



## Are you ready to dive in? A case for Open Data in national libraries

**Gildas Illien**

Director, Bibliothèque nationale de France  
Paris, France

### Meeting:

**181 — National libraries and open data: new discovery and access services — National Libraries**

### Abstract :

*As acknowledged in the 2011 W3C report on Library Linked Data<sup>1</sup>, libraries might have an important role to play in the development of Linked Open Data (LOD), which is also a unique opportunity for them to make their metadata more visible and “reusable” on the Web, to demonstrate their public value after having invested efforts and money in digitizing their resources. This presentation, based on the experience of the Bibliothèque nationale de France (BnF) proposes to review the pros and cons of open data from the specific perspective of national libraries. When it comes to lead advocacy efforts in this field, what should we keep in mind and what experiences can we share? It will present the BnF’s current strategy and experience in building a vision for open data at the National Library and removing obstacles, step by step. Although the French National Library is only half way through the path to open data, it seems interesting to share the lessons learnt at this stage.*

### Introduction: what Linked Open Data means to national libraries

Library Linked Open Data supporters believe that their metadata should be more visible, easier to search, to link and to reuse directly online (either by humans or by machines) rather than within catalogue silos. It is their belief that by engaging in this movement, libraries could serve many more users and organizations and facilitate new business opportunities and collaborations within and beyond their own community. After many years of hearing about the end of cataloguing, it seems that the need for trusted and structured bibliographic information is actually not fading. Quite on the contrary, the deluge of data creates an

<sup>1</sup> <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>

increasing demand for interoperable clusters of standardized metadata. This is about helping users finding more easily authoritative information within the mass of the Web. This is about avoiding duplicating cataloguing efforts in a context of budget and staffing cuts. This is about stimulating innovation and growth by opening and sharing library data more widely.

Could the so called upcoming third Age of the Web – the Web of data – then be the Golden Age of Library metadata as well? As acknowledged in the 2011 W3C report on *Library Linked Data*<sup>2</sup>, librarians may have a special role to play in the development of Linked Open Data (LOD). In return, this new region and these new standards of the Web may provide a unique opportunity for libraries to make the most of their metadata legacy and daily production. A key challenge for many institutions is thus to enforce innovative strategies in order to transform and display their metadata outside traditional catalogues and bibliographies hidden in the Deep Web so that they become visible and linkable in the LOD environment together with other valuable datasets.

This great challenge is especially crucial to national libraries. At times where budgetary constraints increasingly lead heritage institutions to demonstrate their public purpose, value and output, they need to return more benefits to the society at large from the public effort and money which have been invested over the past decades in cataloguing their collections. In the past two years, various experimentations and proofs of concepts have popped up among several national libraries (in Sweden, Germany, France, the UK, the USA...), which show that a growing number of institutions have started exploring this new path. Transforming preexisting MARC records and authority vocabularies in RDF triples ; starting to implement the FRBR model ; playing with the semantic Web standards ; building applications and datasets of a new, Linked Data-friendly type: this is what looking at LOD means to them at this stage.

According to the same W3C report, linking data using semantic web standards is however not sufficient. While “Linked Data” refers “to data published in accordance with principles designed to facilitate linkages among datasets, element sets, and value vocabularies”, that is, the technical interoperability of datasets, “Open Data” focuses on their *legal* interoperability in order to maximize their circulation and reuse.

This paper does not address Linked Data technologies specific challenges as such but focuses on the *Open* data issue, that is, the political, legal, economic and cultural challenges national libraries need to deal with when considering choosing open licenses such as CC-0 or CC-BY for their metadata.

Why should Open Data be a particular challenge for national libraries? These longstanding heritage institutions are usually the host or coordinating point for bibliographic production and control. They play a key role at national and international level in maintaining national bibliographies and authority files. They know very well how to produce and exchange structured metadata in standardized ways. They have usually developed tight connections with national governments and legislation-makers. In sum, because of their audience, authority, responsibilities and reputation, national libraries are more likely than others to drive public opinion, public policy and best practices, and to endorse leadership roles in advocating for Open Data. Without their commitment and resources, it would be difficult to convince decision-makers as well as librarian communities and networks to make Open Data come true.

---

<sup>2</sup> <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>

There are however many good reasons for national libraries to fear the Open Data movement: fear of not being competitive, fear of being robbed from their legacy, and, knowing how perfectionist librarians can be, even the fear or the shame to display “dirty”, poor or incomplete data. Those advocating for Open Data within the ancient walls of national libraries need to be aware that it is not a straightforward message to pass along national library stakeholders. In Europe, giving up for free decades if not centuries of cataloguing work to the great, anonymous cloud may remind us the time when people had to abandon their national currency to adopt the euro. Furthermore, there are legal, economic and political obstacles to overcome, which require that a dialogue be set up on these specific issues at the very top of the library management together with public authorities. This is not an easy task considering the degree of technical and legal knowledge which is required to fully understand the mechanisms and implications of LOD.

How can librarians discuss Open Data with their overseers? This is the main topic of this paper. Our presentation, based on the recent experience of the Bibliothèque nationale de France (BnF), proposes to review the pros and cons of Open Data from the specific perspective of a national library. It tells the story of how the BnF became a supporter of Open Data in a year's time without initially planning to. Although the Library is only half way through the path to Open Data, we wanted to share the obstacles met and the lessons learnt at this stage. As the French situation is far from being definitely settled, this article only reflects the current state of the art in July 2012.

## **1 - Where it all started: BnF's initial framework for metadata reuse, from UNIMARC delivery to Dublin Core repositories**

To clarify the scope and settings of our story, it should first be noted that when addressing data rights in this paper, we're only referring to metadata - not the digital documents. By metadata is meant: bibliographic records created by the BnF to describe documents and resources received under the legal deposit scheme or otherwise acquired and signaled in both its main catalogue (*Catalogue général de la BnF*) and its special catalogue for archives and manuscripts (*BnF Archives et manuscrits*), along with authority records (persons, corporate bodies, titles, uniform titles for music, geographic names, RAMEAU subject authority records). The whole set composes the bibliographic and authority databases of the BnF, whatever the data format (production format, extraction and distribution format, display format).

The BnF did not decide to open its metadata in the first place. As of today, open licenses only apply to certain datasets in RDF format as we will see. Other metadata in MARC and Dublin Core formats are still under more restrictive licenses. Although the situation is rapidly evolving and the ultimate wish of the institution is now to adopt open licenses for all of BnF's metadata (whatever the format, type of reuse or retrieval protocol), the current situation can therefore be described as intermediary. This very situation, where some metadata is open while the rest is not, happens to be the outcome of a process of convergence between internal factors within the Library and recent events outside of it, rather than of a straightforward decision-making process.

## **Terms of use**

Until December 2011, the terms of use for all of BnF metadata were in line with the French Law n°78-753 of 17 July 1978, modified by the *Ordonnance* of 6th June 2005 on the freedom to access administrative documents and re-use public information. This general framework is as follows:

In the case of not-for-profit use, any person or body who retrieves records from the BnF is allowed to use, adapt, modify and disseminate them. In return for that permission, the purchaser is committed to preserve permanently within the record a statement of its source, i.e. the content of field 001 of the BnF record, in the appropriate field of the target format (e.g. field 035 in UNIMARC). In addition, if the data is adapted or modified, the user must notify third parties that it is the case.

In the case of commercial use, the use of data for commercial purposes (dissemination as part of a product or service intended to be made available to third parties, for a fee or for free as long as it is to a commercial entity) is submitted to an agreement with the Library and the payment of an annual fee.

A similar policy applies to digitized reproductions to be found on the BnF digital library Gallica (<http://www.gallica.bnf.fr>): most of those documents are in public domain and can be used freely with obligation of attribution, unless commercial use is made of them, in which case licensing fees apply.

In practice, BnF records have been made available for reuse in a variety of product types (retrospective products, current products, by demand...), formats (UNIMARC, Dublin Core) and retrieval protocols (Z39 50, FTP, OAI-PMH) over the years. The same terms of use apply regardless of the format or retrieval protocol, although the Library never had the technical nor human means to actively monitor the actual use and reuse of its metadata.

The most refined retrieval and delivery services have been quite logically designed for bibliographic and authority records from the National Bibliography, since the BnF has the legal responsibility and obligation to provide information of reference for all materials published in France. However, some of those retrieval services also cover the totality of records present in the BnF General catalogue, including foreign material outside of legal deposit scope. Even in that case, most of the metadata is actually home made, as imports of foreign records from other libraries (through OCLC WorldCat services) only started a few years ago. As a result, BnF is the primary producer and owner of ca. 99.5% of its metadata, which gives the Library the possibility to define its licensing policy without relying on too many legal dependencies.

## **Delivering UNIMARC records**

First, the official National Bibliography (<http://bibliographienationale.bnf.fr>) established by the BnF on the basis of systematic description of materials received under the legal deposit scheme is accessible online. This website facility provides free-of-charge open access to bibliographic records. From each record, it is possible to jump directly to the BnF General Catalogue via a link carried by each record's number (FRBNF). This link provides access to catalogue functions which allow the end user to reuse the record by means of email, printing, or bookmarking. These basic retrieval services work for all types of records. They are

designed for individual users (e.g. researchers, students...) willing to reuse records to produce a bibliography for instance.

Second, the online record transfer service enables all bibliographic and authority records from the BnF General catalogue to be downloaded by creating a "basket" (selection) of records, which will be sent to the user via the BnF FTP server. Although the service is free of charge, users must register and set up a user account. Downloading records via Z 39.50 is another possibility. These services are designed for professionals usually working at a limited scale, typically a small or medium size public library.

Last, current and retrospective bibliographic products (including both bibliographic and authority records) are also available on request. These services only cover the National bibliography production. A subscription is required to use these products. Specific sorting may also be requested. Custom sorting is subject to specific delivery and pricing arrangements. Raw data is sent out on CD-ROM, made available on BnF's FTP server, or made available on the customer's FTP server. This service is designed for larger libraries working at a bigger scale (those are exempted from fees as they do not serve a commercial purpose), or companies from the publishing or library management system industries (to whom fees do apply).

### **Harvesting Dublin Core records in OAI repositories**

A few years ago, the BnF also opened two OAI repositories aimed at facilitating access to its collections and metadata by means of automated harvesting. The first of these repositories, OAI-NUM, contains the records for all documents digitized by the BnF and accessible via the Gallica digital library. The second, OAI-CAT, is intended to contain records for all BnF documents, whether digitized or not. It currently contains over 11 million bibliographic records.

These two repositories are organized into sets of records: by document type, by topic-based file or collection in accordance with the structure used to organize collections, by subject index under the Dewey Decimal Classification System. There are regular updates, and these datasets can be freely harvested under the terms of the OAI-PMH protocol<sup>3</sup>, enabling digitized documents and bibliographic data produced by the BnF to be referenced in other databases. In accordance with the protocol, BnF uses the unqualified Dublin Core for metadata describing its documents. It does not include authority records.

The use of OAI repositories is well designed for larger scale data transfers requiring frequent updates. The OAI-NUM repository has proved especially useful to organize institutional cooperation for regional or thematic digitization collaboration initiatives such as the Europeana digital library. However, several limitations are met as the Dublin Core profile remains quite poor and because authority records are not part of the datasets. For these reasons, metadata transfers required for BnF's participation to major services run by OCLC (e.g. WorldCat or VIAF) still use FTP transfers. OAI repositories are fit for some purposes but can't cover all data exchange needs.

---

<sup>3</sup> <http://www.openarchives.org>

## Contrasted and questionable outcomes

A wide range of technical possibilities has thus been explored over the years to disseminate BnF's metadata, mostly targeting public libraries in order to encourage the reuse of BnF records and therefore avoiding duplication of cataloguing work at the national level. However, the number of re-users and the feed back about usage are far from satisfactory. More than 4,000 users might be currently registered to the online transfer subscription service, but only 780 used it actively over the past two years, while there are more than 16,000 public libraries in France. BnF bibliographic products currently have only 14 paying customers, from which the Library derives financial benefits which are small in comparison with the actual cost of managing the payment and accounting process and in light of the Library's general budget.

In 2012, the BnF conducted an online survey on metadata retrieving among French libraries, to which over 700 institutions replied. Although the results showed that institutions reusing BnF's records by any of the proposed means were generally satisfied with the data and the service (71% of them declared that they derived records from the BnF on a more or less regular basis), most of them tended to say that it was not a very straightforward nor well-known offer. As to the OAI repositories, it is by definition difficult to know how often they are used and by whom.

Looking at things from a different angle, recent annual audience statistics about online traffic bring slightly disturbing news: while in 2011 BnF website ([www.bnf.fr](http://www.bnf.fr)) received 20 million visits, and the Gallica Digital library 9,3 million, the online General catalogue's score was 4,6 million visits and the online National Bibliography only 0,13 million. In terms of reuse and visibility, it doesn't seem unfair to say that BnF bibliographic resources could do better.

Of course, reconsidering the scope, design, look-and-feel and functionalities of catalogues has been a constant matter of concern and an obvious task to the Library system managers and bibliographic experts over the past years in order to adjust to patrons' changing practices and expectations in the digital world. One major concern has notably been the description of electronic resources. Those purchased from major electronic publishers are not described in the BnF General catalogue but in a separate, commercial system which provides direct access to the resources on external servers owned by the publishers themselves. Those collected by the BnF through the legal deposit scheme by means of Web harvesting aren't described in the catalogue either, but available via a separate interface providing search by URL and some full-text limited possibilities. This is a first set of urging issues, familiar to many other libraries, to which the BnF doesn't have an answer yet, and which is not in the scope of this paper.

If we remain focused on non-digital resources, that is, the bibliographic legacy and current cataloguing production, the key question remains: how can the Library improve their visibility? Considering that cataloguing activities currently employ no less than 250 FTE (at least 10% of total staff of the BnF), there is reason to believe that in view of this investment, the outcome for the national community should be more significant. It has become clear to BnF's management that at least some marketing efforts if not an even more radical and ambitious re-thinking of the whole bibliographic activity is needed in the next 5 years in order to better disseminate bibliographic resources. And it is by looking at the latest technological developments, that the Web itself and the Web of data in particular appeared to offer potential answers to this challenge.

It is in this context, and building from these findings, that the the BnF started the data.bnf.fr project which, in return, ultimately confronted the Library with the Open Data discussion.

## **2 – How things changed: RDF metadata dumps with an open license in the Web of data**

In July 2011, the BnF publicly launched its Linked Data proof of concept project “data.bnf.fr” (<http://www.data.bnf.fr>) after two years of conception. We will not present this project in detail here<sup>4</sup>, but only focus on how the making of it actually impacted the Library’s way of thinking about Open data.

### **A proof of concept to explore the Linked Data environment: Data.bnf.fr**

The application was designed to be used by individual, human-driven browsers, navigating through the various pages of the website, which are based on FRBR entities: authors, works and subjects. But the major innovation was that the application was also – and primarily-intended to be used by machines. Data.bnf.fr was indeed to group and expose on the Web data in RDF form coming from heterogeneous sources. Thanks to significant and unique identifiers (URIs), the data could be easily indexed by search engines and densely linked to other resources, either internal to the BnF (the General catalogue, the Archives and manuscripts catalogue, Gallica...) or external (the Union catalogue for French Research libraries *Sudoc*, WorldCat, Wikipedia...). In order to facilitate data dissemination and reuse, all datasets were to be made available for download as RDF dumps.

The use of semantic Web technologies was necessary for linking these resources, but as the project team started designing this new object, many questions remained still unanswered. We did not know whether the application would be able to work at scale with such a variety of data sources. We were not sure how the project would fit in the overall library bibliographic environment. And our terms of use being what they were for bibliographic records, we did not even think of opening its data. In the fall of 2011 however, only a few months after the launch of the project, two events brought the most unusual political attention to the subject of metadata.

### **The Europeana Data Exchange Agreement**

First, the Europeana Foundation asked its contributors to sign a new Data exchange Agreement (DEA)<sup>5</sup>. This agreement, now well known to most librarians, means a shift in vision as to metadata rights and reuse. In Europe, it is the DEA which really forced national libraries to discuss Open Data. Signing this agreement involved accepting to apply a CC-0<sup>6</sup> (Creative Commons Zero) licence for the metadata attached to all digitized collections shared in Europeana. But such a decision would naturally lead to envisage the possibility of similar licensing changes for the other metadata produced by contributing institutions.

The arguments highlighted by the Europeana Foundation to advocate for the CC-0 licence are summarized as follows on the Europeana’s website:

---

<sup>4</sup> For a presentation of data.bnf.fr in English, see : <http://data.bnf.fr/docs/databnf-presentation-en.pdf>

<sup>5</sup> <http://pro.europeana.eu/data-exchange-agreement>

<sup>6</sup> <http://creativecommons.org/publicdomain/zero/1.0/legalcode>

“Why support metadata re-use? - It brings tangible benefits for heritage organizations:

- increases traffic to your site;
- allows the Europeana APIs to be widely used;
- provides income-generating opportunities, e.g. for image licensing;
- enriches your data for your own use;
- increases usability and knowledge generation through Linked Open Data applications;
- enables the development of innovative services and new revenue streams”<sup>7</sup>

Being a strong advocate and major player in the creation of Europeana, and one of its most significant contributors, the BnF could not consider taking its resources away from the European digital library. Although the BnF has not signed this agreement at the moment this article is being written, BnF President has publicly expressed his full support of this initiative.

### **Data.gouv.fr and a new type of open license by Etalab**

Approximately at the same time, the French Government, inspired by similar initiatives in the US and in the UK, launched its own Open Data initiative for the public sector: <http://www.data.gouv.fr>, whose portal website opened in December 2011. Under the authority of the French Prime Minister’s services, this new portal was created to encourage and federate public Open Data initiatives.

One important aspect of this project was the creation of a new open licence type<sup>8</sup>. Although creative commons technically hardly fit in the French code of intellectual property, this licence could be compared to a CC-BY type of licence. It is free, open, encourages reuse and reproduction, including for commercial purposes. But unlike the CC-0 licence, it requires attribution. It is compatible with other foreign or international similar licences such as the Open Government Licence (UK Government) and the OCD-BY and CC-BY 2.0 licences.

### **How metadata rights became a strategic issue**

Both events raised much attention and discussion about Open Data within the Library. They brought ideal conditions to undertake a dialogue with decision-makers. We talked with the Library lawyers to look into the details of different metadata licence types and compare their respective risks and assets from a legal perspective. We talked to the Library accountants to try to demonstrate that putting our metadata in the Open Data cloud would bring more benefits to the BnF than trying to sell them. The latter proved quite difficult a conversation because we did not have (and still do not have) a business plan or model which could demonstrate with figures the benefits the Library would derive from this change of policy. We discussed with the General Direction the strategic opportunities the Library could get from aligning its metadata rights policy both with European and national initiatives.

We also had to convince ourselves and our teams that some risks were to be taken. We had to bet that Open Data would, at some point, leverage our metadata and our professional activities to a level of visibility and audience which they had never reached. We had to bet that this move could help protecting our jobs, since you can’t develop and promote a sustainable model in the Open Data world if you lose the core staff resources you need to produce metadata of quality, that is, cataloguers. We also had to take the chance that no bigger player

---

<sup>7</sup> <http://pro.europeana.eu/web/guest/support-for-open-data>

<sup>8</sup> <http://www.data.gouv.fr/Licence-Ouverte-Open-Licence>



would try to re-sell to us our own data under a different form at some point. Of course, all these points still need to be proven

The RDF sets from data.bnf.fr were the first significant cultural datasets to be published on the data.gouv.fr portal when the latter was officially launched in December 2011. Simultaneously, BnF management took the decision to adopt the French Government Open Data licence, making the Library a pioneer cultural institution in this area. This was an extremely important milestone for the project.

### **The benefits from joining the Linked Open Data: a first impression**

Data.bnf.fr keeps developing, with more contents and functional improvements being added every month. It has reached a stage of maturity beyond proof of concept status. Although still in its infancy, it currently describes more than 200.000 FRBR entities, with 2.5 million linked resources and more than 6 million RDF triples. Its audience is still modest but growing quite fast (102 000 visits altogether over a 6 months period in 2011 after the launch, but 67 000 just in one month last May). Furthermore, it drives extra traffic towards other BnF applications, the Gallica digital library and the BnF General catalogue in particular. Given those encouraging first results, it has been possible to secure staff and budget to continue the project in the next three years.

It is difficult to identify all types of reuse made from this growing dataset. To extend the discussion from the previous section of this paper presenting the more traditional possibilities for metadata delivery and reuse, it should be first highlighted that the data.bnf.fr project has great potential for metadata reuse by libraries specifically. The BnF envisions that 5 or 10 years from now, public libraries from France (and elsewhere) could use Semantic Web technologies to link their local catalogues to the metadata produced by the BnF rather than continue retrieving them via Z3950, FTP servers or OAI repositories.

The technical assumption underlying this vision is that records (or say, bibliographic and authority information at large) will not need to be retrieved “physically” from server to server between the National library and the local libraries anymore. Instead, we imagine that most of the references describing the collections kept by those libraries could be made available by the BnF in a hub building over the technologies developed for data.bnf.fr (which are all open source technologies<sup>9</sup>). Local libraries could then use appropriate APIs to link to the National Library metadata from their local systems rather than retrieving them. In this scenario, local cataloguing could be reduced to a minimum, allowing librarians to change their descriptive tasks by focusing more on linking resources and enriching the metadata with local information rather than describing them.

In order to explore this scenario in a prospective yet realistic fashion, a few months ago the BnF started a use case R&D project together with a local public library near Paris (the Municipal Library of Fresnes<sup>10</sup>) and a software company specialised in LOD technologies (Logilab<sup>11</sup>). This small-scale project aims at exploring and prototyping this scenario in a real life situation, starting from the local library’s functional requirements. This project’s goal, besides starting prototyping solutions, is to ensure that the way the BnF envisions the future of its metadata will ultimately meet practical needs and constraints from the national public

---

<sup>9</sup> In particular, CubicWeb. See : <http://www.cubicweb.org>

<sup>10</sup> <http://bm.fresnes94.fr>

<sup>11</sup> <http://www.logilab.fr>

library network rather than propose a scheme that nobody would use. Communications about this project might also help raising awareness about the new forms of cooperation and the new business models which will need to be envisaged among the librarians as well as the software vendors.

But opening the data also means opening the Library resources to other communities than the librarians themselves. The Library has already heard of, or been approached by several organisations reusing or planning to reuse its metadata for a variety of purposes. The cultural agency of the French Foreign Affairs (l'Institut français) used data.bnf.fr to developed IF Verso, a portal which serves as online registry for all French works translated in foreign languages<sup>12</sup>. An unknown start-up apparently run by a single individual created an iPhone application using the data. Others are planning to use it in order to link metadata about authors (from data.bnf.fr) together with digitized works (some coming from Gallica) and geospatial information (aligned with BnF geographic name authorities). The goal is to develop a new service where tourists and locals from all regions of France could easily read on their smart phone excerpts from books in relation with a monument or a landscape they pass by whenever some famous writers is know to have written something about those places. Those are the very first success stories or plans that can be shared. The BnF hopes that many more will come, as each of these stories helps demonstrating the value of pushing the BnF bibliographic legacy in the LOD.

At the scale and time where national libraries operate, it should however be pointed out that there is no magic in transforming preexisting data. Only a small portion (10 to 20 %) of the BnF General catalogue MARC records is currently made visible and reusable on data.bnf.fr. Much remains to be done not only to FRBRize and display the rest of the legacy records but also to envisage an even more radical and expensive shift on the production side, which is likely to require changing cataloguing code, practices and information system as well.

Additionally, if the Library wants to have a consistent policy towards its users, it would now make sense to consider spreading open licences more broadly so that they apply to all the bibliographic formats and products, whether delivered as UNIMARC records or harvested in Dublin Core through the OAI repositories – and not only in RDF on data.bnf.fr.

### **Conclusion: Lessons learnt and tips to share**

When it comes to lead advocacy efforts in the field of Linked Open Data, what should we keep in mind and what experiences and “tips” may the BnF share with other national libraries? As a conclusion, we propose a short list of suggestions drawn from the Library’s experience in the past months when discussing Open Data issues with decision-makers.

- **Benchmark your national library**

Investigating Open Data issues is a perfect situation to keep a truly open mind on what others do. Because they are usually more directly exposed to discussions and comparisons with external players, management staff tends to be more attentive to models, benchmarks and competitors than to internal arguments.

---

<sup>12</sup> <http://ifverso.com>

International cooperation has proved excessively useful in promoting the data.bnf.fr project. By participating regularly in formal and informal networks and conferences in the field of Library Linked Data, the BnF project team gave international visibility to data.bnf.fr from the very beginning of the project. Such participation also gave first hand information about other ongoing projects often led by prestigious institutions known for their innovative strength. Settling the project and disseminating information about its progress in an international environment highly contributed to demonstrate its value internally.

However, what proved even more efficient from a benchmarking perspective was the national context. With the opening of the data.gouv.fr portal, the BnF project received unexpected visibility and even some attention in the media. All of a sudden, BnF metadata sets were compared and measured with other data from a variety of branches of the public administration, such as public health or transportation. This somehow changed the image of the Library catalogues, showing they could be potentially as useful to the citizens as other valuable public datasets.

- **Make it an economic argument rather than an ideological battle or a technical necessity**

One difficulty in advocating for Open Data is that this movement is often looked at as a naïve, excessively generous trend of opinion very much in line with the values shared by the pioneers and designers of the World Wide Web twenty years ago. Technical arguments, on the other hand, are quite difficult to explain. For instance, explaining why it is virtually impossible to monitor systematically the use of your data once you have exposed it in the Linked Data environment is a challenge when discussing with someone who is not exactly with Web technologies at large.

So, rather than ideological or technical arguments, we found more convincing to explore arguments of an economic nature, which all led to demonstrate that investing in the Open Data could lead to make significant savings over time. For example:

- Data.bnf.fr is not a stand-alone project and investment as it benefits to other applications of the Library by generating more traffic towards catalogues as well as the digital library;
- Data.bnf.fr links BnF resources directly to resources from other libraries and organizations. In the long run, cross-linking these resources will spare the cost of duplicating or retrieving records at all ends.
- Other public administrations can make savings and generate valuable new services by using the metadata from data.bnf.fr. This is obviously true for other French libraries (helping with de-duplication of cataloguing efforts at the national level) but also for other types of administrations and agencies. This means global savings for the State.
- Private companies may use the metadata from data.bnf.fr to develop new products and services, hence using public investments by creating innovation, growth and jobs.

- **Show real use cases and simple facts to explain what you really want to achieve**

Librarians tend to look at change as if they were architects, which they truly are indeed when it comes to build information models and systems: they discuss the blueprints of the basement

before the windows and the view of the house. As a result, in their communication with library stakeholders, they tend to focus too much on the conditions and the cost of change (standards and systems in particular) instead of marketing and promoting the expected benefits from the end user's perspective.

At the BnF, at an earlier stage of the project, the management got scared from hearing too much from librarians about the RDA code, the FRBR model and the need to invest in a new information system while the current system, built 15 years ago, had been a major financial burden.

Building data.bnf.fr as a proof of concept first and then putting it a work, collecting early success stories of reuse and starting the R&D project with a local library have proved most useful in changing the Library management's vision of bibliographical issues and projects. These stories helped illustrating in a simple and tangible way the innovative and strategic value of bibliographic issues with little need for the management to decipher their technical implications.

In short, it seems more efficient to communicate about front end rather than back office issues and more prudent to present things one after the other rather than as a fully integrated new set of standards and systems. Although we know such a comprehensive vision is definitely needed as it will determine the sustainability of change in the long run, we might be more successful with decision-makers when flipping the project upside down: show what the result would look like first then discuss the needs and the implications.

\*  
\*     \*

**Speaker:** Gildas ILLIEN  
Director, bibliographic and digital information department  
National Library of France  
Quai François Mauriac, 75706 Paris Cedex 13  
E-mail: gildas.illien[at]bnf.fr - 33 (0) 1 53 79 59 00

Since 2011, Gildas Illien is director of the bibliographic and digital information department at the Bibliothèque nationale de France (BnF) - the former French National Bibliographic Agency. After 6 years devoted to the implementation of web archiving and digital legal deposit at the BnF, he is now in charge of pushing the Library's catalogues towards the web of data. In the past years, Gildas served as Program Officer and Treasurer of the International Internet Preservation Consortium, organizing cooperation for software development, advocacy and collection building in the field of web archiving among 40 heritage and research institutions over the world. He currently serves as Vice-Chair of the European RDA Interest Group (EURIG) and is willing to drive similar international cooperation in the field of Linked Open Data. A digital curator with academic background in management, political science and sociology, he has published several articles and book chapters about web archiving and digital heritage at large.