models and domain-neutral standards (such as LOD) cross-domain implementations are possible without forcing any of the domains to adopt a different paradigm. While I am clearly calling for the reuse of library data by other domains, there is also a clear advantage of reusing and integrating data from other domains into the bibliographic ecosystem. Several novel bibliographic information systems are, for example, already importing and integrating information about authors from DBpedia, reviews from Amazon and adding user-provided data such as ratings and tags.

4 Suzanne Briet, *Qu' est-ce que la documentation?* Paris: EDIT, 1951.

5 International Federation of Library Associations and Institutions, *Functional requirements for bibliographic records: final report*. München: Saur, 1998.

6 Jan Pisanski; Maja Žumer, *Mental models of the bibliographic universe*, «Journal of documentation», 66 (2010), n. 5, p. 643-667, 668-680. *Id.*, *User verification of the FRBR conceptual model*, «Journal of documentation», 68 (2012), n. 4, p. 582-592.

The question therefore is not whether the cultural heritage institutions will be able to integrate into the semantic Web – they have to. The only alternative, I am afraid, is the road to oblivion.

Ai fini della citazione, utilizzare esclusivamente il testo in lingua italiana, che presenta il DOI, la paginazione, l'abstract e gli altri dati ufficiali.

When citing, please always refer to the Italian translation only, complete with DOI, page numbers, abstract and other data.

[Gruppo di studio AIB CILW, *Old wine, new bottle? Principi e metodi per una reale innovazione nelle prospettive LIS. Il parere di Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer*.
AIB studi, vol. 55 n. 3 (settembre/dicembre 2015), p. 385-403. DOI 10.2426/aibstudi-11384]