

UBC reloaded: remembrance of things past, back to the future

English translation of the original paper: “A la recherche du temps perdu, retour vers le futur: CBU 2.0”

Translated by: Mélanie Roche, Bibliothèque nationale de France. Département de l'Information Bibliographique et Numérique, Paris, France.

The text of this document has been translated into English and differences from the original text may occur. This translation is provided for reference purposes only.

Gildas Illien

Department of Bibliographic and Digital Information, Bibliothèque nationale de France, Paris, France.

E-mail address: gildas.illien@bnf.fr

Françoises Bourdon

Department of Bibliographic and Digital Information, Bibliothèque nationale de France, Paris, France.

E-mail address: francoise.bourdon@bnf.fr



This is an English translation of “A la recherché du temps perdu, retour vers le futur: CBU 2.0” copyright © 2014 by Mélanie Roche. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

ABSTRACT:

The outcome of a conversation between a pioneer of Universal Bibliographic Control (UBC) and a digital native head librarian director of France's national bibliographic agency, this paper celebrates UBC's 40th anniversary in its own way. We wished to read UBC's founding statements once again, and share our questions and thoughts as to what its future might be. UBC used to rest upon a few simple yet ambitious ideas: knowledge, the basis of a nation's wealth and heritage, is expressed in publications, originally printed onto a carrier; every country ought to invest in a national bibliographic agency in charge of collecting and describing any resource of national interest; this description must follow international bibliographic standards, the primary function of which is to facilitate the exchange of standardised records; this exchange works as a worldwide division of labour and enables universal access to a broad set of encyclopaedic resources via bibliographies and catalogues.

UBC's founding principle stood in-between those of the Alexandrina and Google. It adhered both to the Ancient dream of mastering all the knowledge in the world and to the modern-day building up of global, routed access to information. In this vision predating the dissemination of the Internet, the

very idea that a patron should have access to the sum of current knowledge on their own was inconceivable: librarians were still considered key figures in the diffusion of knowledge. And the worldwide library network envisioned by UBC somehow heralded an "Internet of libraries". Is there anything left of this legacy? Has UBC been overtaken? In our opinion, the very story of UBC's history carries the seeds of an unbeknownst overhaul.

Keywords: Universal Bibliographic Control, national bibliographic agencies, legal deposit, semantic Web, international cooperation

Introduction: UBC 1.0

When Dorothy Anderson published her 1974 document "*Universal Bibliographic Control: a Long Term Policy, a Plan for Action*"ⁱ, IFLA's program on Universal Bibliographic Control (UBC) had actually been running on for quite a few years. This "long-term policy" alongside with its "plan for action" took root as early as 1961, when the very same Dorothy Anderson laid the foundation stones of the idea at the first International Conference on Cataloguing Principlesⁱⁱ in Paris. Yet what truly brands 1974 is the formalization, in a text not unlike a manifesto, of UBC's seminal principles and of recommendations addressed both to librarians and governments. The text summarises the stakes and principles of international cooperation based on coordinated policies and allocated public resources, and sets forth the priorities to be observed in order to implement such an ambitious program. UBC at its core rests upon a few ideas usually well-known to librarians, that we will sum up in this introduction so as to set the scene and ground its foundation in a time when technologies as well as cultural practices were not the same as today.

In the early 70s, the concept of "publication" was instrumental in the cultural perception presiding over the invention of UBC. The result of a codified and organized editorial process, publications are considered the very essence of the cultural production a nation ought to preserve for present and future generations. In a time when the Internet did not exist, publishing something meant you had been through established circuits of selection and validation for books or serials, standing as guarantors for their quality. Original ideas, acknowledged as such and publicly expressed, were materialized in the form of a printed artefact: a book, newspaper or journal. Even though "special collections" (objects, archives, images...) have been otherwise preserved for a long time, most libraries were still primarily interested in printed objects – the main bulk of what they collected and catalogued. In the beginning, the discussion around bibliographic exchange between countries thus focused on printed publications.

UBC advocates that every country should invest in an organisation dedicated to collecting and managing publications – a "national bibliographic agency", usually hosted in the National Library, in charge of systematically collecting every publication either issued on the territory or presenting a national interest (by means of legal deposit or otherwise) and describing them in a "national bibliography". We can see how instrumental the concept of nation-state was in the creation of such an organisation: every country, supposedly granted with a certain territorial, cultural and linguistic unity, holds a share in the universal knowledge split between nations. The concept of a national bibliography thus embodies one of the certainties of the time: the sum of all the national bibliographies will eventually lead to the creation of one single universal bibliography describing all the publications ever produced by human invention. UBC's motivation is also economical: if each country describes its national production in an intelligible fashion and so that others may reuse it, all will save

money. The effort agreed upon by each and every one to comply with demanding standards will in turn allow them to benefit from the production of the others. The political, economic, and even diplomatic history of the major national Western libraries is indeed the cultural background against which the invention of the UBC model in the 70s is to be set.

UBC's modernity lies mainly in that it issues stipulations in terms of standardisation and description, namely that cataloguing should respect international standards set by consensus and within IFLA's framework. Standards are everywhere in Dorothy Anderson's text: they take up a whole appendix – "The requirements for UBC: international standards for the bibliographic record". Certainly, the tradition and practice of descriptive standardisation in libraries predate the 70s. But UBC strengthened this professional culture of standardising and exchanging records (either on cards or machine-made), taking it on a broader scale. As libraries grew more and more computerized, they worked together not only to standardize the practices, codes and rules of cataloguing but also to create and maintain computer formats which would code bibliographic information so that a machine could read and exploit it. In that respect, the adventure of MARC formats can be separated neither from the standardisation surge brought by UBC nor from what was not yet called interoperability.

After glorious hours and Cold War-like twists and turns, UBC's governing body was short-lived: its coordinating structures established in 1974 no longer exist. The Office for Universal Bibliographic Control, later known as the Universal Bibliographic Control and International MARC, closed in 2003. The bibliographic coordination and standardisation was then handed over to other IFLA structures, IFLA-CDNL Alliance for Bibliographic Standards (ICABS) and then IFLA-CDNL Alliance for Digital Strategies, whose activities ended in 2011. In 2012, on a joint proposal from the Bibliography and Cataloguing sections, IFLA's Professional Board issued a statementⁱⁱⁱ which, while it reasserted the relevance of UBC's founding principles, did not only fail to back that position up with actual recommendations, but further acknowledged the digital revolution and other legitimate concerns, such as the need to stop considering bibliographic exchanges from a Western perspective. There is therefore no longer a governing authority for the program, which in the life span of international organisations is usually the sign of a first-class burial.

Did the digital tide knock UBC out? Did the Web kill catalogues and bibliographies? Probably, but one shouldn't too hastily believe the same of the invaluable structured data contained in those tools, nor of the skills on which they were built. This paper is no eulogy. By taking a step back and facing the reality of usage outside library-instituted frameworks, it seems obvious that UBC's story still goes on, but under other names, within other frameworks, and with other actors than the sole libraries. Thanks to UBC's dynamics, librarians may have envisioned something – the Web – that eventually developed without them, for lack of thinking far enough, fast enough, and more importantly outside of their community. All they have to do now is to regain their place and advocate for their added value in an environment they can no longer pretend to control. We shall try and understand how libraries, as an international community, have come to that, and by what means they may pick themselves up again. This paper is divided into three parts: the first one traces the evolution of the scope and forms of knowledge concerned by bibliographic description; the second one focuses on the descriptive methods of these resources and their standardisation for interoperability purposes; the third one is an invitation to imagine new forms of international cooperation in the continuation of UBC's ideals.

1- From publications to resources, nations vs. global networks: radical shifts in the forms and scope of knowledge

In our first part we will endeavour to delineate the changes that have radically modified the outline and scope of UBC's "raw material" – the cultural and scientific production national bibliographic agencies are meant to collect and describe. We will focus on the concept of publication as UBC first meant it in 1974, that is in relation with a printed medium, a publishing circuit, and a national culture.

From carrier to resource: releasing content from objects

As seen earlier, in the 70s what was meant by "publications" was essentially books and printed periodicals. On that note, the first two international standards of bibliographic description, dating back to 1971, were ISBD (M-Monographs) and ISBD (S-Serials). It soon became very clear though that printed material is only but one of several carriers of information. The concept of "non book" appeared, and was recognized in 1977 by the publication of ISBD (NBM-Non Book Material), alongside with the concept of "new media": microforms, discs, magnetic tapes, slides, audio-visual documents, and computer files. In 1986, digital documents on carrier were widespread enough that their bibliographic description be codified, leading to the publication of ISBD (CF-Computer Files) in 1990. The history of bibliographic standards issued by UBC therefore follows the chronology of carrier innovation, up to interactive multimedia, optic technologies, digitisation and eventually, the Internet. In 1997 there appeared a radical conceptual shift in the perception of standardised description: in ISBD (ER-Electronic Resources) the word "resources" replaced the word "document".

Bearing in mind that standardisation is always somewhat late on the reality it intends to shape, the semantic evolution of normative documents evidences how swiftly the production of information was upset. Within the course of fifteen years or so, we have passed from "publication on carrier" on to "immaterial resource". In other words, it took less than forty years for the representation of knowledge or information not only to outgrow the printed word and expand to other media, but also to literally hatch out of the object that used to contain it. Nowadays a "resource" is a material or immaterial entity with an intellectual and / or artistic content, including text, music, animated or still images, graphs, maps, sound and image recordings, digital data or programs. Publication forms do not exclude one another but on the contrary coexist and add up.

Resource granularity: infinite shades of grey

The diversification of document types has drastically impacted the way libraries produce, inventory, distribute and preserve information. Publications today range from traditional printed books which are generally speaking objects with a fixed content, to digital multimedia books available off-site in various formats, each of which carrying one or several digital right management systems (DRM). A resource on carrier has to be issued in large numbers in order to meet its audience, whereas an online resource, even though it may be a single file, is remotely available to several simultaneous users. The concept of "print run *n* copies" has disappeared with online publishing. Different methods of content updating coexist today. A new edition remains the "traditional" method for printed resources. A printed edition with an online update is also usual. Loose-leaf publications are rare, but online integrating resources^{iv} (such as websites) are skyrocketing.

The same information can therefore come in many shapes and forms, each of which addressing different needs and uses. For instance, the publisher of an e-book may publish it chapter by chapter, while the homothetic printed book is bound to be published in full. The two forms coexist: a travel guide printed on coated paper may meet its audience, and the “Where to sleep” section may be marketed online so that it can be regularly updated and meet the very same audience – or a different one. On the other hand, entire sets of digital resources – monographs or periodicals – are now regrouped into thematic packages by publishers, and distributed in that form under other conditions than those of unit sales. And one single resource may belong simultaneously or successively to several packages. Lastly, a digital document, whether born digital or digitized, may be re-materialised thanks to the technique of print on demand... The presentation of information can thus be declined into an infinite number of combinations; this is where increasingly complex problems arise in terms of description: which granularity level suits our information best?

National production: the limitations of legal deposit

As said earlier, the division of labour between the national bibliographic agencies rests on the division of the “bibliographic world” between nations. This idea seemed to be taken for granted at the time, so much so that the founding texts of UBC in 1974 failed to provide a definition of “national production”. One can only infer from the stipulations, and somehow by default, that it refers to what is being collected under the provisions of legal deposit arrangements and described in the official national bibliography of each country.

Every country is encouraged to have a national library and to make legal deposit arrangements. Each state may choose what organisation is best fitted to the context, and split responsibilities between various cultural operators. The legislation should also specify what types of documents are being collected, who should deposit and in how many copies. In France for instance, where legal deposit is strictly bound within territory criteria by the Code of heritage, the publishers and printers whose headquarters are located in the country are primarily subjected to deposit. As far as legal deposit is concerned, French law – one of the oldest in the world (1537) – has constantly been adapted to reflect the apparition of new types of resources. Here are a few milestones: photographs in 1925, phonograms in 1938, videograms in 1975, audio-visual documents and digital publications on carrier (software, etc.) in 1992, up until 2006 where legal deposit arrangements were made for the “signs, texts, images, sounds or messages of all kind which are publicly communicated in digital form”^v – namely the Internet.

Nonetheless, the production of a country can decreasingly be said to match the editorial production issued within its territory. Indeed, legal deposit is often completed with foreign acquisitions, some of them actually dealing with works produced by national citizens who live abroad. The issues of importations and the production of national citizens abroad are old news, but they have grown more complex still with digital publishing, taking up proportions that might seriously shake a system based on a geographical principle thought through by and for States. The preservation and indication of what has been published on the Internet naturally come to mind: given the amount of data at stake, it is simply impossible for libraries to collect a comprehensive national production. Some of them have started to archive websites through selection or sampling with harvester robots, and to develop automated identification strategies for national domains amidst the immensity of the Web (by narrowing down the automated collecting to high level domain names, such as .fr for France or .dk for

Denmark). But the very nature of the Web, essentially a transnational media, opposes any idea of collecting a comprehensive, “national” corpus.

International organisations such as the International Consortium for the Preservation of the Internet^{vi} are currently working on a collaborative collection of the Internet (to cover worldwide sport events such as the Olympic Games for instance) and on dividing work according to principles rather close to UBC’s stipulations. Such efforts are still in their infancy though, and the collecting, recording and preserving methods upon which they rest are very different from the processes used heretofore to treat national collections and share data between countries. There is no Internet bibliography – except if we believe major search engines to be the next best thing.

National bibliographies facing the digital age

Another way of defining “national production” would be to consider what is inventoried in the official national bibliography, another cornerstone of UBC. From 1954 onwards, UNESCO has recognised that the inventory of deposited documents and the publication of this inventory are the natural counterpart of the effort made by the depositors (be it the publishers, printers or authors, depending on countries). But here again there is a gap between what is being inventoried under legal deposit and what is recorded in a national bibliography. Authoritative national bibliographies primarily describe commercial resources published in large numbers. More often than not, resources are scarce outside the commercial publishing circuit: they include mostly self-published documents, associative publications, official publications, research publications. Although some libraries collecting e-books and e-journals are considering including those publications in their national bibliography (such is the case at the BnF), this affects but a tithe on online publications for the time being.

There have been several international congresses on national bibliographic services^{vii}, during which the scope and coverage of national bibliographies have been discussed at length. Notwithstanding the particularities of national contexts, it is commonly accepted today that on the one hand national bibliographies need not single out every legal deposit item, and that they may on the other hand inventory foreign material published by citizens abroad or dealing with country-related topics (as is the case for the National Library of Switzerland, who collect every publication issued *on* and *in* Switzerland). Economic pressure has led libraries to set up pragmatic and selective policies, the relevance of which has been acknowledged by IFLA’s international authorities: sampling what material is collected is recommended, as well as splitting the inventory of national production between several institutions.

National bibliographies thus partly define a national production, but they do not comprehensively list it down. Nor can we say any longer that the sum of all national bibliographies equals global production, for resources increasingly tend to elude inventory: for instance, supranational resources such as intergovernmental publications or international associations’ publications, which are seldom catalogued in the headquarters’ country as should be expected. Furthermore, we have to take into account the collaborative resource production issued by Internet users outside the frameworks and prerogatives set out by states and bibliographic agencies. Global actors, digital natives, the Web and its industry’s own children, they have become information producers at least as powerful as great states. No library is mandated to preserve the content of Wikipedia. And although Twitter entrusted the Library of Congress with the totality of their data, these may be of interest for researchers

worldwide. The “grey” production resulting from the activities of such online professional and social networks thus eludes any attempt at inventory when it could evidence the vitality of a nation through the construction of public debate, the formation of opinions and controversies, all topics that have always interested researchers, historians and sociologists, who used to rely on sources such as the written press or broadsheet newspapers – but those tend to be increasingly dematerialised.

Perhaps more worrisome is the twofold trend affecting academic publishing – both of content dematerialisation and commercial concentration. New methods of digital resources management dramatically shatter the condition of their diffusion and preservation in libraries. As long as the documents had a material carrier, a library could ask a publisher for this carrier and organise the cataloguing and access to its content as it pleased them. But e-resources publishers and distributors have increasingly instituted new economic models based on subscription. Libraries pay for accessing off-site resources but, except for special arrangements stemming from large-scale negotiations (in the case of national licenses for instance), publishers do not actually give away the files, which are stored on the publishers’ own servers. Can we still say that such dearly acquired resources belong to the library? Must they appear in the catalogues and bibliographies that traditionally inventory what libraries hold within their walls?

UBC: utopia revisited

After all these social, commercial and technological shifts, CBU’s 1974 framework defining and describing a publication has been seriously shattered. As a background for the conceptual and normative adjustments called for by the transformation of the very nature of UBC’s “matter”, a major economic reality has furthermore imposed itself: mass.

There is no common ground between the bulk of resources in 2014 and that of 1974. Every mutation the publishing market went through has mechanically led to a breathtakingly fast increase in the number of items to describe. An increasing number of actors have publication means increasingly cheaper than they used to be. Social and cultural practices born with the Internet have allowed many a resource heretofore belonging to the private sphere and the spoken word to enter the scope of public heritage to be preserved by public institutions. In e-publishing, every publication may be declined, aggregated or divided up into innumerable parts or manifestations. In the same time, human and technological resources in libraries haven’t expanded thus – far from it. It so results that the technological and normative apparatus of their intervention has to be revised, as must the scope and range of their cooperation: it is time they opened up to other actors and brought the division of the work on a broader scale.

As far as bibliographic description is concerned, we have seen how closely these evolutions were monitored, insofar as possible, by UBC’s communities of standardisers, so much so that we are coming to the end a certain model, overtaken by the speed, diversification and volumetry of resources, but also by shifts in usage. We shall now consider how the initial model of a bibliographic *record* such as defined by the first editions of ISBD has exploded in favour of a new representation of *data*, which in turn induces a new take on standardisation as well as on the general economics of data production and exchange.

2- Adjusting bibliographic standardisation to the information new deal: data sharing vs. record exchange

In this part we will review the conceptual and normative mutations called for both by the information new deal such as was previously defined and by new usage. Naturally we shan't trace back forty years of bibliographic standardisation here. We shall focus on three structuring aspects for the future of international collaboration in the context of UBC as far as description and control are concerned: descriptive standardisation so as to exchange data, information modelling in order to gain more description granularity, and lastly authority control review.

Securing the uniformity of bibliographic information for interchange purposes

In 1974, UBC's assumption was to spread out locally designed records on an international scale. Standardising then meant securing the *uniformity of the form and structure of information* so that it could be exchanged internationally. By "exchanged", we mean "duplicate and distribute". Skipping through the duplication of cataloguing, only to duplicate and distribute records, the layout and content of which have to be standardised in order for them to be exploited outside their local environment.

The uniformity of access points to bibliographic descriptions was one of the main objectives of the Paris Principles established in 1961 during the first International Conference on cataloguing principles. These Principles have thereafter given rise to a great variety of international recommendations on the form and structure of names, by author type (corporate bodies^{viii}, persons^{ix}) or by title type (anonymous works^x, liturgical works, etc.). Spread out as headings in the bibliographic records published in national bibliographies, such standardised forms were meant to be used identically worldwide.

Up until the 80s, headings and the bibliographic description to which they allowed access to were closely knit together. But the gradual computerization of catalogues split that couple up, leading way to a new kind of record: the *authority record*. Initially meant as a cataloguing aid, an authority record gathers together all the information needed to identify the entity in question. The record derives its name from the fact that it is authoritative in regard to the catalogue it manages; in other words, it defines the authoritative form to be systematically applied as an entry to bibliographic descriptions within the said catalogue. Meanwhile though, the authority record has increasingly broken free from its original catalogue, to become an authority in itself. Its value is acknowledged in that it carries information verified by experts, which can be re-used in very different contexts, far outside the gated community of libraries. According to UBC's 1974 principles, authority records are supposed to be exchanged internationally, that is duplicated and exported towards other systems where they can be locally updated, for instance with additional, local forms to the name of an entity, the better to stick to patrons' expectations.

The concept of a standardised international bibliographic description (ISBD) was born in 1969 during the International Meeting of Cataloguing Experts in Copenhagen. The first ISBD for books and serials was published, as seen earlier, in 1971. It listed down the required elements to identify a publication, prescribed the order in which they should appear and the punctuation that should bind them, and established which data transcription rules and information sources should be referred to.

In parallel to these stipulations regarding the form and structure of bibliographic description and headings, data encoding formats were being developed, based on the punctuation prescribed by ISBD: MARC formats (Machine Readable Cataloguing) created and used by and for libraries. MARC bibliographic formats include fixed-length fields where coded information on the processing of the record is stored, allowing for instance to sort out or select records by publishing country, language and script of the described item, creation date of the record, cataloguing language, etc. ISBD thus enables the production and exploitation of records that can be read both by humans and machines. Manually designed records and automated cataloguing can coexist in a single catalogue.

In order to exchange MARC records libraries use a protocol defined in ISO 2709, the first edition of which dates back to 1973 and was entitled *Format for bibliographic information interchange on magnetic tape*. A sign that the times have changed, the latest 2008 edition is entitled *Format for information exchange*. This is an exchange format shaping every MARC format (for every country has developed its own MARC: we call it the “MARC Babel”). It allows the interoperability between the various MARC formats by setting out conversions (unfortunately few conversions are being maintained internationally, not even between MARC 21 and UNIMARC), thus allowing *MARC record exchanges between libraries*.

By aligning themselves with international standards, less technically developed countries were able to actually take part in the international effort and benefit from it in return. But from 1990 onwards libraries have faced the changes mentioned earlier. The technology of international communications developed (with the Internet and the Web technologies), and there arose a need for closer work with information producers, other heritage institutions, copyright holders – all communities unfamiliar with the sophistication of library standards and their almost “proprietary” formats. There began an assessment of standards and practices, for a tension had grown between the maturity of international standards used on a daily basis and the growing need for a description of digital resources (whether born digital or digitised). The item (one single document) had ceased to be the sole object of description: the resource had to be described independently from its carrier. But digital objects must also be described within the context of their full life cycle: descriptive metadata were no longer enough and had to be linked with technical and administrative metadata relevant to preservation and digital rights management. A normative revolution set out to tackle such novel needs.

Modelling bibliographic information to manage granularity

Back in 1974, it was recommended that an authoritative bibliographic description should be made “with the book in hand”, as the expression went. The aim was to copy and later type down or code information already *in* the publication, according to formatting standards and structure. This bibliographic description formed the main load of the cataloguers’ work at the time. Choosing and setting up entries, which supposes that the content of the document has been analysed through controlled frames of reference, was still a secondary step in their work.

When digital resources came up, not only was it impossible to have the resource “in hand”, but the very content of a descriptive record turned into something of a puzzle, so difficult had it become to define the object or item to be described. How to provide easy

access to the actual content of the resource? Which granularity level would best meet patrons' needs? The chapter of a book? The article inside a periodical booklet? A package of monographs? The rematerialized version of a digitised item? Every single digital version of an e-book as defined by its format (PDF, ePUB, etc.), or even by the DRM system attached to it? Where does the bibliographic description give way to the description of the copy? Which standard applies to a printed item that has been digitised: the one for printed monographs or the one for e-resources? All these reasons amounted to the much-needed differentiation between information *content* and *medium*, which in turn motivated the latest evolution of ISBD (creation of area 0).

Elaine Svenonius^{xi} wrote in 2000 that “bibliographic description is an art and therefore will always be approximate. There will always be instances that are undecidable or where decisions must be made arbitrarily”. This is precisely what happened in 1990 during the IFLA Seminar on bibliographic records where it was decided to remodel bibliographic information so that patrons should be the focus of attention and their expectations better met thanks to a better use of technical innovations. These discussions arise in a bibliographic context where the universe no longer revolves around libraries, where the growth of online search engines is considerably challenging catalogues and where librarians realise they need to focus on patrons' new needs: on the Web, little do they care about the library itself – they want to interact with other resources and other communities.

Modelling bibliographic information will serve both assessment and prospection purposes: does the content of bibliographic records such as it used to be defined meet patrons' actual needs – and how can it evolve? This is the core objective of the now famous FRBR model (*Functional Requirements for Bibliographic Records*) published in 1998^{xii}. Each element in the ISBD record is brought face to face with the patrons' activities: find, identify, select, obtain. The new international cataloguing rules published in 2009^{xiii} are organised around those four activities, and describe another fifth: navigate, later on known as “explore” in the revised version of FRBR. Thanks to FRBR's analysis, the descriptive record ceases to be perceived as a compact information block to resemble rather a network of data elements linked to one another by many relationships. A difference is made between data related to the work, the expression, the manifestation and the item on the one hand, and data related to entities such as persons, corporate bodies, titles, subjects, and so forth, which constitute particularly useful concepts (people, things, places...) to allow a patron to find but also discover resources in a more intuitive fashion.

As far as authority records are concerned, the second edition of *Guidelines for Authority Records and Entries*^{xiv} in 2001 introduces in a normative international text the concept of *sharing* authority records – instead of *exchanging* them. This is a grounding step toward the eventual abandon of internationally standardised name forms, an abandon confirmed by the creation of VIAF (Virtual International Authority File) in 2003, for people. VIAF is supplied by the authority records of members – about forty as of today. For a single person, VIAF gathers together the various authoritative name forms in the files of the various countries, regardless of their language or script, and without a pivot form. Every national form of the name is thus placed on an equal footing. From their confrontation the person's identification shall grow stronger on an international level. VIAF can also deal, although with less trustworthiness for the time being, with corporate, geographical, title and expression names. Humans and machines alike may come and identify data elements they need locally in VIAF and, with a little help from technology, choose to duplicate the information in their local applications, or else to link a local application to the right subset in VIAF. VIAF's

pragmatism inspired the authors of FRAD (Functional Requirements for Authority Data)^{xv} and FRSAD (Functional Requirements for Subject Authority Data)^{xvi}, respectively published in 2009 and 2010. Notice how the word “data” has replaced “records” in their respective titles: this change is anything but fortuitous – it is the sign of a new step towards the acknowledgement of the granularity of the information produced by libraries and used by patrons.

Authority control and data sharing

What is authority control? Clack^{xvii} wrote in 1990: “It is a technical process executed on a library catalog to provide structure. Uniqueness, standardisation, and linkages are the foundation of authority control”. According to this definition, authority control is exercised both on definitive bibliographic records describing documents and on authority records shaping the authoritative forms of names for authors and work titles in a given catalogue. At the beginning, the main goals of authority control were to avoid duplicates, standardise the content, shape and structure of data constitutive of records and link these records or data to one another, within a single catalogue, so as to guarantee its consistency and user-friendliness. Originally, the concept of authority control is therefore inseparable from that of catalogue. Authority control is the reason why the author entry of a catalogue gathers together all of this author’s work, and his own work only, and also why every item is described in one single record. This cleaning up, shaping up and matching up of huge datasets is really not unlike quality control in an industrial context.

UBC entrusted the national bibliographic agencies with authority control (stipulations 18.1.1 and 18.1.3): this had actually been a librarian’s work for centuries – for indeed it requires a great deal of human expertise. However, as computerisation and information standardisation went on, technical devices took over decision-making by automating some controls – culminating nowadays in the use of matching algorithms. More importantly, the use and technical background of catalogues have changed a lot in the course of the last fifty years, and authority control gradually shifted away from the concept of catalogue to acquire a new range and scope – becoming truly a strategic tool.

Today, the bibliographic information unit is not so much a *record* as a *data element*. The apparition of XML was instrumental in this shift: a standard understood and used by many other communities than libraries, it singles out data elements much more easily than MARC fields and subfields used to, and opens data up. Authority control is likewise shifting from record to data element. Beside traditional authority control carried out by cataloguers (controlling duplicates and ensuring compliance between data forms and structures), a new *level of control* has thus arisen, based on the consistency of document analysis strictly speaking. Thesauri stemming from various communities are being linked up by valid and reliable matches between similar concepts expressed differently. The point is no longer to standardise data so as to exchange them, but to *link similar concepts so as to share them*. The concept of “title” for instance, is not the same as seen from a librarian, an archivist or a publisher’s standpoint: online linking of these three data elements will allow users to navigate smoothly from the local catalogue where they typed in their “title” search, to an archival inventory or a publisher’s website. Authority control, breaking out from the catalogue to enter the whole Web, has become instrumental in the cataloguer’s work; it is the core of our profession’s added value, the only chance for libraries’ contributions to stand out in a wider context of data production and sharing.

3- A roadmap for CBU 3.0: will libraries break from their silos onto the semantic Web?

What horizon can be envisioned today as far as the organisation of the cooperation between libraries is concerned? Judging from the aforementioned evolutions, which may be the strategic focus presiding over UBC's overhaul? Which places, which governing bodies are to decide upon the bibliographic future of libraries?

Let us keep in mind first that UBC was born from close cooperation between three international organisations: the UNESCO, because governments planned and paid for national information systems; the ISO/TC46, who gathered together the national standardisation bodies in the field of information and documentation; and IFLA, the international federation of library professionals. Coordination brought along by international organisations was deemed essential then, lest national systems should be developed in isolation, but also to guarantee balance between parts of the world unequally developed. In a very different geopolitical context (the end of the Cold War), in which state-controlled information was still a widespread reality, and as instrumental as professional associations may have appeared, states were nonetheless given a leading role. The whole plan indeed was based on ambitious, interventionist public policies, in a context where major, private actors of information globalisation such as we have come to know them did not exist – or took very different forms.

From the start, libraries – and primarily national libraries – were placed at the core of this essentially public cooperation program, arguably because they had traditionally organised knowledge and its access in a way that appeared authoritative to public power. Making the most of the powerful legal leverage provided by legal deposit and national bibliographies, they voluntarily and even militantly established themselves as trustworthy operators who could implement UBC's programme. In 1974, the setting of UBC's Office as a standing organ of IFLA evidenced how singlehandedly the community of librarians was entrusted with international cooperation. Archive services and museums were excluded from the programme, maybe because their heritage consisted mainly in unique items calling for other collecting and managing methods, for which the interest of mutual work was less obvious. Publishers, as mentioned earlier, could legitimately have taken part in the plan, were it only because they were subjected to legal deposit and already provided data elements useful to pre-publishing cataloguing. However, the International Federation of Publishers – although created in 1896 – did not appear as a party involved in UBC's implementation, and publishers have remained outsiders to a standardisation process closely monitored by librarians, as a rule reluctant to take into account the publishers' specific needs in term of cataloguing, especially commercial considerations.

UBC's original plan was therefore essentially a public programme supported by libraries, excluding all other public or private organisations from its governance and network. Such an organisation can no longer exist today. Valuable as their assets may be in other fields, libraries no longer hold the monopoly of metadata structuration, control and exchange. Other operators, mainly private or associative, have come to overshadow both their profession and the states. And as we have seen, information production now includes major trans- or multinational actors utterly out of reach for state control. Internet Archive^{xviii}'s motto "Universal Access to All Knowledge" might as well be that of many an organisation today, both private and public. In the increasingly competitive world of information retrieval and production, libraries must therefore focus on what they do best, and look for ways of interacting with other professional communities that will highlight their specific contribution.

What are the priorities and modalities of their intervention and cooperation today? We will suggest three guidelines for thought and action that echo some topical discussions.

Direction One: Keeping up library interoperability

Building up on the previous episodes where, in good times and bad, they consensually succeeded in putting together a remarkable structure of standards and formats either common or that can be converted into one another, libraries must now be careful to keep up their own interoperability. Only then – as a group and in the long run – can they preserve a critical amount of data standing its own on the web. Echoing the great bibliographic debates currently agitating the profession, we shall focus on a few points worthy of attention and mutual listening^{xix}.

First of all, all major libraries are clearly engaged in a long-term bibliographic transition, usually reckoned to play out over the next ten years, and with a threefold impact: the effective implementation of the FRBR model, the shift in cataloguing code, and the replacement of MARC formats. The take on this transition varies from country and context. Following the example of the Library of Congress, several major libraries historically accustomed to the AACR2 cataloguing rules and MARC 21 format have set out to shift their production process to RDA cataloguing. Others, such as the National Library of France (with the project data.bnf.fr) or the National Library of Spain (with the project datos.bne.es) have chosen for the time being to prioritize turning and exposing data into RDF over actually shifting from one production process to another. Some libraries (such as the Deutsche Nationalbibliothek) work on both levels (production and diffusion), while the less fortunate remain cautiously expectant.

It is now widely agreed that one single production or diffusion format for all kinds of data and all types of usage is a delusive idea that may well prove counter-productive and even dangerous: managing bibliographic information in very different contexts and on various levels of granularity and completion will still be required, depending on whether we consider local management, professional interchanges within a production network, or the diffusion to patrons with more or less generic or specific expectations. In such a context, a plurality of approaches is possible, if not mandatory. Yet converging methods should be kept up as far as *exchange formats* are concerned, the essential pivotal points of future international and inter-professional indexing, exchanging and linking of information. From this point of view, the proposals issued by initiatives such as BIBFRAME or schema.org ought to be closely followed up and discussed on an international scale. Defining new exchange formats constitutes in our opinion a priority worthy of discussion and coordination, on a par with what had been implemented by UBC in terms of standardisation.

In parallel with this work, major libraries and national bibliographic agencies must go on supplying major international bibliographic databases of a collaborative nature with their high-quality production. Worldcat^{xx}, VIAF^{xxi}, older repositories such as ISSN and younger ones such as ISNI^{xxii} immediately come to mind. In this field, attention should rather focus on the forms of governance ruling the treatment and exchange of bibliographic information between local data suppliers and users on the one hand, and global data aggregators and distributors on the other. Among the crucial issues calling for international discussion, let us mention those of *data rights* or status (open data and the legal regime of licenses), the *traceability* of their origin (a matter of quotation but also of trust), and finally their *quality*.

Arguably, some of the new boards which took over UBC's coordinating bodies have moved to the steering and coordinating instances of those initiatives (VIAF Program or the international agency ISNI-IA for instance). If they are ever to remain in charge of their own data in full transparency, libraries have a strategic hand to play in advocating their positions when it comes to setting up the rules for these new forms of data aggregation, transformation and diffusion on the global scene, alongside with private or associative partners such as OCLC, without which such large-scale projects would never have come about.

Direction two: Taking on the semantic Web standards

Nevertheless, libraries should be wary not to lock themselves up inside cooperative schemes and communities that have long held their organisation together, nor to perform modern-day re-enactments of their historical and cultural disputes. The rhythm of international bibliographic standardisation can no longer keep pace with how fast technology and usage are changing, and more importantly the scope of actors either producing or aggregating data and prescribing standards has considerably broadened and diversified. Libraries would be well-advised to keep up or enter into dialogue with the most influential communities in the Web of data – smoothing out their own disputes in the meantime.

On the semantic Web, libraries open up to unprecedented perspectives of visibility and data reconciliation. While the segregation of professional communities has long forbidden it, it is nowadays possible to reconcile library data with those of other cultural operators, but with statistical, commercial, geographic, linguistic and scientific data as well: the field of possibilities seems unlimited, supplying all data producers, distributors and users in widely open innovation leverage and potentially cost-effective solutions. The technical and legal conditions to enter and be visible on the semantic Web are well known. Yet we would like to stress how some of them have become truly strategic prerequisites for the presence and the cooperation of libraries on the Web. Expressing and publishing data in RDF is a minimal requirement. Legally opening them up so that they may be used again is a real plus. Yet as far as libraries are concerned, it is our opinion that attention should now focus on two aspects: *identifiers* (URI) and *authoritative referrals* (or vocabularies).

In the architecture and “grammar” of the semantic Web, data sharing is built upon the existence and maintenance of unique identifiers: standardised, perennial and actionable, these identifiers are bridges without which nothing can be found, linked or built on. Standardising and routing trusted identifiers can thus be seen as a major challenge for libraries and their partners. From now on, as a group they must therefore consider how to allocate identifiers to their resources from a practical and economic standpoint, and they also have to make the right choices if they are to retain some autonomy in the identification of their collections while gaining in visibility and interoperability. From that point of view the rise of ISNI for public identities or the ongoing discussions around ISTC for textual works are issues that call for international watchfulness and consultation. Admittedly, issues of identification bring along scientific and normative stakes (which identifier for what kind of resource?), but economic stakes as well (at what price can we get and route standardised identifiers?).

Meanwhile, as seen earlier, the distinctive added value of libraries has gradually shifted from *bibliographic data* (easily retrievable at the source from the producers) to *authority data* (our war treasure indeed, piled up from a truly document-oriented analysis of our collections, and presiding over information reconciliation in the FRBR era). On the semantic Web, and as

is evidenced by the international success of the VIAF project, vocabularies and specialised referrals issued by the world of libraries are the most sought for and re-used by other communities. At the BnF, we have noticed that the most popular dataset delivered via our service data.bnf.fr and published in SKOS is RAMEAU, an authoritative repository for subject headings which incidentally comes in quite handy for a multilingual search because of its alignment on its Anglo-Saxon and German counterparts – namely LCSH and SWD. International collaboration between libraries in this domain should therefore be a priority field of consultation and investment, which raises once again issues of governance and cost-effectiveness. Indeed, for lack of means, many institutions tend to decrease their production and control of authority, increasingly relying on collaborative repertoires such as VIAF. If every library follows the same trend, the very value of such repertoires will be lost and the trust in libraries will not be as strong.

It is therefore no coincidence that initiatives linking the issue of identifiers and that of authorities should stand among the most debated in the profession today: indeed, they are truly strategic for the collective presence of libraries on the Web of tomorrow. We believe that building *collaborative authority registries linked to standardised identifiers* is one of the fundamental cornerstones of the new universal bibliographic control. The status, inner workings, economic model and governance of such registries call for international attention.

To conclude on that note, allow us to brush on one last strategic issue that we think would merit further examination: *how can libraries take part in the governing and standardising bodies of the Internet?* From that point of view, it is of vital importance to think about the place of libraries within the W3C Consortium. Anecdotally, the word “metadata”, now so firmly grounded in librarian jargon, was introduced to our community by key actors of Internet standardisation in the mid-90s. The word was indeed forged in 1994 (twenty years after the creation of UBC) by Tim Berners-Lee during a conference in Geneva that led to the foundation of the W3C.

A few years ago, some professionals succeeded in bringing about the reflexion and consultation on the place of libraries on the Web of linked data within the very W3C – a welcome initiative indeed. The subsequent report^{xxiii} became a major reference for libraries within a few months’ time although it wasn’t issued by the usual bibliographic standardisation bodies. Does that mean that tomorrow library standards should find themselves standing by those of the Web, which is the main destination for the data we produce? We have to consider this possibility for we certainly should seek to further integrate the Web’s standardising bodies instead of close-circuiting standardisation in our midst. Reloading UBC will doubtless mean thinking in terms of the governance of the semantic Web.

Direction Three: Joining new networks to build and link data

Among the sectorial lines of collaboration that we think may be of particular interest for libraries as far as cost-effective production and the quality of service are concerned, let us suggest a last few points libraries would be wise to consider and coordinate not only on a national but on an international scale as well.

First of all, we believe that other public operators producing cultural data (chiefly archive services and museums) are the natural partners of libraries. We explained in our first

part how the digital context has brought the trades and management processes of libraries and archive services closer together. Authority standardisation provides once again an interesting instance of collaboration and interoperability. Archivists are diligently working on standardising authorities for names of persons, families and corporate bodies (XML schema EAC-CPF)^{xxiv}, an initiative that might well coincide with libraries' efforts in turning their own authority records into pivotal axes in the organising and sharing of information. National and even international strategies might thus be envisioned to share or co-construct authority files, especially to identify corporate bodies. The resources of libraries, archive services and museums often complement each other from the patron's standpoint but their description follows logics and managing requirements set far apart from the start – wouldn't it be smart if one single page provided online access to all the relevant references stored by all these institutions for one single creator or work? The semantic Web could allow (using primarily authorities and identifiers) to gather together resources described in very different production systems, languages and standards. There is a real potential for cooperation in this sector, and an opportunity to increase the sharing of cultural data, thus enriching the offer of services to patrons. As for the public powers which fund and manage these various cultural or scientific institutions, this would be a fine display of efficiency and patron-oriented service on their part.

Secondly, we believe that publishers, book trade stakeholders (rights management societies, distributors, aggregators...) and information producers in the broadest sense of the word should be far more integrated into the production and distribution of information than they have been thus far. Some libraries and national bibliographic agencies have already widely implemented collaboration schemes with publishers by overhauling their organisation in such a way that would favour pre-publishing cataloguing. This has not yet happened in France, where such cooperation is still at an early stage, through the dematerialisation of legal deposit statements and the implementation of legal deposit for e-books (which remains at an experimental stage to this day). The digitisation of collections, either from public domain (with the growth of digital libraries such as Gallica at the BnF) or under copyright (with the growth of the commercial offer, especially ePubs) has created a new context of cooperation between libraries and publishers, especially regarding copyright identification and management.

For economic reasons, libraries need to get more metadata at the source. For visibility and valorisation reasons, publishers need to enhance their metadata for their online offer. In other words, libraries need to get bibliographic data while publishers need authority data. Some novel projects (the digitisation of out of commerce French XXth century works, or the ARROW project in Europe) necessarily imply to cross-reference databases issued by both communities and which include complementary information: structured reference documents abounding in authority data on the libraries' part, and commercial and availability information on the publishers' part, or information related to the authors and their work preserved by rights management societies. Finally, the development of e-books and the ONIX format has created a new cooperative context in which the descriptive logics and reference documents of libraries and publishers are all the wiser by getting closer. In France, not so long ago, a public report^{xxv} advocated for the creation of *open metadata registries* that would jointly put together and route metadata issued by the worlds of publishing and heritage institutions in charge of legal deposit. We have to envision new metadata flows between private and public sectors, so that mutual benefits may arise.

Lastly, as some libraries have already set out to consider, why not turn patrons into the co-producers of our metadata? Major online bookstores did not wait for the libraries before they enhanced their catalogues with users' reviews and feedback. Generally speaking, Internet users have grown accustomed to being participating actors through their keyboards (sharing, liking but also commenting). From now on they expect to find other users' opinions in addition to authoritative descriptive resources issued by commercial firms and public institutions. The Web is essentially a business of massive co-production, the economics of which are increasingly based on *crowdsourcing*. The question of course should be considered with caution as far as the quality of metadata is concerned, for libraries wish to carry on being branded as trustworthy and only expose data the origin of which is certified. Yet nothing stops us from linking authoritative information issued by a library with information directly produced by users, as long as the origin of the data can be clearly traced back and identified by a user, so that they may be able to assess the status of the information by themselves. Thus, for data.bnf.fr and as early as 2011, the BnF readily created multiple links between the data in its catalogues and those of the online encyclopaedia Wikipedia, judging that users would be smart enough to make a difference between those two kinds of information but would nonetheless appreciate the library's linking of two complementary data repositories. Such takes contribute to putting libraries' data at the very core of the Web's ecosystem.

Conclusion

In 1974, UBC duly acknowledged that one single library cannot describe all the knowledge in all the world all by itself, and that it needs to cooperate with other libraries (the mythical model of the Alexandrina had long faded out). In 2014, libraries need to recognise likewise that they cannot describe all knowledge either, joining forces as they may. They absolutely need to cooperate with the rest of the world. UBC's original assumption – bringing what has ever been produced worldwide to each and every one by dividing the work up between nations – is still valid today. Such an assumption meets with ideals of cultural democratisation and equal access to resources, as well as with the reality of increasing economic pressure befalling state agencies. However, this ambition has widely expanded with the digital revolution and the opening up of unprecedented possibilities in terms of technology and usage. With the advent of the Internet, one might say that what used to sound like a librarian's dream forty years ago has now become a dream for all.

Given the current state of technologies and standards, the semantic Web has firmly imposed itself as the best possible environment to think over the cooperation and opening up of libraries. The possibilities of sharing and co-producing data have indeed multiplied in the wake of this new environment. The semantic Web allows organisations to work with very different production systems and formats while bringing their online resources together shaped as links, when complementarity appears to benefit all parties involved. This new paradigm paves up the way for cross-fertilisation between many areas of activity hitherto stacked up in silos. For libraries, this new opportunity is in our opinion the nodal point from which UBC's ideals may be invented once again.

Acknowledgments

The authors would like to thank **Mélanie Roche** for the translation of this paper into English, and **Gordon Dunsire** and **Mirna Miller**, whose book referenced here^{xxvi} was truly inspirational.

References

- ⁱ Anderson, Dorothy. *Universal Bibliographic Control: a Long Term Policy, a Plan for Action*. Pullach bei München : Verlag Dokumentation, 1974
- ⁱⁱ The « Paris Principles » taken on by the International Conference on Cataloguing Principles, Paris, October 1961
<http://www.nl.go.kr/icc/paper/20.pdf>
- ⁱⁱⁱ IFLA Professional Statement on Bibliographic Universal Control, December 2012
<http://www.ifla.org/files/assets/bibliography/Documents/ifla-professional-statement-on-ubc-en.pdf>
- ^{iv} Integrating resource: “a resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole. Integrating resources can be finite or ongoing.” (cf. ISBD Glossary, Consolidated edition, 2011, p. 329).
- ^v « The state of e-legal deposit in France : looking back at five years of putting new legislation into practice and envisioning the future », Stirling P., Illien G., Sanz P., Sepetjan S. In: *IFLA journal*, 2012, vol.38, n° 1.
http://www.ifla.org/files/assets/hq/publications/ifla-journal/ifla-journal-38-1_2012.pdf
- ^{vi} International Consortium for the Preservation of the Internet (IIPC)
<http://www.netpreserve.org>
For a presentation (in French) of IIPC’s history and activities, see also:
Illien, Gildas. Une histoire politique de l’archivage du web. *Bulletin des bibliothèques de France*, n° 2, 2011.
<http://bbf.enssib.fr/consulter/bbf-2011-02-0060-012>
- ^{vii} Proceedings of the National Bibliographies Seminar, Brighton, 18 August 1987, held under the auspices of the IFLA Division of Bibliographic Control / ed. by Winston D. Roberts. -London : IFLA UBCIM Programme, 1988.
International Congress on National Bibliographies, Copenhagen, 25-27 November 1998
<http://www.ifla.org/publications/international-congress-on-national-bibliographies-icnbs-recommendations>
- ^{viii} Form and Structure of Corporate Headings. London : IFLA, 1980
- ^{ix} Names of persons: National Usages for Entry in Catalogues, compiled by A. H. Chaplin, Dorothy Anderson. Sevenoaks (Kent), International Federation of Library Associations, 1967
- ^x International List of Uniform Headings for Anonymous Classics, compiled by R. Pierrot, 1964 (2nd ed. 1978)
- ^{xi} Svenonius, Elaine. *The intellectual foundation of information organization*. Cambridge, Mass.: The MIT Press, 2000 (p.51)
- ^{xii} <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
- ^{xiii} http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf
- ^{xiv} Guidelines for Authority Records and Entries. London : IFLA, 1984 ((2nd ed. 2001)
- ^{xv} <http://www.ifla.org/publications/functional-requirements-for-authority-data>
- ^{xvi} <http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf>
- ^{xvii} Clack, Doris Hargrett. *Authority Control: Principles, Applications, and Instructions*. Chicago & London: American Library Association, 1990

^{xviii} <http://www.archive.org>

^{xix} For a more detailed presentation of these ideas, see:
Illien, Gildas. Interview with Gildas Illien, Director, Bibliographic and Digital Information Department, Bibliothèque nationale de France (BnF). *Information Standards Quarterly*, Winter 2013, 25(4): 22-29.
http://www.niso.org/apps/group_public/download.php/11938/IP_Illien_BnF_isqv25no4.pdf

^{xx} WorldCat
<http://www.worldcat.org/>

^{xxi} Virtual International Authority File (VIAF)
<http://viaf.org/>

^{xxii} International Standard Name Identifier (ISNI) International Authority
<http://www.isni.org/>

^{xxiii} W3C. Final report of the incubation group Libraries and the Web of data. October, 25 2011
<http://www.w3.org/2005/Incubator/1ld/XGR-1ld>

^{xxiv} Encoded Archival Context for Corporate Bodies, Persons, and Families - EAC-CPF XML-Schema
<http://eac.staatsbibliothek-berlin.de/>

^{xxv} On the creation of open metadata registries, see proposal 79 of the Lescure report: Lescure, Pierre. Mission Acte 2 de l'Exception culturelle. Contribution aux politiques culturelles à l'ère numérique. May 2013.
http://www.culturecommunication.gouv.fr/var/culture/storage/culture_mag/rapport_lescurer/index.htm

^{xxvi} "Bibliographic Information Organization: a view from now into the past". In Miller, Mirna & Dunsire, Gordon. *Bibliographic Information Organization in the Semantic Web*. Oxford : Chandos Publishing, 2013 (Chandos Information Professional Series), pp. 1-85