

Legal deposit of ebooks in France and its bearings on cataloguing and the National Bibliography

Sophie Derrot

Legal deposit department, Bibliothèque nationale de France, Paris, France
sophie.derrot@bnf.fr

Mathilde Koskas

Legal deposit department, Bibliothèque nationale de France, Paris, France
(presenter)
mathilde.koskas@bnf.fr



Copyright © 2015 by S. Derrot and M. Koskas. 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:

French law on digital legal deposit covers websites and online content as well as ebooks. It imposes no obligation to produce a bibliography in the traditional form, indexing being sufficient. But despite their innovative characteristics, ebooks are still books, and their bibliographic metadata is closer to that of printed materials than to the indexation of web archives. To set up a complete ebook deposit workflow, the Bibliothèque nationale de France (BnF) benefits from its experience with digital documents and its tradition of legal deposit. This paper aims to present the questions that it faces when dealing with the cataloguing of ebooks and the management of their metadata, and the solutions that are emerging.

Keywords: ebooks; legal deposit; national bibliography; automated metadata retrieval; cataloguing standards.

The nature of ebooks is dual, between electronic documents and traditional books, and their status is ambiguous, as the debate within the European Union about their tax rate reveals. The

Bibliothèque nationale de France has the legal mission of collecting online content as well as printed books. For a long time, this online content has been mostly in the form of websites, but since 2012, the BnF has been working on a specific legal deposit for ebooks. Such a legal deposit falls within the law from 2006 on digital content, but the workflow which allows a better treatment by the BnF and an easier way to deposit for publishers is very close to the traditional legal deposit for printed books. The major issue is that the digital legal deposit law differs on several points from printed legal deposit: there is no obligation of comprehensiveness or of producing a national bibliography. These differences have to be kept in mind because they draw a line between the treatment of ebooks and that of printed books.

The dematerialisation of the legal deposit process has begun with the possibility for publishers to fill in their mandatory declaration online, thanks to a dedicated Extranet. More recently, they've also been able to mandate their distributor to send a metadata flow that acts as that declaration. The next step is the processing of ebooks and their metadata arriving at the BnF through an automatic process. To achieve this new evolution of legal deposit, the BnF has to adopt a position on important bibliographic questions and to resolve tensions between the bibliographic theory and the organisation of such an institution. In many ways, the postulate of a proximity with printed books has its limits, which we must take into account when we decide how best to create and disseminate bibliographic records for ebooks.

Working with publishers' metadata

The project of a legal deposit of ebooks began in mid-2012, with a dialogue with our partners from the publishing world, the Syndicat national de l'édition (National Publishing Syndicate). This first phase was important because the ebook market was quite unknown to the BnF in many of its dimensions: how many ebooks are published per year? What formats are the most common? How does the publishing and commercialisation workflow work? All these questions were crucial to the organisation of a viable legal deposit. One of the first questions we faced during these interviews was that of metadata: how does the information about an ebook go from the author and publisher to the final seller and consumer? As our knowledge of the workflow increased, we realised that some of the main actors add metadata during the process and that the need of a standardised way to transport metadata has therefore pre-existed the BnF demand.

The standard widely used within the publishing world (in France and worldwide) is one of the ONIX standards, dedicated to books. *ONIX for Books* was originally developed by EDItEUR and is maintained today by this company and an international steering committee¹. It is designed to support computer-to-computer communication, in XML language. An ONIX file can describe printed books as well as electronic books, and can express bibliographic, technical and commercial metadata thanks to a great variety of fields precisely described in an up-to-date documentation available online. Within the French ebook market, the ONIX file describing an ebook is created by the distributor based on the information supplied by the publisher.

The BnF already has experience of processing distributors' ONIX files for printed books. Since late 2014, publishers have the possibility to make their mandatory legal deposit declaration directly through an ONIX flow provided by their distributors to the BnF. Books

1. <http://www.editeur.org/8/ONIX/>

are sent separately and when they arrive at the BnF, their digital declarations are checked and then automatically transferred to the cataloguing application for treatment by the cataloguers. This experience was very helpful for the implementation of the ebook legal deposit workflow.

All of our partners use ONIX 3.0 files to transfer ebooks metadata from the distributor to the online sellers. They provided us with samples from their production, so we could have an idea of the quality of these ONIX files. Meanwhile, we conducted a reflection on the form that an ebook record should take within the BnF's General Catalogue². We compared the information from both sources and thus made an ideal ONIX model, with mandatory, optional or wished fields. This model was sent to our partners; it tends to evolve according to the ongoing reflection on the subject and the information provided by distributors.

Samples of the BnF ONIX model, concerning the ebook format and its title:

ONIX 3 fields	Values and description ³	ONIX standard ⁴	BnF wishes
<DescriptiveDetail>		M	M
<ProductComposition>	00: Single-item retail product	M	M
	10: Multiple-item retail product		
	11: Multiple-item collection, retailed as separate parts		
<ProductForm>	EA: Digital (delivered electronically)	M	M
	EB: Digital download and online		
	EC: Digital online		
	ED: Digital download		
<ProductFormDetail>	E101: EPUB		M
	E107: PDF		
	E108: PDF/A		
	E200 (if E101): Reflowable		
	E201 (if E101): Fixed format		

[...]

<TitleDetail>		R	R
<TitleType>	01: Distinctive title	M	M
	03: Title in original language		
	Other values are optional		
<TitleElement>			
<TitleElementLevel>	01: Product	M	M
<TitleText>	Text of a title element	M	M
<TitlePrefix>	Text at the beginning of a title element	O	M if prefix
<TitleWithoutPrefix>	"No prefix" indicator	O	M if prefix
<Subtitle>			M if existing
</TitleElement>			
</TitleDetail>			

2. The BnF already buys ebooks (and access to ebooks). However they are not described within the General Catalogue but within a specific application dedicated to "Electronic resources", such as databases or e-journals.

3. These values refer to code-lists attached to the standard.

4. M is for Mandatory; R for Repeatable; O for Optional.

Thanks to the ONIX format, the BnF and all its partners have a common way to handle metadata. The ONIX file describing an ebook is mandatory for the legal deposit of that ebook. Both metadata and ebook files will be uploaded to the BnF sFTP platform by the distributor. The automatic treatment begins by checks (presence of both files, checks of EPUB and PDF, validation of the ONIX file) and once the files are certified to be well-formed and matching the BnF's requirements, the metadata is converted for the next step, cataloguing.

Organisation of the cataloguing workflow

The aim of this agreement with the publishers was to fit a workflow by which metadata enters automatically the cataloguing channel of the library. The BnF already had such an experience dealing with ONIX files describing printed books, as mentioned earlier. This work is intellectually still quite close to the traditional method: printed books arrive in the usual way and are handled by cataloguers.

The ebooks workflow is quite similar in terms of the treatment of metadata. The ONIX file is converted into an INTERMARC record in the same way, whether it describes a printed or an electronic book. The result of this conversion is a rather full basis for a bibliographic record, already mentioning a majority of the needed information (title, contributors, publisher, etc.). The cataloguers then have to check this information and enrich the record to reference level with all authorities, indexing and correct bibliographic description.

The organisation of the cataloguing of ebooks is still in its prefiguration stage, but the major elements are already in place. During the prototype phase, a taskforce of five cataloguers and a coordinator will undertake this cataloguing. Joining the library between June and December 2015, they will be an integral part of the unit which already catalogues all the printed books arriving to the legal deposit. This team will be able to catalogue printed books as well as ebooks; at the beginning, they will be the only ones to catalogue ebooks, but the aim is to spread this expertise and to make every cataloguer of this unit able to deal with ebooks.

However, beginning with a small team will enable us to study and experiment a new way of working: what kind of work station do these cataloguers need (two screens, for example)? How do we organise their training? As the whole workflow is automated, the cataloguers can have access to the ebooks only when they are available in the digital library, after the preservation step⁵. So their work on the record really finishes and closes the process. Thus they will have the possibility to catalogue with the book “in hand” – or, more precisely, before their eyes. But some questions remain open, such as the suitability of the public interface of the digital library to a consultation in order to catalogue.

Besides these very pragmatic points, we had to think on an almost conceptual subject: the definition of a record describing an ebook. This definition determined the mapping of the conversion from the publishers' to the library metadata and thus the information the cataloguers will have to deal with.

What could an ebook record within the catalogue of the BnF be like?

To help answer this question, a specific working group on the bibliographic aspects was created from the beginning of the project, to study the needs of the ebooks workflow in this

5. On the global workflow, see [1].

matter. Establishing the bibliographical context had indeed been useful to support other aspects of the project (human resources for cataloguing or preservation, for example). This working group still meets regularly, to follow evolutions on the subject.

The people participating in this working group reflect the complexity of the nature of ebooks: members of the Legal Deposit department (leader of the project and department of the cataloguers of deposited books), the Audiovisual department (which takes care of the legal deposit of multimedia documents) and the Digital and Bibliographic Information department (which is in charge of metadata consistency and standards compliance).

To define what an ebook record could be, this group had to face ontological questions: should the bibliographical description of an ebook be closer to that of a printed monograph or of an electronic document? One of the major elements of context was that these ebooks will be treated by cataloguers who deal with printed monographs. For the time being, the decision was made to treat these documents more as books than as electronic documents.

Once this choice had been made, the next question was how to integrate new metadata in the BnF model. On the one hand, some information doesn't yet exist within the General Catalogue of the library (a file's format and its versions, for example). On the other hand, some bibliographical uses already exist to describe the very diverse collection of the BnF and had to be taken into account: for instance, the notion of the weight of a document is declared in a specific field of our format, but that field is currently used to record the weight of a coin or medal, so another field had to be found for the "weight" of a file (its size).

Meanwhile, this reflection on the ideal ebook record allows innovations that are a driving force for printed books. For instance, when working with the publishers on the ONIX data model, we asked them to provide us with the European Article Numbering (EAN) of other versions of the same book, a field that exists in ONIX (<RelatedProduct>). Thanks to this information, a link is created between the different records during the automated conversion of ONIX into MARC records, if the catalogue already holds a record with the said EAN. The same will also be possible with authority records for authors, by means of the International Standard Name Identifiers (ISNI). Using the links between records for the same book in different formats, it will then be possible to retrieve all the relevant information from the most complete record to automatically enrich the other one(s). Depending on which version of the book is deposited and catalogued first, this means not only that a record created for a printed book can be used to enrich a record for a PDF or an EPUB, but that the reverse will also be possible.

Consequently, records of different levels of completeness and standard compliance will coexist within the catalogue. It is a common situation, but it will be particularly intricate because it won't be a simple case of printed and electronic books having different quality records. Some ebooks will have richer records because of the data retrieved from a pre-existing printed book record, and the reverse might be true too. Thus, the very process will induce such a symbiosis between records that we won't be able (should we want it) to separate workflows for printed and electronic books.

The overlap also exists in terms of standards: the French standard applicable to ebooks is NF Z 44-082 *Catalogage des documents électroniques*, which dates back to 1999. It only gives indications about recording the specific characteristics of electronic documents, the rest of the resource being described under the specifications of the relevant standard (printed books, maps, serials, etc.). Of course, that standard, which predates the rise of ebooks, is now out-dated. That is another point we had to take into account to devise the cataloguing framework for ebooks.

“Within a catalogue and beyond” [2]

Our first response to the question “How do we let ebooks into the catalogue?” was expressed in terms of format, because it is how we catalogue, in a most concrete way. But the answer lies with standards.

There is no set of standards specifically dedicated to ebooks in France today. The rules are derived from those for printed books and for electronic material, both French cataloguing rules and ISBD. The issue of updating French cataloguing rules is broader than that of ebooks, especially in the global context of switching to RDA. France has had a working group on the adoption of the latter cataloguing rules since 2010. In late 2014, both national bibliographic agencies⁶ decided on a policy of converging towards RDA: adopting as much of the code as possible, without compromising our national practices where they are richer or more conform to FRBR than RDA currently is.

Under the new name “Bibliographic transition”, the group is now tasked with updating French standards into what we call RDA-FR. As it is a very long process, parts of the new standard are published and gradually replace the corresponding parts of French standards. It means that we are currently working with two standards at the same time. But RDA’s consolidated approach, with general rules that encompass all types of documents, could be useful toward a convergence of records.

A big hurdle for librarians is the nature of ebooks. We tend to see them as books first and foremost, and are tempted to deal with them as we do with printed legal deposit material: create reference-level records for all ebooks, and publish them in the National Bibliography.

However it is not what the law tell us. Ebooks fall within the scope of the law on digital legal deposit, which was created with the web in mind. It means no obligation for the National Library of instituting a comprehensive legal deposit for ebooks, or of publishing a national bibliography. But our whole reflection on ebooks led us to emphasise the convergences rather than the differences between printed and electronic books, and to create a descriptive model closer to printed books than the web.

Furthermore, we know that there is a need for a comprehensive, open national register or database for ebooks, with reference-level, trustworthy records; this was expressed in the Lescure report to the minister of Culture in 2013 [3]. The commercial database Electre began providing metadata for ebooks lately. But the BnF has a unique position of observer of the publishing world, both printed and digital. Moreover, the very principle of producing individual records for ebooks in the library’s catalogue seems an argument in favour of taking the next step and publishing them in the French National Bibliography.

The question of course is how?

In terms of quality, the National Library is a service provider to a national community, especially through the National Bibliography, which, unlike the General Catalogue, traditionally holds only records of the highest quality. But the *Guidelines for bibliographies in the electronic age* [4] have identified the problem and accepted that, with ebooks, different levels of description are acceptable. It remains for us to determine what level of description to adopt, bearing in mind that our process makes the records for ebooks and printed books tightly interdependent.

6. Agence bibliographique de l’enseignement supérieur (ABES), the Bibliographic Agency for University Libraries, and the Bibliothèque nationale de France.

In terms of display, should records be displayed in the books section of the bibliography, mixed with printed books, with only a dedicated title index? In terms of IT developments it would be the quickest and simplest solution. It might be a temporary one though, until a new section of the bibliography or a new website for the whole publication can be created. If we decide on updating the current website, what form would it be better for it to take? Once we have (potentially) several versions of the same title, a “FRBR” display might be considered. A look at the other websites the BnF uses to display metadata would be the first step of that particular study.

Conclusion

Once we look past the first impulse of doing just the same as we’ve “always” done, we must ask ourselves what service is expected from us. With ebooks, the context of legal deposit changes: ebooks carry with them formatted and standardised metadata. Even if the standards are not those we librarians use, we are provided with more than basic records that already provide a service to users. But is it the service we need to provide? At this point, it looks like a very good step in the right direction, but with still some way to go to be quite up to standards. So how do we close this gap? Automation will be a great help, retrieving data to have the same level of description between records for the same title, but it has its limits. That is where human intervention is necessary.

This project is highlighting the convergence of questions libraries are confronted with: the adoption of RDA, the revision of standards, the automation of cataloguing. To answer these questions, we have to take a more holistic view of our activity, transcending the library’s sometimes rigid structure.

Acknowledgments

The authors acknowledge Peter Stirling for his language review and his great patience.

References

- [1] S. Derrot and C. Oury, “Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France”, in IFLA WLIC 2014, 16-22 August 2014, Lyon, France. [<http://library.ifla.org/830/>]
- [2] International Federation of Library Associations and Institutions, Tillett, B. B., & Cristán, A. L., *IFLA cataloguing principles: The Statement of International Cataloguing Principles (ICP) and its glossary : in 20 languages*. München: K.G. Saur, 2009. [http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf]
- [3] Pierre Lescure, *Contributions aux politiques culturelles à l’ère numérique*. [Rapport à la ministre de la Culture], 2013. [<http://www.ladocumentationfrancaise.fr/rapports-publics/134000278/index.shtml>]
- [4] M. Žumer (ed.), *National bibliographies in the digital age: guidance and new directions*. München: Saur, 2009. [<http://public.ebib.com/choice/publicfullrecord.aspx?p=454007>]