



Mg. Szabolcs Dancs
Budapest – NSL

WHO WILL BUILD THE LIBRARY CATALOGUES? MARC AND ONIX: AN OLD CONCEPT AND A NEW PERSPECTIVE

Abstract

Standard data exchange formats have been created to enhance data flow among stakeholders of the book value chain. MARC already is well-known in the library area, other metadata schemes such as ONIX – which is more ambitious in the sense of containing data fields on other information than bibliographical, e.g. copyright status and commercial availability of a work – are much less utilized at the moment, however use of ONIX is recommended and encouraged by ISBN User Manual and the standard is mentioned directly by ISO 2108:2005, also, information about the format is distributed by a several of EU funded projects (e.g. Linked Heritage, ARROW). The article outlines how ONIX might transform traditional library data flows in a way which makes them more efficient than they are now.

Introduction

Optimizing data flows, making information management cost and time effective by exchanging data is far from being a new idea. As it is applied to the products of the book industry, it's mainly libraries that have experience in applying standard metadata for the description of publications. Bibliographic records created by librarians are typically being used for commercial purposes. An example for that is Amazon, which uses data provided by OCLC's WorldCat.

The more actors in the value chain of the book are involved in the process of data exchange, the more information should be added to the book descriptions. Specific data fields need to be defined according to the contents of the interoperating databases. Publishers, retailers, reprographic organizations, etc. use some core data on publications but have their own needs and demands as well. Publishers extract copyright and other relevant information from their contracts signed with authors and organise it into databases. Retailers register – among others – commercial data. Collective rights management societies need to store information on licences in a way that it could be easily retrieved.

While old standards and data exchange formats proven to be sufficient to libraries' needs are being enhanced and adjusted by different establishments responsible for easing data flow in the book value chain, the transforming role of librarians in modern technological environment has become a subject of high interest in professional discussions. Will the job of the librarians remain the same? Will it be renewed only in a technological sense, by implementing modern tools and acquiring new skills that enable to use modern technology in the most efficient way possible? Or will innovation cover the whole area of library work, forcing librarians to reconsider their role in the digitised era? What's more, are they threatened to be replaced by other experts or, being optimistic, to find their place in another part of the book industry?

The First Revolution of Data Exchange

MARC (Machine-Readable Cataloguing) format arose in the sixties of the last century. Then it seemed that a new era would come for the libraries which could save much time and significantly reduce their costs by implementing a method that enables the re-use of bibliographic descriptions already prepared. The idea was quite simple and very fascinating at the same time:

- one single bibliographical description created once – ideally – by the national bibliography provider (most commonly a national library),
- multiplied utilization of bibliographic records (by dissemination among libraries through protocols such as Z39.50).

Oddly enough, since the concept of exchangeable bibliographical record formats was born, many national standard formats have appeared (INTERMARC, IBERMARC, HUNMARC, etc.). Thus libraries have drifted away from the principal aim which was something like constructing „One Common Catalogue” of bibliographic records all around the world. Albeit WorldCat developed nowadays by OCLC seems to accomplish this concept by building an international level library catalogue. So the idea is still alive, and tendencies, such as the spread of MARC 21 and RDA, show that libraries do their best to get closer to the realization of the original goal.

The Second Revolution of Data Exchange

It should be noticed however, that the aforesaid conception of „create once – use many times” principally concerns libraries. Other possible utilizations of these bibliographical data, such as for commercial purposes, haven't been taken into consideration. By identifying this problem ONIX (Online INformation Exchange) standard brought about a new perspective to the library sphere, and what's more to the whole market of publishing. Owing to the structure of an ONIX record, commercial and copyright data can also be stored in addition to bibliographic data.

This standard provides that once created data can be reutilized for different purposes, such as:

1. publishers' databases,
2. ISBN databases
3. library use (as bibliographic record)
4. commercial use (databases of online retailers)
5. Books In Print databases
6. copyright information management tools (ARROW)
7. etc.

For these purposes ONIX has the compatibility with the FRBR (=Functional Requirements for Bibliographic Records) model. The first group of entities in FRBR are the following: work, expression, manifestation, and item. Copyright refers to the creator of the work (author's copyright) and of the expression (translator's copyright). Commercial data should be defined on the manifestation-level, i.e. edition-level.

Also, ONIX complies with MARC 21 which is crucial to enable data exchange between library systems and other databases in the book value chain.

As a result of implementing ONIX, the old data flow concerning published documents will be replaced by a new one.

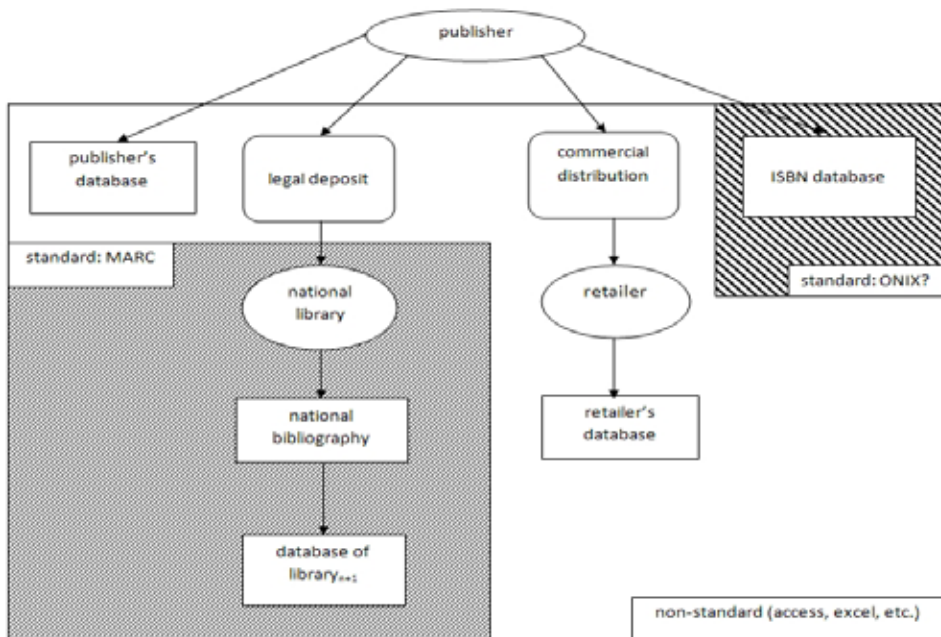


Figure 1 Old fashioned data flow

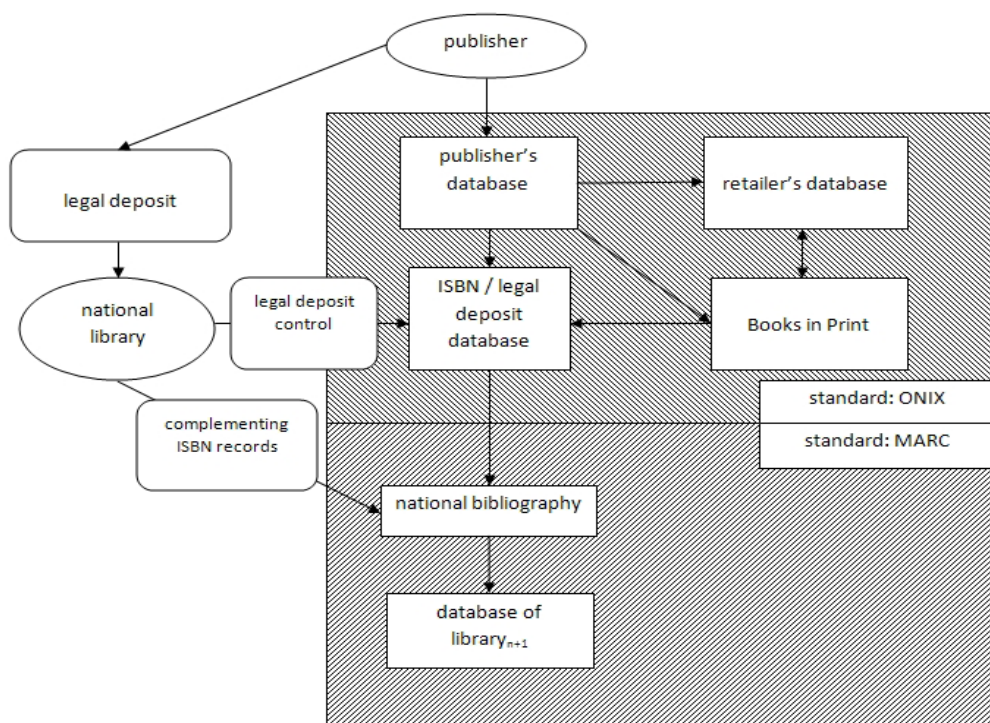


Figure 2 A data flow renewed

As it is well shown by Figure 1, traditional data flow has its primary disadvantage in multiplying manual work with data. No need to emphasize, that typing the same data more than once is much less cost effective than typing it once and reutilizing it several times.

One of the significant differences between the two models is that in the new data flow (represented by Figure 2) core data about publications are created in the private sector, by the producers of the documents, for their own purposes, for their own databases. Some elements of the core records provided by the publisher are identical with those found in a bibliographic record. Fields such as „Author’s name”, „Title”, „Publisher’s Name”, „ISBN” are a few to mention.

Books in Print database is a new element in the data flow outlined. From the viewpoint of a national library, information derived from Books in Print databases helps to improve legal deposit control. Reports about books issued but not deposited by the publishers enable us to identify missing titles in a more efficient way than ever. Also, for current publications, a more complete listing than previously possible can be constructed on the basis of these titles.

Nevertheless, it's important to notice that the second model is only a suggestion, not the single possible method to realize a data flow based on the use of ONIX standard.

Metadata control

Using ONIX might have another advantage besides optimizing dataflow, that is supporting retrieval of information by improving the quality of metadata (see Figure 3). In this model retailers use metadata that is controlled by the national library (or national bibliography agency) on the basis of the documents deposited.

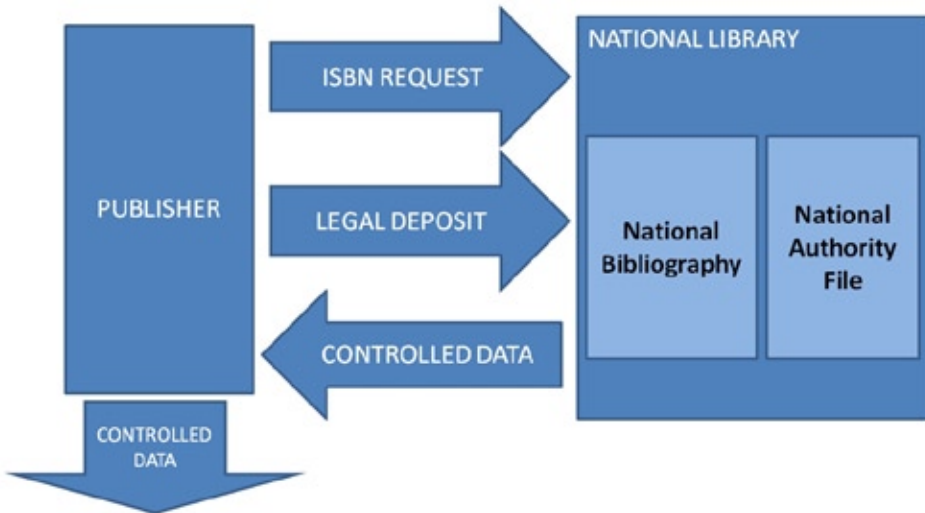


Figure 3 Metadata control

By means of disseminating new standards and methods in order to optimize data flow in the private sector and to guarantee high quality of metadata libraries might have a direct stimulating impact on the economy instead of the current indirect influence such as the support of education and research. This new mission of a library as a kind of economic catalyst (or infrastructure provider) mostly depends on its ability to exploit the potential lying in the use of ONIX which is being supported and encouraged not only by some international projects and initials such as ARROW (= Accessible Registries of Rights Information and Orphan Works Towards Europeana) and Linked Heritage but also by ISO 2108:2005 which clearly states that “[t]he metadata requirements for the ISBN system should be compatible with the ONIX International product information standards maintained by EDItEUR and its associated organizations”.

What to do with librarian skills?

What else do we need the librarians for if they're divested of one of their essential functions, i.e. creating bibliographic descriptions? In a world where self-check systems are quite common and the registration of patrons is supported by self-service devices there still remains some places where librarians' skills can be exploited.

In Web 2.0 participating culture has its undeniable impact on the librarian work as well. Many thought that our patrons would catalogue the books instead of the librarians in the near future. There are online public access catalogues where one has the opportunity to add subject terms to bibliographic descriptions, which are made searchable for other users. However, these subject access vocabularies of tags provided by users, built up in an automated way, so called folksonomies, have proved to be insufficient for replacing controlled vocabularies such as thesauri mainly edited by librarians. Skills of an average patron don't seem to be enough to substitute in an efficient way the knowledge of a professional. In spite of this fact linking folksonomies and controlled vocabularies is an idea to consider, also, libraries should support producers of publications, such as publishers and authors, in categorizing their own publications by using classification systems developed by library professionals.

Conclusion

To answer the question in the title of the article, we can conclude that in the not too distant future bibliographic descriptions will probably be created by persons of the private sector, employees of publishers for instance, who, of course, may arrive from the library area, bringing with them and revitalizing their professional skills. In this new environment the library will have its role as a sort of check-point, a place for data control. Library professional will be responsible for constructing the means of this control by editing authority files, controlled vocabularies, thesauri, all the stuff that is crucial for creating the semantic web.

Readers interested in the topic might consult the following sources as well:

Controlled vocabulary – http://en.wikipedia.org/wiki/Controlled_vocabulary

EDItEUR – <http://www.editeur.org/>

Folksonomy – <http://en.wikipedia.org/wiki/Folksonomy>

FRBR — Functional Requirements for Bibliographic Records – <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

MARC 21 Format for Bibliographic Data – <http://www.loc.gov/marc/bibliographic/>

MARC standards – http://en.wikipedia.org/wiki/MARC_standards

ONIX — Online INformation Exchange – <http://www.editeur.org/8/ONIX/>

RDA — Resource Description and Access – http://en.wikipedia.org/wiki/Resource_Description_and_Access

WorldCat – <http://www.worldcat.org/>