

Digitization at the Bibliothèque nationale de France, including an interview with Bruno Delorme

The Bibliothèque nationale de France (BnF) is the national library of France and is located in Paris. It is intended to be the repository of all that is published in France and is a public establishment under the supervision of the Ministry of Culture. Its mission is to build collections – especially copies of works published in France that must, by law, be deposited there – to conserve them, and make them available to the public.

‘Gallica’, the digital library for online users, was established in 1997. In early 2011, the BnF announced plans to digitize and make available for sale online a corpus of 500,000 books within five years. This article provides an overview of the digitization work of the BnF and features an interview with Bruno Delorme, International Sales Director of Jouve Group, the company tasked with this huge digitization undertaking.

Introduction

The Bibliothèque nationale de France (BnF) can trace its origin back to the royal library founded at the Louvre by Charles V in 1368. Expanded under Louis XIV, the library first opened its doors to the public in 1692. With the seizure of the private libraries of numerous aristocrats and members of the clergy during the French Revolution, the library’s collections swelled to over 300,000 volumes. Under the First Republic, the ‘Bibliothèque du Roi’ (‘King’s Library’), as it was then known, was declared to be the property of the nation, ending four centuries of ownership by the monarchy. It was renamed the Bibliothèque Nationale.

Following a number of regime changes, it became known as the Imperial National Library and in 1868 moved to newly constructed buildings on the Rue de Richelieu. By 1896, following further expansion, the library had become the largest repository of books in the world (although it has since been surpassed by other libraries for that title).

On 14 July 1988, the French President, François Mitterrand, announced “the construction and development of one of the largest and most modern libraries in the world ... (which) should

cover all fields of knowledge, be available to all, using the most modern technologies of data transmission, can be accessed remotely and interact with other European libraries”.

It became part of the new generation of major new libraries that were being built around the world, in the United States, Great Britain, Germany, Japan and, of course, in Alexandria. An international jury, chaired by renowned architect I M Pei (architect of the famous glass pyramid at the Louvre), shortlisted four projects, from which, on 21 August 1989, the President finally selected Dominique Perrault to receive the architectural commission.

Construction on the site in the 13th arrondissement of Paris ran into technical difficulties related to its high-rise design and financial overruns, so much so that one journalist referred to it as the ‘TGB’ or ‘Très Grande Bibliothèque’ (i.e. ‘Very Large Library’, a sarcastic allusion to France’s successful high-speed rail system, the TGV). However, eventually the major collections moved from the old Rue de Richelieu site, and the new library was inaugurated on 15 December 1996. It contained more than ten million volumes.

The strategic move to online delivery

The BnF began its digitization activities in the early 1990s, when, recognizing the changes in internet user behaviour, search engines and digitization techniques, as well as the implementation of mass digitization projects (such as the European Digital Library, Francophone digital library, Google Book Search, Open Library, etc.), the strategic decision was made to step up digital activities, in particular via 'Gallica', its digital library for online users, which launched in 1997.

The BnF digitizes a wide range of materials – printed materials (monographs and periodicals), manuscripts, prints, maps and plans, photographs, sound recordings – though, particularly in the context of the European Digital Library, it has three major current focuses:

- **national heritage** (e.g. philosophy, history, literature, religion, science, etc.)
- **international programmes** (including comparative collections such as daily press, legal documentation, history and ancillary sciences such as genealogy, languages and language learning, etc.)
- **a corpus of materials dedicated to all aspects of Europe** (political philosophy, national identities, history, economics, trade, labour, professions, science and technology, social issues, major cultural events, etc.).

From the inception of Gallica, a detailed charter focusing on national heritage was drawn up to precisely define the criteria used to select the documents for digitization. The library also created a specific OAI warehouse – OAI-NUM – which consists of data exported from the library's native catalogues, the BnF Master Catalogue and BnF Archives et Manuscrits. This approach has the advantage of being technically undemanding and of allowing the library to be able to offer other services linked to the digital collections, i.e. the OAI-PMH protocol enables the BnF to publish its own metadata to the wider community and also to harvest metadata from other repositories worldwide, such as the Library of Congress.

The work that Jouve will be doing in conjunction with the BnF sits squarely within this strategic framework. However, the BnF are aware

that any development of online services must include a qualitative evaluation, whether this be the assessment by users of existing services or an exploration of user expectations of new services. Consequently, the BnF regularly carries out usage studies and online questionnaires, interviews individuals and organizes focus groups, with the findings of such studies then documented in highly detailed and instructive reports.

The new digitization project

In a press statement issued on 16 May 2011, the BnF announced that it had signed a framework agreement to digitize copyrighted out-of-print 20th-century French books. The agreement, signed by Frédéric Mitterrand (Ministry for Culture and Communication) and others, including Bruno Racine (President of the BnF), Antoine Gallimard (President of the French Publishers Association) and Jean-Claude Bologne (President of the French Society of Literary Authors), set out an ambitious five-year plan to digitize and make available for sale online around 500,000 books.

The digitization project will centre on the legal deposit collections stored at the BnF, which will retain a digital copy of each title for its own use. Through the Gallica website, enriched bibliographic records will be made available, allowing readers to access excerpts and to be redirected towards online retailers from where they will be able to buy digital copies.

The project has government financial support within the framework of the 'Development of the digital economy' programme. This €4.5bn scheme is one of the main components of the €35bn being mobilized by the government for 'Investments for the future', and includes €750m earmarked for the development of new ways of promoting and digitizing cultural, educational or scientific content. The digitization of the books will enable content to be exploited by means of a common management structure, guaranteeing publishers and authors a fair remuneration in line with intellectual property rights. It has also highlighted the fact that copyright law may need to be modified.

Interview with Bruno Delorme

BRUNO DELORME

International Sales Director
Jouve Group



Speaking to Bruno Delorme, International Sales Director of the Jouve Group, your Editor began by asking about his thoughts on how public and academic libraries are embracing the digital age. He responded: “Most libraries launch digitization projects for preservation purposes only. Very few are ready to invest significant money to build digital products optimized for new devices, such as new generation mobile phones or tablets.”

Bruno went on to say that “All libraries strive to give maximum visibility to their content. For many libraries, working with Google and providing content via their search function has been seen as an interesting option at no cost. However, it is not designed to build digital products ... Companies like Gale/Cengage built programmes to monetize library content and marketed rich products based on selected collections. In this case, the financial risk was covered by a private company, not by the libraries”.

When asked more specifically about the new flagship project with the BnF, Bruno said enthusiastically: “The National Library of France (BnF) is the repository for all that is published in France, and its digitization project is an exception in this market. It is a very ambitious digitization programme aimed at delivering multiple high quality outputs for print on demand (POD), e-book and online databases. They also have a dynamic social media policy and they invest significant time and resource into search engine technologies. This strategy has been successful and drives important traffic to their website. The BnF is also a key contributor to the European library, ‘Europeana’.”

[For more information on the Europeana project, see Jill Cousins’ article in the March 2011 issue of ‘Serials’.]

Looking at the wider context of digitization in

libraries, Bruno was asked why he believes that e-book migration is important, and how he feels that libraries can benefit in both the short and longer-term. “Libraries have to go to digital but not necessarily through e-books. Their content has to be available through different distribution channels and through different formats. As a matter of fact, the improved and cost-effective POD services are a great opportunity. Intuitive image-based apps are another.”

He went on to say “... producing e-books does make sense for some libraries’ collections but, since producing e-books can be expensive, we would advise libraries to have a selective approach. Only valuable and suitable content should be converted to e-book formats. In short, libraries have to build a digitization programme based on their strategic goals. Making the content discoverable requires rich metadata, a social media presence and being listed in major retailer sites (including those specializing in second-hand books). Matching readers’ expectations implies that you can deliver the content in all the formats a reader would expect: paper through POD, through an app on the iPad or similar tablets as well as mobile devices. This drives the way libraries have to go digital to make the most out of all the current trends.”

Having talked about digitization within libraries, your Editor asked what Bruno feels the wider publishing market and retail can learn from the advances in library digitization. “Digitization programmes for libraries require a high level of automation to scale and guarantee the quality required by libraries. Jouve, for instance, has invested a lot in R&D to optimize the image treatments and perfect automated segmentation and structuring of text (for newspaper content, for example). Volumes are also very significant. This know-how and expertise gained from large library

digitization will benefit publishers who plan to go to digital. In France in particular, there is a highly ambitious programme to digitize books which are out of print but still under copyright (the volume is estimated to be between 500k to one million books). This project would benefit directly from the production lines that we built for the BnF and other European libraries."

Your Editors were keen to find out how the Jouve/BnF deal will change the way that the BnF works, especially in terms of what they are able to offer to their customers. "The new BnF programme is very ambitious. The BnF capitalized on what they learnt from the previous programmes to change their digitization specifications. For instance, they understood the opportunity of POD and changed the scanning requirements. As a result, the rendering of digitized books will be much better, making the books available through the main retailer channels providing POD services. The BnF also increased the OCR quality and the percentage of books converted to ePub format. A significant portion of books will now be available on e-reader and tablet devices. The BnF put particular care into metadata to support their curators' work. This will ease the promotion of BnF through search engines and social media. Finally, the BnF opened its

programme to other French libraries. More than 30% of content will come from other libraries who partner with the BnF."

But, projects of this magnitude are not without their problems and can be littered with administrative or technical challenges, so it was interesting to find out what plans Jouve have in place to tackle such a large project. "Almost 300,000 books will be converted by the BnF in this new programme. The system put in place is very workflow intensive. It will help to manage the inventory, which was a major issue. It will also provide detailed reporting on the project status as well as quality tracking and monitoring tools. Since the production process is quite complex (multiple formats, quality requirements, multiple outputs, etc.), increasing production volumes can be a challenge."

Inevitably, another problem faced by such projects is the handling of the materials themselves. Bruno was asked about the selection of books to digitize and whether handling concerns had affected the choice of material, and he replied that "All the precious books will be digitized on site at the library to minimize the logistics. The rare books will be digitized with a limited opening device in order to preserve the binding in dedicated rooms where natural lighting and humidity is suitably controlled".



Careful digitization of ancient books at the BnF



When asked about the metadata standards that are being adopted by the project, Bruno confirmed that Mets and Alto are the preferred formats, before going on to say that "... we have used MARC21 cataloging formats for some other libraries, but it is not part of the BnF's requirements". In response to this comment, your Editor asked whether the BnF is employing other contractors to produce the metadata. "For the metadata, all the objects will be covered by the usual standard of cataloguing procedure of the BnF. The elements coming from the regional libraries, who are partners as part of this new contract, will be indexed at the initial reception stage so it is consistent with those coming from the national collections."

One of the other major administrative problems faced by any digitization project is that of ensuring compliance with copyright and IPR legislation. In the case of the BnF project, Bruno was confident that robust safeguards have been put in place. "The BnF content which is going to be digitized is out of copyright. There is also a strict and rigorous control process carried out by the BnF. In the case where a problem occurs with a book, there is an opt-out clause which will allow the book to be taken out of stock and circulation ..." though he went on to say "... however, this has not yet happened since we started collaborating with the BnF".

In bringing the interview towards a close, your Editor asked about the next steps as the project moves forward. Bruno noted that "The French

National Library will develop up to 12 dedicated partnerships from 2012 to accelerate the electronic availability of its collections. All the digitization work, including the precious objects restored by the library, will take place in France and follow the same specification as the new contract awarded to Jouve in May 2011 ..." before going on to say that "... each project should show how it can be profitable in the mid-term by balancing the data capture and infrastructure costs with the revenue coming from these related electronic assets, with a limited duration of their exclusivity."

Finally, Bruno was asked what, in the face of ever increasing levels of digitization and online provision of information resources, he sees as the future role for libraries. "There is a huge potential for libraries to leverage digital programmes. Having the content digitized in a smart way opens many opportunities in terms of business models. For instance, 'freemium' models could be built online with full access to services through online subscription. Some formats can be delivered for free, and richer, more sophisticated products can be charged for. Libraries shouldn't try to do it alone. It is a great opportunity to build public/private partnerships."

Your Editor thanked Bruno for his time. As a community, we must now wait and see how the work progresses, especially in relation to the more 'commercial' aspect of this project. This may well prove an extremely interesting direction for other libraries around the world.

The DOI for this article is 10.1629/24261. Click here to access original article via DOI:

<http://dx.doi.org/10.1629/24261>

For a link to the full table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=24&issue=3>