

{BnF

# Digital roadmap

March 2016

# Table of Contents

<b>INTRODUCTION</b>	<b>03</b>
The strategic priorities of the BnF	05
A vision for the future	10

<b>01. BUILDING A DIGITAL REFERENCE COLLECTION</b>	<b>14</b>
----------------------------------------------------	-----------

<b>MAIN TOPICS</b>	
01.A_ The web legal deposit	22
01.B_ The digitisation of audio-visual material	23
<b>RESOURCES, TOOLS AND METHODS</b>	
01.C_ The digital collection development strategy	24
01.D_ Digitisation funding strategies	25
01.E_ In-house digitisation studios	26
<b>DEVELOPMENT OPPORTUNITIES</b>	
01.F_ Newspapers: a special case	27
01.G_ The digitisation of copyrighted works	28
01.H_ The legal deposit of e-books	29

<b>02. CATALOGUING AND PRESERVING</b>	<b>30</b>
---------------------------------------	-----------

<b>MAIN TOPICS</b>	
02.A_ Data.bnf.fr	36
02.B_ SPAR: the Scalable Preservation and Archiving Repository	37
<b>RESOURCES, TOOLS AND METHODS</b>	
02.C_ Evolution of catalogues: towards the FRBR model	38
02.D_ E-resources: description and access	39
02.E_ Ensuring long-term access to audio-visual material	40
<b>DEVELOPMENT OPPORTUNITIES</b>	
02.F_ The bibliographic transition	41

<b>03. INNOVATING FOR THE BENEFIT OF USERS</b>	<b>42</b>
------------------------------------------------	-----------

<b>MAIN TOPICS</b>	
03.A_ The www.bnf.fr website	50
03.B_ Gallica content curation targeting different user types	51
03.C_ A digital communication ecosystem	52
03.D_ Online research tools	53
03.E_ Circulation of digital collections for commercial purposes	54
03.F_ Scientific publications	55
<b>RESOURCES, TOOLS AND METHODS</b>	
03.G_ Profiles and expectations of Gallica users	56
03.H_ The digital cultural policy	58
03.I_ Digital accessibility	59
<b>DEVELOPMENT OPPORTUNITIES</b>	
03.J_ Dissemination of Gallica resources and API strategy	60
03.K_ Developing the general public's use of Gallica	61
03.L_ Artistic and cultural education	62
03.M_ Digital lending	63
03.N_ Improved access to digitised newspapers	64
03.O_ Wi-Fi access	65

<b>04. FEEDING INTO A NETWORK OF NATIONAL AND INTERNATIONAL PARTNERS</b>	<b>66</b>
--------------------------------------------------------------------------	-----------

<b>MAIN TOPICS</b>	
04.A_ The national shared digitisation programmes	74
04.B_ The French Union Catalogue (CCFr)	75
04.C_ Third party archiving	76
04.D_ International digital cooperation	77
04.E_ Europeana, the European cultural heritage digital library	78
<b>RESOURCES, TOOLS AND METHODS</b>	
04.F_ Gallica vanilla	79
04.G_ Involvement in international bodies	80
04.H_ Online information for librarians and professionals	81
<b>DEVELOPMENT OPPORTUNITIES</b>	
04.I_ Nationwide access to web archives	82
04.J_ The dissemination of documents digitised by other heritage institutions	83

<b>05. SIMPLIFYING TOOLS AND PROCESSES FOR THE BNF AND ITS PARTNERS</b>	<b>84</b>
-------------------------------------------------------------------------	-----------

<b>MAIN TOPICS</b>	
05.A_ The digitisation workflow: processes and system	92
05.B_ The legal deposit extranet	93
05.C_ The <i>Espace Coopération</i> extranet	94
05.D_ PLATON: Access for people with disabilities	95
<b>RESOURCES, TOOLS AND METHODS</b>	
05.E_ The document management system	96
05.F_ Web archiving tools	97
05.G_ Quality management	98
<b>DEVELOPMENT OPPORTUNITIES</b>	
05.H_ Digital collection management tools	99
05.I_ Ingest of purchased and donated digital documents (ADDN)	100

<b>06. ENGAGING IN SUSTAINABLE INNOVATION</b>	<b>102</b>
-----------------------------------------------	------------

<b>MAIN TOPICS</b>	
06.A_ New e-book formats: EPUB, DAISY	108
06.B_ Research and development in the field of OCR processing	109
<b>RESOURCES, TOOLS AND METHODS</b>	
06.C_ Human resources	110
06.D_ IT investment	112
06.E_ Internal communications and assistance for change	113
06.F_ Professional development	114
<b>DEVELOPMENT OPPORTUNITIES</b>	
06.G_ Text and data mining	115
06.H_ 3D digitisation	116
06.I_ A cooperation project with the French Atomic Energy Commission	117

# Introduction

---

In less than two decades, the digital revolution has caused major upheavals, representing a constantly evolving challenge for our public authorities. In the cultural domain, work on digital development was undertaken very early. Our heritage – memory and collective history – provide fertile ground for innovation, with the digital era giving it a deeper and broader impact than ever before.

The objectives of a national library – to preserve, collect, enrich and transmit – are directly affected by the fundamental transformation of tools, objects and uses in the digital environment. They highlight a broader reflection, relating to the continuity and adaptability of library services for the community in order to promote cultural democracy, equal opportunities and, beyond free access to knowledge and culture, the widest possible dissemination of new digital content, media and tools. This presents a number of challenges in terms of adapting content to users, cultural aspects, diversification of users and accessibility.

---

As the foundations of a new era are created, how do we envisage the future of our national historical and cultural heritage? To what extent, and for what purpose(s), can digital technologies be recognised as a heritage tool and object? How can continuity be ensured over the ages when content, media and scale are constantly evolving? How can traditional and new heritage forms be brought together, when the latter are yet to be defined?

True to its set objectives, the Bibliothèque nationale de France (BnF – France’s National Library) has always remained committed to facing these challenges. In 1988, during the development phase of the new library project, Jacques Attali foresaw that the “totally new kind” of library desired by François Mitterrand would be completely digital. The idea took shape in stages: first digitisation initiatives, creation of the Gallica digital library, start of mass digitisation, setting up of a dedicated organisation etc.

These years of questioning and action were punctuated by several major reports submitted to the ministers of Culture and Communication, on *La numérisation du patrimoine écrit* (*The digitisation of written heritage*, Marc Tessier, 2010), *Le Schéma numérique des Bibliothèques* (*Digital roadmap for libraries*, Bruno Racine, 2010) and *Les Contributions aux politiques culturelles à l’ère numérique* (*Contributions to cultural policies in the digital era*, Pierre Lescure, 2013). Each dealt in its own way with the implications of the digital environment – for the authorities, legislation, the BnF and the public.

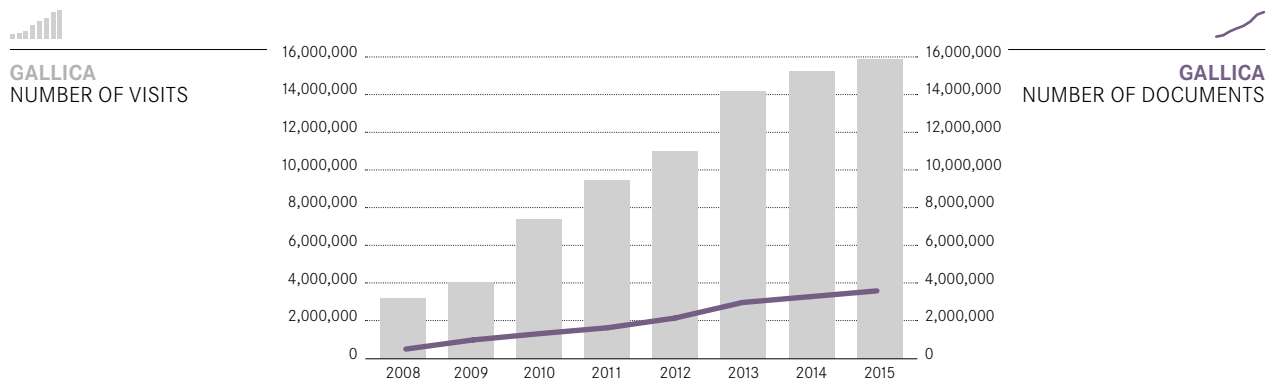
The BnF Digital roadmap is a reference text complemented by a mind map. It encompasses activities and ideas that have supported the development of new heritage policies both in France and abroad.

Libraries entering the digital era are confronted with numerous challenges which the BnF analyses, integrates and, in most cases, anticipates. This practical and forward-looking roadmap sets out the route that the BnF will take over the coming years, in order to provide a service that is in line with current expectations and uses, while anticipating and participating in the development of a constantly changing technological environment.

## The BnF Digital roadmap encompasses activities and ideas that have supported the development of new heritage policies both in France and abroad.

This digital strategy could not be implemented without an ambitious and assertive policy regarding the development, preservation and cataloguing of collections; innovation; user-oriented content management; cooperation and training. The challenge is to maintain the BnF’s ability to preserve a shared heritage, to develop and enrich it and to encourage its transmission, for the benefit of a wide audience.

We have translated and divided these challenges into six strategic priorities that form the basis of its Digital roadmap and that are summarised in this introduction.



## 01

BUILDING A DIGITAL  
REFERENCE COLLECTION

The creation of a digital reference collection covers three main types of operation.

**The first of them is digitising documents currently in a traditional format:** books and journals, video tapes, silver film photos, medals etc. This digitisation, known as “retrospective”, targets documents of a very diverse nature, with a focus on books and periodicals that have entered the public domain. Out of print or not, copyrighted works can also be digitised due to their interest in terms of history and cultural heritage, their scientific relevance or their material fragility.

**The second operation is to collect works already in a digital format:** born-digital documents entering via legal deposit, purchased electronic resources (e-books, journals, databases, audio-visual documents etc.), donations etc.

**The last of these operations consists of archiving the web.** Implemented in August 2006, the web legal deposit plays a leading role in the growth of BnF’s digital collections. National web archiving campaigns are regularly carried out by harvesting sites, pages and files online, through snapshot harvests (domain names ending “.fr”) or more focused crawls (elections, newspapers, other topics).

**All these ways of enriching collections are aimed at combining volume with encyclopaedic knowledge, in the interests of the scientific and educational community, as well as the general public.**

All these ways of enriching collections are aimed at comprehensiveness and, where the abundance of online content implies being selective, at combining volume with encyclopaedic knowledge, in the interests of the scientific and educational community, as well as the general public.

## 02

CATALOGUING  
AND PRESERVING

The creation of a reference heritage collection has two consequences: the long term preservation of documents and the management of their metadata, in other words their ability to be identified, described and found. These activities form the core of the BnF’s historical expertise. In the age of the internet and open data, the Library must develop and sometimes reinvent this expertise by applying it to new objects, new media and new uses.

**The creation of a reference heritage collection has two consequences: the long term preservation of documents and the management of their metadata.**

The BnF has created a sophisticated preservation system, SPAR (Scalable Preservation and Archiving Repository) in order to deal with the constant evolution of data formats and the obsolescence of digital media. SPAR is a digital repository offering the best guarantees of preservation and data security. Certified by the SIAF (the Interministerial Department of the French National Archives), it creates multiple copies of documents and regularly checks their integrity. It also helps to identify technologically obsolete formats, thereby allowing for their transformation.

Guaranteeing the longevity of documents is one thing; making them visible is another, which requires to open up and disseminate metadata - a key challenge. An ambitious project has been underway for several years to ensure the circulation of bibliographical information in the digital environment. Several projects undertaken by BnF teams aim to make library data more visible to web search engines. These projects include: the creation of new cataloguing standards and data models, experimentation with semantic web technologies, convergence of existing catalogues and systems.

Emblematic of this desire to improve access to the BnF’s bibliographic metadata was the launch of the data.bnf.fr service in 2011, under the French state’s open licence since January 2014. The BnF’s web audience has been strengthened, thus bolstering the transmission of French language and culture throughout the world.



## 03 INNOVATING FOR THE BENEFIT OF USERS

The main responsibility of a library, especially when its vocation is universal, is to provide the widest possible audience with sustainable access to a critical mass of creation and knowledge resources.

Cornerstone of the BnF's digital collection, the Gallica platform is the main instrument of this mission. With more than three million documents, it ranks among the world's leading digital libraries and contributes to the online distribution and promotion of a significant part of the BnF's collections. Such an extensive access to resources involves numerous challenges.

### The BnF has implemented a digital hospitality policy for its readers on-site.

The first challenge is to ensure that the considerable amount of data, options and applications doesn't prevent the ease and efficiency of queries. While silence used to be the main barrier to accessing library content, it has now been superseded by excessive noise. Qualifying metadata, simplifying interfaces, optimising the search engine, developing thematic portals and creating online help services are all relevant ways of leveraging opportunities offered by new technologies to improve user experience.

BnF users are by nature diverse, different communities seeking services tailored to their needs: researchers, students, professionals, individuals, those with limited cultural access etc. Editorial programmes in the literary, scientific and historical fields are designed from collections and documents held by the BnF in accordance with this diversity of users. Differentiated access levels are also established, allowing readers of the research library to consult copyrighted documents via Gallica intra muros.

Finally, innovating for users implies taking into account the habits of the digital citizen. The BnF has implemented a digital hospitality policy for its readers on-site: Wi-Fi, self-service photography facilities, services on personal devices including mobile features (seat and document reservation, room capacity information, access to digital resources, online orders of reproductions, personal spaces etc.). In addition, new research practices (social networks, blogs and websites, digital corpora) allow the scientific community to become a stakeholder in its own right of the BnF's collection development policy. Eventually, part of the data could be co-authored with researchers, thanks to the development of content enrichment, correction and linking tools, with a service-oriented approach. More generally, the BnF is increasing the visibility of its services and the circulation of its resources by encouraging the creation of user communities like the *Gallicanautes*, by being highly present in social networks (Facebook, Twitter, Pinterest etc.) and by facilitating the reproduction of its content on other sites.

### Eventually, part of the data could be co-authored with researchers, through the development of content enrichment, correction and linking tools, with a service-oriented approach.

Gallica, Sindbad (the online and telephone reference service) and AVEC (Bring Your Own Device portal) are powerful vehicles for innovation and improvement of the services provided to the public. The richness of their features and the prospect of their future development contribute to greater mobility and personalisation of uses.

## 04 FEEDING INTO A NETWORK OF NATIONAL AND INTERNATIONAL PARTNERS

Due to its history, the richness of its collections, its institutional status and its cooperative mission, the BnF is a driving force in digital policy.

This is true on a national scale, where it feeds into a network of partners including regional libraries, research institutions and cultural organisations. Consisting of a succession of digitisation, cataloguing and curation operations, these cooperative activities respond to a demand to redistribute the financial resources allocated each year by the public authorities – the French National Book Centre firstly – for the digitisation of France’s cultural heritage. In return, these partnerships allow the BnF to enrich the national digital collections and strengthen its visibility.

### The “Gallica vanilla” service illustrates perfectly the knock-on effects of the BnF’s projects in terms of digitisation.

By offering partner institutions the option of having their own digital library (in exchange for the integration of a copy of the partner’s digital resources into the national collections), the “Gallica vanilla” service (*Gallica marque blanche*) illustrates perfectly the knock-on effects of the BnF’s projects in terms of digitisation.

The expertise and creativity of the BnF teams are also shared on an international level. They play a decisive role in the development of Europeana, the digital library of European historical and cultural heritage. They exercise a major influence over international bodies which create standards, establish digital cooperation programmes and develop new tools: *Réseau francophone numérique* (French-speaking digital network), IIPC (International Internet Preservation Consortium), IFLA (International Federation of Library Associations), LIBER (Association of European Research Libraries), the Europeana Foundation, CENL (Conference of European National Librarians) and, recently, the IIIF consortium (International Image Interoperability Framework).

Partnerships have been forged with libraries, universities and cultural organisations over the five continents, with successful outcomes such as shared digitisation programmes, skills transfer, development of thematic portals etc. These partnerships value French expertise in terms of preservation and heritage transmission. They are powerful levers for the reputation of our country and for its culture and language.

### The BnF exercises a major influence over international bodies which create standards, establish digital cooperation programmes and develop new tools.

## 05 SIMPLIFYING TOOLS AND PROCESSES FOR THE BnF AND ITS PARTNERS

Libraries, cultural institutions, professionals from the cultural industries and associations: the BnF is at the crossroads of practices and requirements, with an increased density, intensity and diversity due to digital technology. Therefore, simplifying tools and processes is required to facilitate exchange within this ecosystem.

Via improved sharing services, the BnF is committed to simplifying its relationships with its partners in order to involve them as fullfledged and self-sufficient stakeholders. An *Espace Coopération* was therefore established in 2011. Assisting partner operations for monitoring collections referenced or digitised by the BnF, it also allows for the transfer of digital documents into Gallica and provides information to the beneficiaries of third party archiving. It will soon allow “Gallica vanilla” digital libraries to be configured remotely.

**Via improved sharing services, the BnF is committed to simplifying its relationships with its partners (libraries, cultural institutions, professionals from cultural industries, associations) in order to involve them as fullfledged and self-sufficient stakeholders.**

The legal deposit processes have also been simplified, via the implementation of an online declaration service (“extranet”).

**A significant amount of administrative processes are now fully digital and a document management system (DMS) has been created.**

Bibliographic data provided by publishers is ingested into the BnF information system before enrichment by cataloguers: publishers thereby become co-authors of the metadata.

Implemented in 2010, the adapted publishing service for people with disabilities (PLATON platform) is also testament to the development of partner relationships towards more self-sufficiency and greater sharing: digital files used when publishing a work are deposited by the publisher in the BnF platform, which then plays the role of a trusted third party to ensure the transmission of these files to the approved bodies who have made the request and conduct the appropriate publication (Braille, large print etc.) of these works.

The incentive for simplifying and optimising tools and processes is also valid for the BnF itself, in terms of managing internal exchange and organisation. A significant amount of administrative processes are now fully digital and a document management system (DMS), updated regularly, has been created. The implementation of the system was accompanied by a significant training and adaptation process, to ensure a modernised and more fluid operation of the organisation. Internal management tools and reliable indicators govern the contractual relationship with service providers, especially regarding digitisation outsourcing.



## 06 ENGAGING IN SUSTAINABLE INNOVATION

The BnF would not be able to effectively assist the changes brought about by digital technologies, much less anticipate them, without engaging in innovation and research, monitoring new usages and creating new tools. While leveraging significant resources, both in terms of human workforce and IT developments, the BnF is assisting change and ensuring its capacity to keep evolving in line with the outside world.

### **The BnF is partner to several top laboratories and has signed cooperation agreements with research and higher education institutes.**

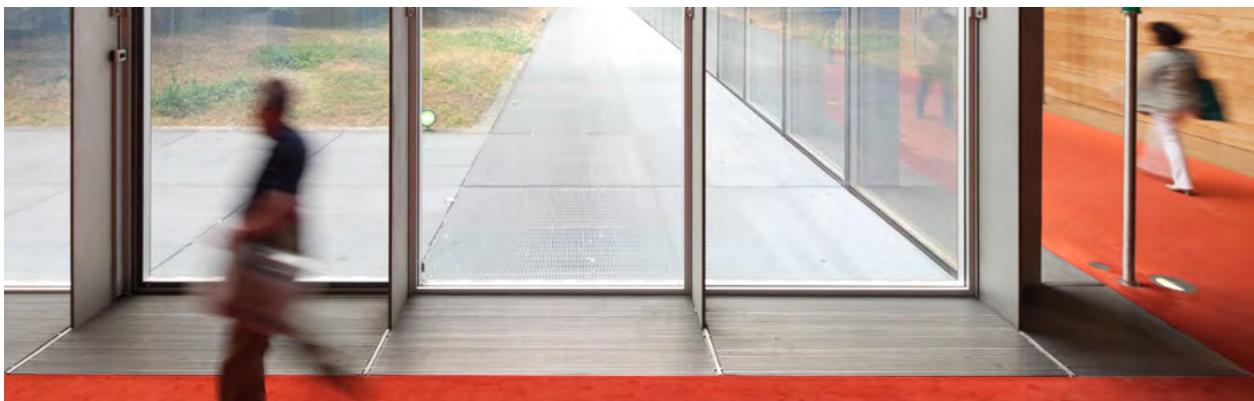
Innovation is part of all the initiatives undertaken by the BnF to concretely improve the preservation and transmission of its heritage. From the point of view of users and their needs, innovation mainly consists of widening the options for research and access to collections, as well as improving the data on an ever growing number of documents. The creation of a digital

lending system and the improvement of access to digitised newspapers could figure among the major innovations introduced by the BnF in the near future.

In terms of research too, the Library registers an intense activity. The BnF is partner to several top laboratories and has signed timely cooperation agreements with research and higher education institutes. These partnerships explore topics like the use of digital heritage, text and data mining and the exploration of targeted collections, thereby contributing to the development of digital humanities.

The BnF's engineers and technicians are also conducting major technical experiments, most often in partnership with public and private bodies. This work targets the development of new production systems for digital information, for instance the crowdsourcing of content production and the improvement of optical character recognition (OCR).

Thanks to these developments, the BnF will better meet users' needs in the future. These programmes also stand as a significant contribution from the BnF teams and their research and cooperation partners to the showcasing of French excellence in terms of digital innovation.



# A vision for the future

---

The cartography of digital activities at the BnF, as described in this document, demonstrates strengths, methods and development opportunities for each strategic priority, thus outlining a medium-term vision. If the work on innovation and change were to continue at the same pace and with approximately the same resources, what would the digital BnF look like within five years? What milestones would be achieved and which hurdles surmounted? Without wishing to pre-empt the actions and objectives due to be defined by the new performance contract from 2017, suggested below are some of the areas that may form the outline of the BnF in 2020.

---

## The BnF, home of digital hospitality

The BnF user, both on the François-Mitterrand site and at Richelieu, will benefit from attractive tools: clearly highlighted exclusive digital resources, an updated website to better answer his or her needs, and services that are accessible and usable on the go. Users will be offered personalised services, from Wi-Fi access to searching through digital corpora. The BnF will also be able to impress its visitors, especially at Richelieu, by offering them an enhanced heritage experience, a renewed source of feelings and emotions. The BnF will therefore use digital technologies as a lever to increase on-site visits. Online users will be invited on new virtual tours covering the history of the BnF's sites and collections: a digital cultural transmission

strategy likely to arouse the curiosity of young users and strengthen the BnF's relationships in the educational domain and more generally in the social domain.

## Gallica, a source of culture and a melting pot for all types of users

Researchers, students and amateurs as well as genealogists, enthusiasts, children and families: the *Gallicanautes* will be more numerous and their uses more diversified thanks to a content adaptation strategy aimed at the general public, both on the BnF sites and social networks and via mobile applications. The BnF will strengthen its role as an organisation that explores new ways of ensuring cultural access for all, by inventing new services: digital lending, crowdsourcing platforms (Gallica studio).

## Gallica and Gallica intra muros, melting pots for all types of expertise

The ever growing wealth of content offered by Gallica and the quality of its digital collections will contribute to the appeal of the digital library. The most qualitative digitisations from an intellectual and technical point of view will be displayed on Gallica, where user experience will be improved by providing advanced browsing tools. Extracts from copyrighted documents will also be available. On Gallica intra muros, readers inside the BnF's reading rooms will access the entirety of the digitised heritage collection: over five million documents thanks to the continuing national digitisation effort. Some 500,000 additional newspaper pages and 200,000 out-of-commerce books from the twentieth century, digitised as part of the ReLIRE project, will be added to several tens of thousands of e-books ingested by legal deposit, an increasing amount of digitised special collections, digitisations carried out in partnership with private stakeholders and digital collections from partners of the national cooperation network.

## Digital heritage, continued

The BnF will be fully adapted to the new transmission methods of cultural assets and other digital content intended to become part of its collections, both from a legal and technical standpoint. The legal deposit of electronic publications, including audio material, videos and other web and multimedia documents which are currently inaccessible, will be developed. The Library's heritage collection will thus undergo a decisive shift, reflecting the development of society. Thanks to secure tools and highly efficient workflows, the BnF will be identified as a trusted partner by stakeholders from the publishing chain, allowing for fluid data and content exchange. Donations and purchases of digital documents of all forms and types will extend the Library's traditional collections and enable the exploration of new forms of knowledge.

## The excellence of digital preservation

The BnF will continue to develop its digital preservation infrastructure and will integrate all of its workflows into SPAR, the Scalable Preservation and Archiving Repository. All the files resulting from digitisation, web archiving and digital legal deposit, as well as the BnF's administrative records, will be preserved over two sites. A third site, secured to face the most acute risks, will be sought out, in order to accommodate at least the most sensitive collections. The merger with the system dedicated to preserving audio-visual material will be started.

Finally, the BnF will continue to share its expertise and its infrastructure in this area with other public stakeholders, notably through third party archiving.

## Towards the appropriation of big data

The BnF will capitalise on the experience of data.bnf.fr to significantly develop its big data production methods. This work will be more collaborative: the production of authority files and data will occur at a national level, at least in part, thanks to shared tools and standards like ISNI, which brings the community of archives, museums, libraries, publishing and higher education and research closer together. Algorithms, data flows and, more generally, big data tools will be developed for the sake of efficiency, leveraging opportunities for the use of data other than bibliographic (user logs, management data etc.) for the development of the Library's tools and services.

## With the legal deposit of electronic publications and audio material, the Library's heritage collection will undergo a decisive shift, reflecting the development of society.

## The value of open data

The BnF's expertise through the development of data.bnf.fr will also contribute to the visibility of its resources and increase its audience. In the general catalogue, as well as on Gallica, online users and readers will have direct access to the works, people, places and concepts that they are looking for, thanks to semantic clusters and visualisation experiments previously tested on data.bnf.fr. The service will be open to even more external data, both to help promote partners and to better respond to users. The proliferation of links to and from the Library, via platforms like Wikipedia and Wikidata, will reinforce the BnF's position in the web ecosystem. The BnF's bibliographic data being legally open since 2014, then technically open via APIs such as the IIIF and other dissemination mechanisms, third party reuse of BnF data will thrive and contribute to its image and reputation. This impetus will also benefit partners, thanks to a similar initiative for accessing data in the French Union Catalogue (CCFr) and the aggregator role played by the BnF on behalf of the Europeana digital library.

## Gallica, a shared infrastructure

Gallica will become more than just the digital library of the BnF and its partners. The digital library will become the core of a shared infrastructure for an ever bigger network of public stakeholders, both in France and abroad, especially in the French-speaking world. The BnF will contribute to the digital influence of its partners through different initiatives: the Gallica vanilla system, improved for greater efficiency, and other digital cooperation tools like the “shared heritage” curation portals, developed with international partners. Collections digitised by partner institutions that are still lacking an access system will become visible to the public. Through these projects, the BnF will share with its network of partners both its expertise and the investment accorded over many years, thus helping to increase the number of collections available online and enabling global access to knowledge.

## The management of professional tools

The BnF's information systems will continue evolving to better support the digital transition. Experts from the BnF will be equipped with monitoring and management tools for better management of both physical and digital collections, ensuring the necessary continuum between different media for preservation as well as access. Via its extranets, the BnF will extend this

management to its partners who will be able to act independently to ingest and access workflows for their documents and meta-data. The improvement of internal processes will benefit BnF readers, via measures like emergency digitisation in internal studios, as much as it will benefit staff, who will be better able to manage their digital collections, including authorisations for dissemination. Finally, publishers and other book chain professionals will become strong advocates for the dissemination of the BnF's iconographic resources, through the new pictures collection (*Banque d'images*).

## The BnF, innovation laboratory

The BnF will continue to innovate, acting as a driving force for change and inventing new ways of conducting its missions. In the wake of its collaboration with the Ideas Laboratory of the CEA, it will set up a dedicated organisation, exploratory measures and appropriate methodologies in order to encourage the emergence of new tools and services. Already a major stakeholder in the academic field, especially in the digital humanities, the BnF will also open itself up to the experience of other professions, including start-up companies, by benefiting from its proximity to the Halle Freyssinet. New ways of assisting change and collaborative working methods will enable the Library's staff to adopt these new developments and contribute to designing and building them.

---

In addition to facilitating its main missions of collecting, preserving and cataloguing, the digital transition empowers the BnF to encourage access to its collections, but also their circulation and promotion, to support scientific research and even to strengthen cooperation between cultural institutions. This document attempts to show how the BnF has been able to seize the opportunities presented by the digital revolution.

It is through meeting these new challenges that the BnF, loyal to its historic mission, continues to assume the responsibilities involved in transmitting our heritage.

01

BUILDING A DIGITAL  
REFERENCE  
COLLECTION - P14

02

CATALOGUING  
AND PRESERVING - P30

03

INNOVATING  
FOR THE BENEFIT  
OF USERS - P42

04

FEEDING INTO  
A NETWORK OF NATIONAL  
AND INTERNATIONAL  
PARTNERS - P66

05

SIMPLIFYING TOOLS  
AND PROCESSES  
FOR THE BNF  
AND ITS PARTNERS - P84

06

ENGAGING  
IN SUSTAINABLE  
INNOVATION - P102



{ BnF





01

# BUILDING A DIGITAL REFERENCE COLLECTION

---

**3.5 million**

out of copyright documents online

---

**1 million**

copyrighted documents digitised  
including more than 700,000 audio-visual documents

---

**1,519 terabytes**

of documents on Gallica

---

**668 terabytes**

of web archives

---

**305**

digitisation partners

---

**15.7 million**

digitised images in 2015

---

The dissemination of knowledge and the democratisation of culture in a digital world lead to the expectation that every publication, past and present, should be readily available to the user, in one click. For a national library, this challenge translates into the creation of a digital heritage reference collection that seamlessly integrates born-digital documents with early-printed books and manuscripts, in order to facilitate access to these resources for everyone, today and in the future.

---

---

#### **MAIN TOPICS**

- 01.A\_** The web legal deposit
- 01.B\_** The digitisation of audio-visual material

---

#### **RESOURCES, TOOLS AND METHODS**

- 01.C\_** The digital collection development strategy
- 01.D\_** Digitisation funding strategies
- 01.E\_** In-house digitisation studios

---

#### **DEVELOPMENT OPPORTUNITIES**

- 01.F\_** Newspapers: a special case
- 01.G\_** The digitisation of copyrighted works
- 01.H\_** The legal deposit of e-books

# The digitisation of collections

The digitisation of written, audio-visual and graphic heritage aims to progressively cover all periods, documents types and disciplines, and to grant access to them via Gallica. Nowadays, thanks to the mass digitisation started in the middle of the last decade, 3.5 million documents are online on Gallica. The digitised collection grows regularly at a rate of several million pages or images per year, all sectors combined – printed works, special collections etc.

This work has been carried out in three major stages, in line with technical developments and collection development policies:

**2002-2005:** Creation of a **Library of the *Honnête Homme***, encompassing the most significant works, discipline by discipline, from Antiquity to our times. The digitisation was done mainly as images.

**2005-2009:** Emergence of **mass digitisation**, due to the Google initiative. Since 2007, optical character recognition (OCR) has become systematic and is applied retrospectively to those collections already digitised. In contrast to the previous period, mass digitisation was carried out via de-selection: once the collection is selected, documents that are considered out of

scope for digitisation (documents still in copyright, foreign documents and those too fragile to undergo the digitisation operations) are removed. History of France (26%) and literature (25%) are the most represented disciplines quantitatively.

**Since 2009: Digitisation has become a collective effort, diversified** in its content and funding. Mass digitisation con-

tinues, all the while striving to balance objectives in terms of quantity and selectivity; initiated in 1998, the digitisation of audio-visual material has appeared on Gallica.

Along these steps, priority was given to unique documents: those preserved in no other institution. The BnF has relied increasingly on partnerships with French and foreign institutions and opened its digitisation workflows to collections from other institutions.

Digitisation is no longer limited to works in the public domain, which excluded publications from much of the 20<sup>th</sup> century: the ReLIRE project (Register of out-of-commerce books republished in digital form) allows for the digitisation of out-of-commerce works published before 2001, while managing intellectual property rights.

**The digitised  
collection grows  
regularly at a ratio  
of several million  
pages and images  
per year, all sectors  
combined.**



# Legal deposit of born-digital media

The BnF's digital collections must, in addition, make the link between the retrospective digitisation of heritage and the future of published media, which is increasingly available in digital form.

As part of its legal deposit mission, the BnF receives and preserves sound recordings, videos, multimedia documents and electronic documents. The first audio-visual material collected in a digital form was in the 1980s, taking advantage of the commercialisation of software discs and tapes, then optical media. By communicating audio-visual material on digital video servers from 1996, the BnF was a leader in experimenting access to digital material. With the quick development of new publishing and circulation channels for online digital material, the BnF had to adapt its ingest processes. It currently receives digital files in varying amounts depending on the types of media, sometimes they can amount to up to 40% of the deposits. In spring 2016, born-digital audio-visual material ingested as legal deposit therefore represented over 25,000 documents. The volume of this collection is expected to increase quite considerably in the coming years,

**The BnF has been able to create, via web archiving, a collection of 668 terabytes of data.**

taking advantage of the implementation of an automated input workflow for audiovisual material, starting with audio material.

In 2006, the French law relating to copyright and related rights in the information society, known as the DADVSI law, extended the legal deposit to “signs, signals, written documents, images, sounds or messages of any kind that are communicated to the public electronically”. The BnF has therefore been able to create, via web archiving, a collection of 668 terabytes of data, which, for example, allows researchers to analyse how the internet is used in the development of militant movements, or to relive the Tunisian revolution of 2010-2011 by looking into its impact via social networks.

In 2016, the legal deposit domain is extending to e-books, whether they are homothetic to printed works already deposited or only available in digital form. After an initial experiment, a complete processing workflow will allow for the ingest of several thousands of titles into the collections from the first year.

## E-resources acquisition

This new heritage must finally be understood in the wider context of the development of knowledge and digital services. Since 1995, the BnF is committing to an acquisition policy for three types of born-digital resources:

- **databases:** reference works (dictionaries, encyclopaedias, indexes and directories), corpora of texts, specialised portals (164 subscriptions were acquired in 2015, representing 204 database titles available to readers);

- **journals:** nearly 1,600 current titles are available. Accessible online, they are combined, or not, with a paper subscription. Exclusively digital subscriptions are mainly for scientific and technical journals;

- **e-books:** 91,900 titles were acquired in bundles from major publishers (Springer, Elsevier, Wiley, Royal Society of Chemistry). Simultaneously, the BnF also purchases individual e-books with permanent access (2,120 titles since 2012).

Faced with the rapid growth of e-resources, a national digital acquisitions policy was developed in France in order to purchase national licences, which the BnF benefits from for journal archives. Moreover, at the end of 2013, the BnF joined the University Consortium of Digital Publications (Couperin). It can therefore take part in national negotiations relating to the purchase of new resources and give its readers remote access to the resources acquired.

## An encyclopaedic collection

The BnF's digital collection development policy is complex, due to its numerous criteria: selection principles, technical constraints linked to the media and the nature of collections, funding strategies, historic sedimentation of digitisation policies, partnership opportunities, emergence of new knowledge transmission methods etc. It is nevertheless covered by a comprehensive project, which defines the concept of comprehensiveness as "relative" and "reasoned", with the aim of offering today's reader and tomorrow's researcher as comprehensive a vision as possible of national heritage. The excellence of this policy lies in covering all the disciplines and serving current uses as much as future needs that cannot be anticipated. It relies on collaborative partnerships, with a national and international objective: to increasingly enrich digital collections in order to circulate them as widely as possible.

**The aim  
is to offer today's  
reader and tomorrow's  
researcher  
as comprehensive  
a vision as possible  
of national heritage.**





{BnF

entree



# 01.A

## The web legal deposit

Established by the French law of 1<sup>st</sup> August 2006, the web legal deposit plays an important part in the growth of the BnF's digital collections. The entirety of online French publications is supposed to be collected by the depository institutions (the INA – French national Audio-visual Institute – for radio and television websites, the BnF for all other French sites). The mass of documents on the internet and their continual update prevents comprehensiveness; the notion of representativeness determines the harvest scope, with the aim of offering future generations an image of French online culture over the ages.

To build representative collections, the BnF has defined an integrated model which combines snapshot harvests (of all the French websites, once per year) and focused crawls of websites selected by the Library staff or by partners. The plan was developed, extended and shared (with libraries responsible for legal deposit by the printer especially), in order to more effectively fulfil its objectives and to better serve the interests of collection users, researchers, amateurs and other interested parties.

### THE QUESTION OF SELECTION

The coverage of snapshot harvests is still insufficient, in spite of the extended adoption of automatic identifying methods and cooperation with other

national stakeholders. The BnF cooperates with the AFNIC (the .fr registry) and complements the .fr domain list with other websites which are automatically identified as French (.com, .org, .net): the BnF is therefore able to identify and archive 4 million sites.

The focused crawls, concerning 20,000 sites to date, can be enriched with a more widely open selection of the BnF's partners, like the National and University Library of Strasbourg, which selects its local heritage on the web, the "Alsatiques" (see part 4 section I). Such an approach led to a study launched with the Interministerial Department of the French National Archives (SIAF) about the harvesting of French state and local authorities websites. Moreover, a national selection network has been created, complementing the network of web legal deposit curators already present within the BnF.

### THE QUESTION OF PRESERVATION

The preservation of heterogeneous and complex content is crucial. The most recent collections are subject to transfer and preservation in SPAR, the digital repository of the BnF. The early collections are about to be ingested into SPAR, they are the true "early printed books" of the web at the end of the 1990s and the start of the new millennium, and require optimal preservation conditions.

**The notion of representativeness determines the harvest scope, with the aim of offering future generations an image of French online culture over the ages.**

## The regional daily newspapers

The regional daily newspapers represent an essential source of local, national and international history. 25 are now harvested by the web legal deposit and correspond to over 200 local publications, including *Les Dernières nouvelles d'Alsace*,

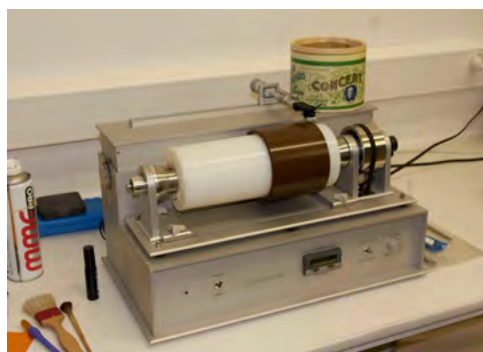
*La Dépêche du Midi* and even the online daily *Médiapart*. The newspapers harvested by the web legal deposit servers are most often made up of PDF files. Quality and access control are ensured, day to day, by the same staff as for the printed versions.



# 01.B

## The digitisation of audio-visual material

Encompassing over 1.5 million documents, the BnF's collection of audio-visual material is extended, every year, by approximately 40,000 documents via the legal deposit, but also thanks to donations and acquisitions. In order to ensure its long-term conservation and communication missions, the BnF undertook, in 2000, a digitisation plan known as "safeguarding" for its audio, audio-visual and multimedia heritage collections – documents threatened by the fragility and natural deterioration of their media as well as the material and software obsolescence of reading equipment. This systematic and wide-scale operation concerns, above all, the digitisation of the most fragile media: the content of instantaneous discs ("Pyralis"), wax cylinders, magnetic (tapes, cassettes and floppy discs) and optical (audio CD, DVD, CD-ROMs) media is transferred onto standardised and durable formats. Analogue-audio material is digitised in high definition, without compression or correction, in order to conserve all the information contained on their media. Born-digital material is copied bit by bit identically. 'Mass' digitisation of standard formats is carried out by external service providers. Such providers also process rarer media, which require rendering devices that are more difficult to find (magnetic wires, Tefi tapes, 1 or 2 inch video tapes etc.).



**Digitisation of cylinders from the Prince's Palace of Monaco with the Archéophone phonograph.**

### AUDIO-VISUAL DIGITISATION STUDIOS

The BnF has four digitisation studios for audio material (François-Mitterrand and Bussy sites) and two for video material (Bussy). They deal with documents that cannot be subjected to mass processing: instantaneous discs, cylinders. Lastly, an internal branch allows for the digitisation of DVD media.

These studios, in part, ensure the restoration of the signal. A videogram restoration workflow was implemented in 2014 in order to process endangered documents.

### AUDIO-VISUAL EXPERTISE

The expertise of the audio-visual department is not only used to process the BnF's own collections, or collections that have been deposited at the Library, but also sometimes external collections processed on-site in the context of service provision (digitisation of cylinders from the Prince's Palace of Monaco, for example). Over the years, this has given the BnF a high level of expertise in the digitisation and restoration of videograms and phonograms.

## A partnership for the digitisation of LPs

A project led by BnF-Partenariats since 2012 has enabled the digitisation of 46,000 LPs (from 1949 to 1962 inclusive), or more than 300,000 tracks across all genres (classical, jazz, pop, world music etc.). LPs were digitised in 2014 via the Memnon Archiving services technical partnership. Once their metadata had been enriched, more than 36,000 albums were put onto online music distribution platforms

(downloads and streaming, iTunes, Qobuz, Deezer, Spotify etc.) by Believe Digital. Snippets are accessible on Gallica and full content on Gallica intra muros inside the BnF's reading rooms.

The first revenue generated by the sale of LPs will allow for the digitisation of 1,000 78 RPM records of French songs from the 1930s cinema.

**Over the years, the BnF has developed a high level of expertise in the digitisation and restoration of videograms and phonograms.**



# 01.C

## The digital collection development strategy

The BnF strives to support a multimedia-oriented digitisation, reflecting the diversity of the heritage that it protects: manuscripts, books, periodicals and newspapers, prints, drawings, photographs, maps, sheet music, currencies, audio material and moving images therefore find their place on Gallica.

There are several methods of digitisation: a comprehensive method where the entirety of a collection is digitised; a selective method that gathers thematic corpora and intellectual collections; and an on-demand, piece-by-piece method on a limited scale.

### THE DIGITISATION OF BOOKS, JOURNALS AND MAGAZINES

Since 2007, digitisation of printed material focuses on French and regional language documents which are no longer under copyright and were printed on French soil. The quality of the scanning equipment has been constantly improving: the selection options have thus expanded. Digitisation campaigns now extend to works from the 17<sup>th</sup> and 18<sup>th</sup> centuries, as well as to valuable and fragile items. At the same time, the creation of two collections has continued: the digitisation of major texts (selection of a collection of approximately 10,000 editions of emblematic works classified by theme and by century) and the digitisation of the most consulted documents,

representing a collection of 5,000 documents by topic (history, philosophy and humanities, law, economics and politics, science and technology, literature and art). Finally, the digitisation of newspapers remains a priority.

### THE SPECIAL COLLECTION DIGITISATION PROGRAMMES

Although they focus on the reproduction of “treasures and masterpieces” (prints by Pierre Roche, Berthe Morisot or even Vincent Van Gogh, documents from the Marcel Proust collection at the manuscripts department, the Rameau programme in the music department and even the digitisation of medieval illuminated manuscripts from the reserve of the Arsenal Library etc.), these programmes also extend to collections of documents selected for their documentary interest, frequently in partnership with research programmes. These include the digitisation of currencies, which has already enabled a quarter of the 700,000 currencies held by the BnF to be put online. Specific work has been undertaken in accordance with the partnerships and agreements made by the BnF in very diverse areas, like the 3D digitisation of 55 globes from the maps and plans department by the DNP company (Dai Nippon Printing), in the context of skill-based patronage (see part 6 section H).

### As part of its 2014-2016 performance contract, the BnF has decided to formalise its digitisation development policy in a charter.

As part of its 2014-2016 performance contract, the BnF has decided to formalise its digitisation development policy in a charter, in order to communicate it more widely to user communities and partners in France and abroad.

## Three main founding principles

The principles which govern the decision to digitise a document relate to three areas:

- **Heritage interest:** this applies to all the documents which can be considered as masterpieces or treasures, given their aesthetic quality, historical value, origin or scarcity;
- **Documentary interest:** endorsed by research practices, this characterises highly requested collections;
- **Need for conservation:** this drives the digitisation of fragile documents and those in a state that makes them unfit for viewing: acidified documents (in particular newspapers), large formats, glass plates, objects with mobile items etc. In terms of conservation, audio-visual materials are a special case, since magnetic, optical and mechanical media are fragile and at risk of accelerated degradation.

*Spider* (print), by Odilon Redon



# 01.D

## Digitisation funding strategies

Several funding strategies have been implemented, in order to give the BnF a capacity for action of an unrivalled breadth regarding heritage digitisation – even if, for certain collections like newspapers, there are still needs to be met.

### THE BnF'S OWN RESOURCES

(from 1997)

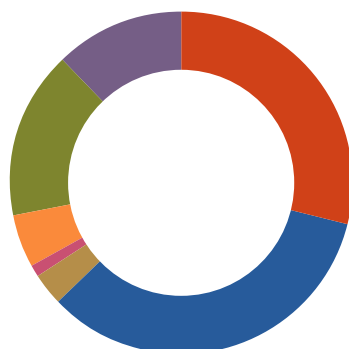
17 million Euros were allocated to digitisation by the BnF from 2002 to 2012, including 3.2 million Euros for agreed libraries and partners.

### GRANTS FROM THE FRENCH NATIONAL BOOK CENTRE (CNL)

(from 2007, for printed material)

46.7 million Euros were received from 2007 to 2014 (an annual average of nearly 6 million Euros) to finance the digitisation of printed material, digital preservation in SPAR, dissemination on Gallica and, recently, the ReLIRE project.

**DIVISION OF DIGITISED IMAGES AND ITEMS PRODUCED - 2015**



- CNL - Printed material 29%
- CNL - Out-of-commerce books 34%
- CNL - Rare books 3%
- Newspapers 1%
- Special collections 5%
- In-house studios 16%
- BnF-Partenariats 12%

### CORPORATE SPONSORSHIP

Several projects have benefited from funding since 2009: for example 150,000 Euros for the digitisation of 350 Arabic, Turkish and Persian manuscripts supported by the Total Foundation in 2009, and the digitisation and 3D reproduction, in very high definition, of a collection of terrestrial and celestial globes dating from the 11<sup>th</sup> to the 19<sup>th</sup> century, supported by skills-based patronage from Dai Nippon Printing in 2015 (see part 6 section H).

### INDIVIDUAL SPONSORSHIP

The “Adopt a book” operation was initiated and launched by the Association des Amis de la BnF (Friends of the BnF) in 2011. By the end of 2015, more than 300 works had been digitised using this method.

### PROJECTS

Since 2009, collaborative projects have seen the light of day at national and international levels, financed by private European funds, or linked to the Grand Emprunt (Investments for the Future programme). These projects are a chance to associate French libraries with the creation of digital collections, sometimes with the purpose of uniting former libraries that are now dispersed (examples include the International Dunhuang Project, Le Roman de la Rose, BHL-Europe, Europeana Regia, LABEX OBVIL and the digitisation of Christine de Pisan’s manuscripts).

### THE MOBILISATION OF “INVESTMENTS FOR THE FUTURE”

In order to access funding from the French Commissariat-General for Investment (CGI), the Library has created a subsidiary, BnF-Partenariats, which now pilots public-private partnerships on digitisation projects for specific collections. Moreover, funding for the digitisation of out-of-commerce works from the 20<sup>th</sup> century, is planned from three parties – the CNL, the CGI and publishers (notably the Cercle de la Librairie).

### REPRODUCTION SERVICES

These contribute to the digitisation of collections: a million images are produced every year from 13,000 individual and professional orders. In 2015, the earnings linked to reproduction work rose to 1,104,179 Euros for the reproduction operations, 541,683 Euros for commercial use fees and 21,402 Euros for partnerships, corresponding to a total of 1,667,265 Euros including tax and VAT.

## BnF-Partenariats

This subsidiary, created in 2012, has led to the creation of several partnership agreements for digitisation:

- with ProQuest, in order to digitise approximately 28,000 old printed books from 1470 to 1700 (or 10 million pages);
- with Believe Digital and Memnon Archiving Services, to digitise and restore the audio signal of nearly 46,000 LPs;
- with Immanens, to speed up the digitisation and online circulation of written newspaper collections, since their creation.

# 01.E

## In-house digitisation studios

Three in-house reproduction studios, located at the Bussy-Saint-Georges, Sablé-sur-Sarthe and Paris (François-Mitterrand) sites are dedicated to conservation reproduction. Launched in 2005, their gradual transition from silver (microfilm and microfiche) towards digital is now almost complete. A fleet of over 25 scanners allows for the processing of a wide range of documents (texts, images, objects), from different collection departments that cannot be covered by the digitisation markets: large formats (Sablé-sur-Sarthe), very fragile and valuable documents (Paris and Sablé-sur-Sarthe), fragile newspapers (Bussy), objects (mobile studio) and the process known as “*urgent lecteur numérique*” (“emergency on-demand digitisation”) at the Paris site.

In addition to these studios are four digitisation studios for audio (François-Mitterrand and Bussy sites) and two digitisation studios for video (Bussy).

Each site has therefore developed a technical specialisation or area of excellence without restricting other activities. The quality of the work provided by the studios stems from the implementation of synergies with those responsible for the collections, the restoration and physical preparation workshops and even with the cataloguers. The versatility of the equipment must also be highlighted, which offers numerous processing solutions.

At the François-Mitterrand and Richelieu sites, two workshops for digitisation and printing on-demand annually process approximately 13,000 orders from customers wanting to acquire a reproduction for a professional, personal or research project (see part 3 section E).

**Each site has developed a technical specialisation or area of excellence.**

**Recourse to machines that enable digitisation via partial exposure – here, Qidenus**



# 01.F

## Newspapers: a special case

A vital resource for research, teaching and information, newspapers, originally designed for temporary use, raise specific conservation and safeguarding issues. The paper and ink, of a poor quality, accelerate the deterioration of the document: wood pulp, a widely used material, makes the paper very acidic. The reproduction of newspapers is therefore crucial.

Successive digitisation campaigns have made several dozen titles partially accessible on Gallica: major national dailies plus specialist and local newspapers. With more than 5 million pages of digitised newspapers since 2004 (including nearly 4 million processed via OCR), the BnF provides one of Europe's most significant online newspapers collection.

### AN UNEXPLORED DEPOSIT

The considerable rise of newspapers in France during the mid-19<sup>th</sup> century means that reproduction needs exceed available resources. Only 10% of titles have been microfilmed to date and less than 3% digitised, while only 2,500 of the 110,000 publications held by the BnF, representing at least 100 million pages, are currently available online.

### CONTENT SELECTION BASED ON USE

Content selection takes into account the viewing statistics of newspapers on-site and the suggestions of individual users and academics.

There is also a focus on collections in the public domain (printed before 1945), in order to display them immediately on Gallica. An opening up of the subsequent period is also to be encouraged, in partnership with the publishers and rights holders.

The innovative digital file enrichment techniques (article segmentation, named-entity recognition, headings, title correction) are subject to a second selection level, to retain those types of newspaper that would most benefit from them.



The BnF provides one of Europe's most significant online newspapers collection.

## Regional newspapers

In 2013, the BnF undertook a survey into the digitisation of regional newspapers. Nearly one third of France's major public libraries and local authority archives indicated that they had digitised nearly 5.5 million pages, or approximately 1,000 titles. The funding comes, for the most part, from the French National Digitisation Plan,

sometimes with an extra contribution from the BnF's agreed regional partners (indexation of digitised newspapers in Languedoc-Roussillon).

The [presselocaleancienne.bnf.fr](http://presselocaleancienne.bnf.fr) site, launched in 2016, consists of an enriched database of digital links to titles of early printed, local and regional newspapers, digitised in France.



# 01.G

## The digitisation of copyrighted works

In the context of national digital cooperation, various documents still under copyright can be accessed on Gallica, thanks to rights negotiation or specific digitisation and circulation agreements.

### SOME EXAMPLES

Several important current journals in the area of humanities and social sciences (*Études*, *Revue française de psychanalyse*, *Revue historique*, *Revue d'histoire littéraire de la France*, *Revue d'histoire moderne et contemporaine*, *Le mouvement social*, *Le Moyen Âge*, *Hérodote*, *XVIII<sup>e</sup> siècle*) and 100 publications from local and regional learned societies, academies, national and regional societies (science, medicine, agriculture etc.) have granted the BnF permission to display their content on Gallica free of charge. Gallica also gives access to government publications from the BnF's collections (official journals, county council meeting minutes etc.) and from partners (parliamentary debates from the National Assembly and the Senate, INSEE, Banque de France, Observatoire de Paris, the City of Paris, Cédias-Musée social, Institut catholique de Paris etc.).

### MAKING UNPRESENTABLE ITEMS ACCESSIBLE

Thanks to the digitisation of copyrighted works, material declared unusable for conservation reasons has become accessible again via Gallica intra muros, in the reading rooms of the research library.

The selection of items to be digitised is usually part of a preservation plan relying on systematic verifications of the conservation state of stored material. A priority digitisation can also be launched via the “*urgent lecteur*” process – a dedicated

workflow which provides a substitute digital version of an unpresentable item within at most one month of the reader's request. The substitute is then made available online, on Gallica or Gallica intra muros. 130,000 works still in copyright have been digitised by the BnF using this process.

**The selection of items to be digitised is part of a conservation plan.**



### ReLIRE

The ReLIRE project is the result of the French law of March 1<sup>st</sup> 2012 relating to the digital use of out-of-commerce books from the 20<sup>th</sup> century. Its objective is to publish in a digital format books that have not yet entered the public domain.

The selection of books nominated to appear on the ReLIRE registry is ensured by a scientific committee, composed equally of representatives of authors and publishers. Published yearly by the BnF, the list of works selected is recorded in a registry that is freely available online ([relire.bnf.fr](http://relire.bnf.fr)). The registry included more than

215,000 works on March 21, 2016 and should identify 500,000 titles by 2023. The digitisation of books in the BnF's collections began in 2014, with 15,000 items. FeniXX, a subsidiary of Cercle de la Librairie that was specifically created for this purpose, ensures the commercial distribution of e-books. The digitised works are made accessible at the same time on Gallica intra muros and a snippet of the first 15 percent can be browsed on Gallica. Upon their entry into the public domain, books on the ReLIRE register will be made completely and freely available on Gallica.



# 01.H

## The legal deposit of e-books

---

Born-digital books sold and distributed online are within the scope of the legal deposit. From 2016, the first digital files will enter the collections on a trial basis.

### A SPECIFIC DISTRIBUTION CHAIN

Unlike for websites, e-books cannot be automatically harvested online. In the distribution chain, the owner of the website through which the sale is conducted (usually the e-book store) is never in possession of the e-book file itself, which remains on the servers of the e-distributor. It was therefore decided that a specific ingest method would be created, resembling the model of the printed work deposit.

### AN ADAPTED LEGAL DEPOSIT CHAIN

Echoing this distribution chain, a complete legal deposit chain was designed. Although the publishers stay legally responsible for the legal deposit, they are allowed to mandate their e-distributor to carry out the deposit on their behalf.

This collaboration facilitates the lifting of digital rights management technologies, which are put in place by the e-distributors. DRM technologies, just as proprietary formats, prevent easy management and proper preservation of e-books by the BnF. But e-distributors manage the files in open formats (PDF and/or EPUB), before they are converted into these proprietary formats or before the DRMs are affixed.

Also, in order to be usable by online sales platforms, the e-books are described in metadata files (in ONIX format) formatted by the distributors from information provided by the publishers. The harvesting of these files enriches the BnF's e-book processing chain.

## The processing chain

The processing chain to manage the e-book legal deposit has been designed according to the existing applications and tools. The Extranet of the legal deposit ensures the processing of metadata, which is then transferred into the general catalogue. The digital preservation system, SPAR (see part 2 section B), manages the long-term preservation of files. Only the management

system of this processing chain had to be developed, with the rest of the tools simply adapted to receive these new documents.

The entire procedure is designed to work automatically.

To secure access to collections, the e-books will only be accessible on Gallica intra muros in the research rooms.

**All of the processing procedure is designed to work automatically.**



{ BnF





02

CATALOGUING  
AND PRESERVING

---

**13 million**

bibliographic records and 2.5 million authority records  
in the catalogues

---

**2.4 million**

pages in data.bnf.fr including 2 million author pages

---

**8.5 million**

documents linked to pages from data.bnf.fr

---

**813,000**

audio-visual materials in the audio-visual system  
(or 1.7 petabytes)

---

**32 PB**

theoretical capacity of the tape libraries of SPAR (2 sites)

---

The longstanding expertise of libraries for processing collections applies especially to two areas, often referred to as the core of librarianship: cataloguing – in other words, techniques allowing a document to be identified, described and found – and conservation. In the digital age, in order to properly carry out its conservation and dissemination missions, the BnF must develop and sometimes reinvent this expertise by applying it to new objects, new media and new uses. The BnF catalogues, with over 13 million records, have started their revolution towards the web of open data. The digital repository, SPAR (Scalable Preservation and Archiving Repository), which now contains over 5 million documents, is open to third party archiving and was certified by the SIAF (the Interministerial Department of the French National Archives) in 2013.

---

---

#### **MAIN TOPICS**

**02.A\_** Data.bnf.fr

**02.B\_** SPAR: the Scalable and Preservation Archiving Repository

---

#### **RESOURCES, TOOLS AND METHODS**

**02.C\_** Evolution of catalogues: towards the FRBR model

**02.D\_** E-resources: description and access

**02.E\_** Ensuring long-term access to audio-visual material

---

#### **DEVELOPMENT OPPORTUNITIES**

**02.F\_** The bibliographic transition



# Cataloguing in the era of open data

The bibliographic description data or “metadata” are at the heart of librarianship. They feed into all physical or digital workflows of the library, from storage to curation of its collections, including conservation and digitisation. This is why the BnF must thoroughly adapt its tools and processes if it wants to effectively support the development of digital activities.

In the era of big data, qualified, reliable and structured metadata is a strategic asset: they allow library resources to be recognised, referenced and finally highlighted among the mass. The modernisation of production and dissemination methods for metadata therefore becomes an essential corollary of an ambitious digital policy. This area is rightly one of the strengths of the BnF, which has shown its ability to create and distribute a critical mass of reference metadata by relying on its assets: a solid foundation of trust from its users and partners; a wealth of millions of rigorously established records; its multimedia expertise, adapted to the richness and diversity of national and international cultural production. The open government data movement and the development of the semantic web foster technological opportunities for the development of the BnF’s public service missions.

**The open government data movement and the development of the semantic web foster technological opportunities for developing the BnF’s public service missions.**

Evolutions are technical (models and formats of metadata), computer related (data processing procedures, production applications and dissemination interfaces) and organisational (transformation of jobs and activities). The BnF is therefore undertaking a real “bibliographic transition”, characterised by the progressive development of traditional catalogues towards internet search engines and by a new way of organising access to information.

The BnF also intends to make the most of this bid to stimulate and facilitate the reuse of metadata by third parties. Since January 2014, its metadata is available under the open licence of the French State (*Licence publique de l’État*). Increased visibility of the BnF’s bibliographic resources has been noted since, also improved by their integration in the main international commercial knowledge bases which feed into numerous library systems (OCLC, Ex Libris, EBSCO etc.). The BnF’s long-term objective is to be part of a genuine

cultural metadata public service, jointly created with public (the national archives, the bibliographic agency for higher education etc.) and private (publishers, start-ups, cultural industries etc.) stakeholders and offering innovative services with high added value.

# Digital preservation

Conservation is one of the BnF's key missions and is developed in continuity with physical conservation when it comes to preserving digital content. As argentic photography and microforms are progressively abandoned, digitisation is becoming the only reproduction method for conservation purposes.

From the 1980s, the BnF launched major reproduction campaigns via microform of its most fragile collections, as part of a large-scale safeguarding process. Digitisation offers several advantages over microform: better image quality, easier viewing of reproduced documents and remote viewing. However, the question of its reliability in terms of conservation still remains: the longevity of digital media is linked to its sustainability, whether technical or in terms of costs.

Moreover, the breadth of resources devoted to digitisation over the last few years, as well as the enrichment of collections via ever more significant and heterogeneous masses of born-digital documents (see part 1), have broadened the issue of digital conservation. The challenge is no longer about safeguarding originals at risk of deteriorating or disappearing by providing a substitute media, but rather about considering born-digital material as part of the collections as rightfully as earlier media – physical (manuscripts, printed works, objects, LPs), magnetic (discs, video and audio tapes), optical (CD-ROMs, DVDs) – which constitute the heritage of the BnF in all its diversity.

**The breadth of resources devoted to digitisation, as well as the enrichment of collections via masses of born-digital documents, has broadened the issue of digital conservation.**

Faced with this new challenge, the BnF has gradually implemented, since 2005, a system for making digital data sustainable (SPAR). SPAR is unrivalled in France in that it promises, for very significant volumes of data (current capacity of 2 x 16PB), a longevity beyond the lifetime of the media and the technology (software, formats, operating systems, terminals etc.). SPAR was deployed in several stages and continues to evolve to meet the growing needs of digital safeguarding. It is an open and modular system relying on the international [OAIS](#) standard (Open Archival Information System), capable of evolving over time, as computing environments become obsolete and are replaced by new ones.

However, technical solutions are not sufficient to improve the management of digital collections or ensure their longevity. The different workflows that are supposed to be ingested in SPAR need to be prioritised in terms of resources, based on scientific criteria: priority to the most fragile documents, differentiated risk management depending on the vulnerability of documents, their scarceness or the importance of their content etc. Permanent access to

content with guaranteed authenticity and confidence in the conservation system also rely on the documentation and archiving of the access platforms: archiving a video game or a web page cannot be limited to just saving its content. Preserving digital heritage also means preserving the historical breadth of its immateriality: the sensory experience that accompanies it, the memory of the way in which culture and knowledge impregnate the daily life of a society.

# 02.A

## Data.bnf.fr

The pioneering [data.bnf.fr](http://data.bnf.fr) application is now facing great challenges and offers high expectations regarding open data. Although the BnF will still need to rely on distinct applications, with functionalities adapted to its different resource types (Gallica, the general catalogue, BnF archives and manuscripts etc.), data.bnf.fr aims to act as a hub, providing access to all of these resources on the web.

### A TOOL INTEGRATED INTO THE INTERNET ECOSYSTEM

This tool offers internet users rich pages, organised into relevant topics (authors, works, places etc.), as well as numerous links, both between these pages and to other sites. This strategy, differing from the traditional display of information in library catalogues, encourages search engine optimisation, generates traffic for other BnF applications and, lastly, is able to engage a considerably larger audience than that currently enjoyed by libraries.

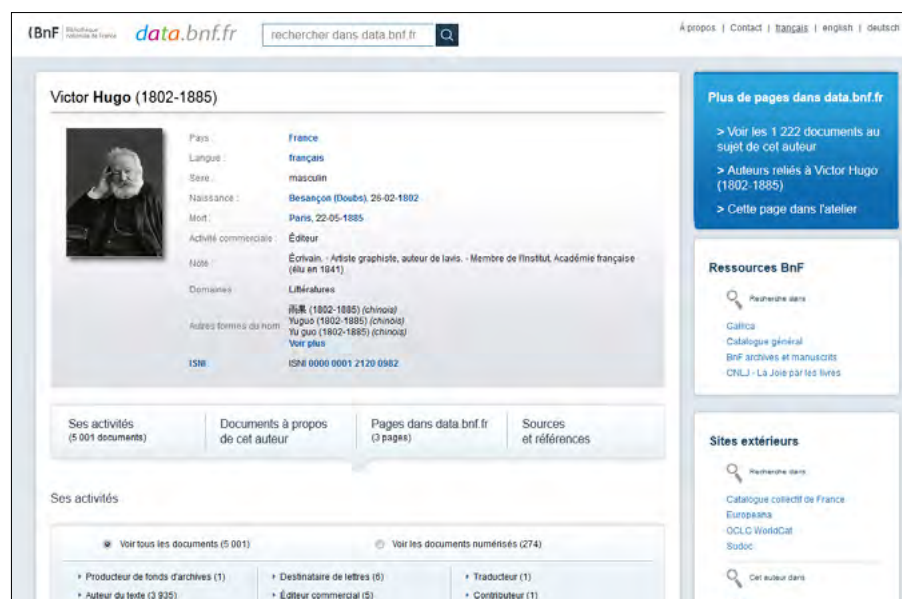
### A SMASH HIT

Launched in 2011, data.bnf.fr is a smash hit already: daily visits doubled between the end of 2014 and the end of 2015 to nearly 10,000, making it the BnF's fourth most consulted site.

**data.bnf.fr is now facing great challenges and offers high expectations regarding open data.**

### THE CORNERSTONE OF A NEW COOPERATION METHOD

Relying on technologies and formats more open than those traditionally used by libraries, data.bnf.fr creates the possibility of data co-production implying a more diversified professional community. In the years to come, the BnF intends to unify an ecosystem of partner producers and re-users of cultural scientific and educational data around data.bnf.fr, thanks in particular to an active policy of links and the integration of data coming from partners in the cultural domain. In 2016, data.bnf.fr will have reached technological maturity with a comprehensiveness and robustness that will make it the cornerstone of new cooperation methods, new uses and new services. In 2017, it will be updated monthly.



## An open data approach

Since 2014, all bibliographic metadata from data.bnf.fr are available under the open licence of the French State (Licence publique de l'État) supported by the Etalab mission.

By encouraging wide distribution and free re-use, including commercial re-use for its metadata, the BnF seeks to demonstrate its ability to serve more users and to support the creation of new services, such as mobile applications and future connected objects.

# 02.B

## SPAR: the Scalable Preservation and Archiving Repository

The growth and diversity of BnF's digital collections – several thousand formats – raise the crucial question of the repository for these digital objects. To this end, the BnF launched the SPAR project (Scalable Preservation and Archiving Repository). A complete digital repository, SPAR answers the issues of sustainability and security of digital data. Initiated in 2005, SPAR relies on authoritative international standards: [OAIS](#) (Open Archival Information System) and [PAIMAS](#) (Producer-Archive Interface Methodology Abstract Standard), [METS](#) (Metadata Encoding & Transmission Standard), [PREMIS](#) (Preservation Metadata Implementation Strategies).

### PREVENTION AND GUARANTEES

SPAR is not just a secure data storage system for making multiple copies – three copies over two sites – and ensuring the continuous monitoring of storage media. Through precise and comprehensive format recognition mechanisms of the transferred data, it also ensures permanent access by planning and undertaking the necessary transformation in the event of technological obsolescence of the dissemination tools (software, operating systems, reading equipment etc.). The SPAR system is organised along multiple channels (digitisation, web legal deposit, third party archiving etc.), in order to account for the diversity of the collections which it has to preserve. From May 2010 to January 2016, 2.23 petabytes of data representing 5 million archive packages were ingested into SPAR. On 29<sup>th</sup> March 2013, SPAR was certified for three years by the Interministerial Department of the French National Archives for the archiving of current and semi-current public electronic records (published in the Official Journal of 2<sup>nd</sup> May 2013).

### SERVING A COMMUNITY

SPAR is not designed just as an in-house tool: it was also designed to be of service to a community. Therefore, the BnF has implemented a third party archiving solution, to give organisations the same safety and sustainability guarantees as those applied to its own files deposited in SPAR (see part 4 section C).



The infrastructure of the SPAR system: tape libraries for archive storage

From May 2010 to January 2016, 2.23 petabytes of data representing 5 million archive packages were ingested into SPAR.

Investment in the development of the digital store has been significant, to cover the increasing volume and system maintenance – the hardware (discs and tapes) represents 500,000 to 1 million Euros per year, while the software represents 600,000 Euros per year. These costs are supported by the BnF's

own budget and by fundings from the CNL. The search for a third site by 2020 is to combat the effects of a major disaster (e.g. a magnetic solar storm, nuclear accident or major climate event), which could simultaneously affect both current sites – François-Mitterrand and Bussy-Saint-Georges.



# 02.C

## Evolution of catalogues: towards the FRBR model

In 2013, the BnF undertook a makeover of its [general catalogue](#)'s public interface, to adapt it to internet users and provide specific entry points into the collections (Children and youth, Images and maps, Music). From 2014 to 2016, the public interface of the [BnF's archive and manuscripts catalogue](#) was redesigned to increase user-friendliness, with a better-performing system.

### A MODEL ADAPTED TO USES

The catalogue data structure is required to evolve along with the web. The internet user is interested in people, concepts, works and places, more than in viewing records or locating documents. The FRBR model is the cornerstone of this change.

The goal of this new model is twofold: to present bibliographic information in a more intuitive way for the user and to facilitate the management of information by clustering different types of publications associated with the same work (translations, digital versions, different formats, different editions etc.).

### A FULLY FLEDGED PROJECT

For an institution like the BnF, this fundamental change in the structure of data first implies a vast retrospective processing of its bibliographic resources, which represent over 13 million records. Data.bnf.fr provided technologies such as alignment and clustering algorithms, which were instrumental in initiating the first steps of data transformation.

This processing is still of an experimental nature and requires the use of complex algorithms as well as the support of experts in order to define confidence levels. The transfer of the generated data back into the source catalogues is the end goal of such an approach and contributes to the proper preparation of the bibliographic transition. 150,000 records of printed books, sheet music, audio material, films and electronic and multimedia documents were therefore automatically linked to a work record when the first data was loaded into the general catalogue at the start of 2016. The next stages will concern the creation of new records for works.

#### Castor et Pollux . RCT 32A-B opéra

Jean-Philippe Rameau (1683-1754)  
Livret de : Pierre-Joseph Bernard (1700-1775)



Langue : français

Date : 24-10-1737

Note : Tragédie lyrique en 5 actes et 1 prologue. - Livret de Pierre-Joseph Bernard. - 1re représentation : Paris, Académie royale de musique, le 24 octobre 1737. - L'oeuvre fut révisée pour une reprise à l'Académie royale de musique, le 11 janvier 1754 : suppression du prologue, nouvel acte 1, condensation des actes 3 et 4

Domaines : Musique

#### Éditions de l'œuvre

☒ Voir tous les documents (167)

☐ Voir les documents numérisés (8)

▸ Partitions (123)

▸ Enregistrements (33)

▸ Spectacles (10)

▸ Livres (1)

Example of information  
structuring following the  
FRBR model in data.bnf.fr

## THE FRBR MODEL

This international standard was designed within [IFLA](#) almost twenty years ago. By multiplying the links between the different entities (authors, works, subjects), it prepares the bibliographic data for its integration into the semantic web.

Nowadays, libraries mainly apply the FRBR model in their catalogues by automatically grouping the documents according to specific rules. The development of cataloguing standards in the years to come will allow for a more systematic adoption of FRBR in production workflows, using more reliable criteria.

Thanks to the clustering  
of different types  
of publication associated  
with the same work,  
bibliographic information  
is presented in a more  
intuitive way.

# 02.D

## E-resources: description and access

### TWO INGEST METHODS

Beside the legal deposit, digital resources are acquired via two major types of operation. The first is subscriptions to publisher platforms, with or without file acquisition; the second consists of donations and digital file acquisitions, for which access to content must be ensured by the BnF.

The acquisition of electronic documentation from publishers can be done individually or in packages. It concerns a great diversity of content: e-books, periodicals, databases, but also images, maps, sheet music etc.

These resources, accessible on-site and remotely depending on the licences, are offered to readers according to two protocols: by alphabetical or thematic list (over 240,000 entries were recorded in December 2015 just for books and periodicals); by the EBSCO integrated search engine (EBSCO Discovery Service), but only for a subset of the electronic resources acquired by the BnF.

The digital files which are not accessible on publisher platforms will be available on Gallica intra muros via a new channel, "Ingesting purchased and donated digital documents" (see part 5 section I), currently under development.



### THE CHALLENGE: DESCRIPTION

Nowadays, these resources, whose description is basic and disconnected from the general catalogue, are hardly known by readers. Their consultation rate therefore remains very low, even though a quarter of the annual budget of current acquisitions from the collections directorate is dedicated purely to electronic documentation and some foreign periodicals are now only accessible online. This is why priority is now given

to the description, in the general catalogue, of electronic resources to which access is sustainable as well as of documents from the ADDN channel.

In 2016, technical solutions such as the OpenURL protocol and data integration from external sources will be investigated.

**Accurate description is one of the cornerstones of the development of access to these e-resources.**

# 02.E

## Ensuring long-term access to audio-visual material

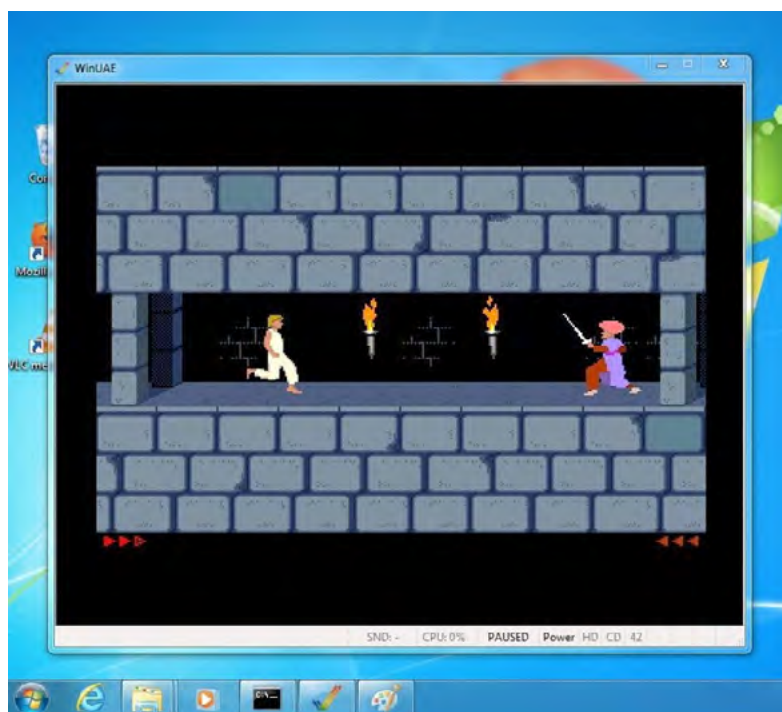
### THE EMULATION OF ELECTRONIC DOCUMENTS

Following digitisation (see part 1 section B), some documents require additional intervention to be fully readable and viewable. Multimedia and interactive documents and even more so software and video games, protected by anti-copying processes, generally require a specific intervention. A system has had to be developed to carry out the emulation of these old documents on recent computer workstations, by simulating the operation of rendering devices and operating systems from the period. A programme to copy and emulate the original software environments, system by system, console by console, has therefore been underway since the middle of the last decade: documents on Thomson platforms (games and especially educational documents in partnership with the 1985 IT for all plan, for TO7 and MO5) were able to be processed, as well as games on floppy discs from the 1980s for the Amiga and Atari platforms and all of the PC floppy discs. A new project launched in 2014 concerns CD-ROMs.

### THE AUDIO-VISUAL SYSTEM

The audio-visual system was created in 1996 for the archiving and consultation of audio-visual material. With an annual budget of 450,000 Euros, the system is maintained day-to-day by a department of six (engineers, a systems administrator and a technical operator), and is gradually upgraded in order to respond to ever larger and more diversified file storage needs. This infrastructure doubles up as a scalable interface for consulting collections and is regularly adapted according to the needs and uses of its audience. The audio-visual system conserves over 813,000 documents, corresponding to 1.7 petabytes. Currently, this dedicated infrastructure covers 80% of consultations.

A link must be established between access to audio-visual material and consultation of the INA's (French National Audio-visual Institute) collections



Emulation of the *Prince of Persia* video game (floppy disc, Amiga platform, 1990).

which are viewable at the BnF, at the François-Mitterrand site (general public service in the reference library; French radio and television legal deposit at the INAthèque in the research library).

The integration of digitised audio-visual material from the BnF into SPAR also represents a challenge to be met in the years to come. The system was designed to ingest them and already stores some of them, corresponding to the recent digitisations carried out within partnerships. A massive migration must, sooner or later, be envisaged: it will be a complex and very large-scale project likely to last several years.

**The integration of digitised audio-visual material from the BnF into SPAR represents a challenge to be met in the years to come.**

# 02.F

## The bibliographic transition

In France, the major public stakeholders in charge of bibliographic information standardisation have jointly committed, under the aegis of AFNOR and via the French National Bibliographic Transition Programme ([transition-bibliographique.fr](http://transition-bibliographique.fr)), to a vast revision of the current cataloguing standards and instructions, all the while ensuring the retrospective processing of historic data in order to ensure their compliance with this new model.

### A NEW CATALOGUING CODE

The outcome of this work is a new cataloguing code (RDA.FR), which aims to adapt the international RDA (Resource Description and Access) code to French cataloguing practices, thus allowing gradual adoption. Training and awareness raising actions among professionals such as library system vendors will be carried out, in order to support change over an estimated period of 10 years.

### THE CHALLENGE OF THE SEMANTIC WEB

In a constantly growing production context it is essential to take advantage of the possibilities offered by semantic web technologies, like sharing data by combining them with each other instead of copying them. In order to reduce its cataloguing



costs, the BnF will develop production networks in order to concentrate on its own added value. This represents a profound cultural change. The integration of a growing number of metadata coming directly from publishers and distributors, or even the derivation of international records during acquisitions of foreign monographs, are emblematic examples of this openness. It is also the path jointly envisioned by the ABES and the BnF, who want to build national authority files together.

**In a constantly growing production context it is essential to take advantage of the possibilities offered by semantic web technologies.**

The sustainable and standardised identification of resources constitutes another key challenge for information organisation and sharing. The BnF will continue to invest heavily in the maintenance and development of historic identifiers (such as the ISSN or the ARK identifier – Archival Resource Key), while new projects will be launched, especially [ISNI](#)

### REBUILDING THE PRODUCTION TOOLS

In 2016 an initial study was started into the rebuilding of production tools. A first step will be the prototyping of a national authority file in partnership with other public operators, the ABES being at the forefront. Although set and irreversible, these advancements will be done step by step. The BnF does not wish to precipitate a cataloguing revolution; it prefers to demonstrate the benefits of change while preparing the bibliographic transition via gradual stages.

## ISNI: an identifier for people and organisations

The BnF was the first library in the world to integrate people's and organisations' name identifiers conforming to the [ISNI](#) (International Standard Name Identifier) standard into the general catalogue authority file. Nearly one million ISNIs were loaded during summer 2013 and, since 2014, have been recoverable in the bibliographic products and via [data.bnf.fr](http://data.bnf.fr). The BnF became an ISNI registration agency in 2014, thus leveraging a very

tangible collaboration with a large number of stakeholders: the inter-professional book organisations (Dilicom and the Commission interprofessionnelle du livre), the ABES and even the French National Archives. These very promising collaborative projects could precede a national governance for author and community repositories, in the form of a national federation of public operators, among which the BnF would be a leading figure.



{BnF





# 03

INNOVATING  
FOR THE BENEFIT  
OF USERS



---

**3.5 million**

documents on Gallica

---

**3.8 million**

documents on Gallica intra muros

---

**15.8 million**

Gallica visits in 2015 corresponding to 330 million  
pages viewed

---

**+170,000**

followers on the official Twitter account of the BnF

---

**+110,000**

fans on the Gallica Facebook page

---

**13,000**

reproduction orders per year

Gallica is one of the world's leading digital libraries, with a secured audience and user satisfaction. Moreover, in 2013 Gallica and data.bnf.fr obtained the innovation prize awarded by the libraries of Stanford University. However, the BnF's mission for disseminating culture and developing knowledge is also challenged by digital technologies. The link between services offered on-site and those available online is at stake, leading to a rethinking of content curation via new tools, and encouraging the conquest of new territories, notably through social networking.

In 2015, the BnF launched a revitalisation project for its user policy, with a spirit of sharing and openness, targeting a broad spectrum of users. The challenge is to offer a positive, enriching and adapted experience. This strategy is articulated on-site and online.

#### MAIN TOPICS

- 03.A** The www.bnf.fr website
- 03.B** Gallica content curation targeting different user types
- 03.C** A digital communication ecosystem
- 03.D** Online research tools
- 03.E** Circulation of digital collections for commercial purposes
- 03.F** Scientific publications

#### RESOURCES, TOOLS AND METHODS

- 03.G** Profiles and expectations of Gallica users
- 03.H** The digital cultural policy
- 03.I** Digital accessibility

#### DEVELOPMENT OPPORTUNITIES

- 03.J** Dissemination of Gallica resources and API strategy
- 03.K** Developing the general public's use of Gallica
- 03.L** Artistic and cultural education
- 03.M** Digital lending
- 03.N** Improved access to digitised newspapers
- 03.O** Wi-Fi access



## Gallica, a multi-faceted product

Until 2012, Gallica was a single digital library providing access to the same documents and offering the same features to all internet users, whether they connected remotely or within the BnF. At the start of 2012, Gallica intra muros was opened to the public, thus extending the horizons of the digital library by making copyrighted content available on the BnF premises. Gallica intra muros therefore became the main digital platform for accessing the BnF's digital resources, especially as it will soon include part of the acquisitions and donations of digital documents (ADDN – see part 5 section I), as well as e-books entered via legal deposit. Simultaneously, the implementation of a test platform, Gallica Labs, in 2012, the launch of mobile applications (iOS in 2012, Android in 2013), the Gallica mobile website in 2013 and the first vanilla websites (Numistral in 2013, La Grande Collecte in 2014), have

made access to digitised collections even more diverse in terms of content as well as features. Gallica has therefore become a diversified service, adapted to users on the go and to all kinds of devices and responding equally well to the expectations of those in reading rooms and to internet users around the world.

**An entire  
community,  
the Gallicanautes,  
has emerged  
combining  
scholarship with  
creativity.**

Driven by a very active social networking strategy, the digital library also pushes content via a blog, a newsletter, social networks like Facebook, Twitter and Pinterest and through APIs like IIIF (International Image Interoperability Framework), with the aim of engaging an ever wider and more diverse audience, from the most specialised researcher

to the curious reader browsing the immense richness of the digitised collections in search of serendipity. An entire community, the *Gallicanautes*, has emerged combining scholarship with creativity.



# Towards the digital reading room, everywhere and for everyone

In the BnF reading rooms, the challenge is to provide readers with a customisable working space, seamless access to physical documents and digital resources, and an optimal environment for conducting their research. The cable and Wi-Fi internet access allow researchers to easily access free and open services ([www.bnf.fr](http://www.bnf.fr) website, catalogues and Gallica), as well as digital resources under copyright (web archives, protected works from Gallica intra muros, purchased e-resources), using their own equipment. Launched in 2014, the AVEC (BYOD – Bring Your Own Device) portal intends to satisfy this demand by providing access, on users' personal devices, to all of the digital documents in the research library, including those still under copyright.

Beyond already available office software (Word, Excel, PowerPoint), the BnF's computers also offer specialised software tools, for instance within the audio-visual creation studio in the reference library, which offers readers the chance to try out audio and visual manipulation and 3D modelling. This full range completes a set of measures deployed a long time ago – the audio-visual service in the François-Mitterrand reference library has been entirely digital since 2004 – which enable

**Researchers  
and readers are  
increasingly engaged,  
showcasing a more  
collaborative  
knowledge economy.**

access to carefully curated content (video games, multimedia documents etc.). Users still need support to appropriate tools and resources: workshops to discover specific collections like that of PRISME (Resource and information centre on the business world) are available to jobseekers and those wishing to set up their own business. In the future these initiatives will extend to other software used to manipulate digital resources of a varied nature by researchers as well as the general public (NASA World Wind, MARBLE, object recognition, photo retouching, and image confrontation software, 3D object viewers, 3D design etc.).

The tools made available to researchers in the years to come must enable them to better understand the masses of data produced and collected by the BnF, thanks to big data technologies: named-entity recognition, semantic search engines etc. Researchers and readers are increasingly engaged in the consultation and indexation of documents, showcasing a more collaborative knowledge economy. Interactive tools like the experimental Correct platform, allowing for the correction of the OCR of Gallica documents, demonstrate this disruption of practices.

# Creating ambitious, varied insights from content

A specialised driver for the diversification of users, user-oriented content curation has taken on a growing importance in an institution that has, for a long time, been directed mainly at researchers and readers. The BnF's cultural project (exhibitions, educational service, events and visits) was designed 20 years ago as an extension of the reading rooms, offering the public an alternative access to heritage.

Opening up to new users – secondary school and college students and even, more recently, children at the French National Centre for children's Literature (CNLJ) – is one of the remarkable consequences of content adaptation for literacy and education purposes, which is now at the heart of the BnF's practices. The Library is positioning itself, via the digital resources that it circulates, as one of the major stakeholders of artistic and cultural education.

In the digital world, this strategy takes various forms. On Gallica, digital corpora corresponding to a collection or selection of reference works are created by the collection departments. Other products are created for specific user groups as part of collaborations and partnerships with external institutions, such as schools ([classes.bnf.fr](http://classes.bnf.fr)) or through virtual exhibitions

([expositions.bnf.fr](http://expositions.bnf.fr)). Addressing students and apprentices as much as teachers, this service, with an educational goal, also aims to engage those who would not dare or dream of entering a cultural site and requires other intermediaries.

Portals were developed within international digital cooperation programmes, for the creation of thematic collections and the

reconstitution of dispersed collections e.g. the Europeana Regia project, a European digitisation programme for Middle Age and Renaissance manuscripts from the great royal collections. New uses for these collections, such as research, may also emerge.

With an abundant and varied offer on the one hand and on the other the introduction of tools improving reading quality, digitisation has become, in a

short period of time, the main accessibility driver for heritage documents, thus benefiting people with a visual disability.

Finally, new territories have been conquered, in particular thanks to mobile devices – tablets, telephones etc. – which allow everyone to promote BnF content. The mission to extend the use of Gallica by the general public is being developed as part of the BnF's 2016 user development policy.

**Digitisation  
has become,  
in a short period  
of time,  
the main accessibility  
driver for heritage  
documents.**







# 03.A

## The www.bnf.fr website

At the heart of the digital communications ecosystem, [www.bnf.fr](http://www.bnf.fr) is the shared portal through which a considerable volume of information and services are exchanged, aimed at users who are themselves numerous and varied.

### THE PORTAL

The portal introduces the Library and provides access to all services: librarianship tools, professional resources, the cultural and events programme and even practical information, all content which benefits from high visibility via a website recognised by users, the rebuilding of which will be a major challenge in the months to come. This portal emphasises the sub-domains of [www.bnf.fr](http://www.bnf.fr) which provide access to the Library's core services (catalogue, Gallica etc.), as well as specific services (reserving spaces and documents).

### THE MOBILE WEBSITE

Available online since October 2013, the mobile version has its own design and architecture while using the same content as the portal, reorganised according to its usefulness in a mobile context. Since April 2015, this version has included a personal space for reserving areas and documents.

### FEEDBACK FROM THE SATISFACTION SURVEY

A qualitative survey undertaken in spring 2015 by the Creative Works agency highlighted site users' areas of satisfaction and criticism. This survey gives a better understanding of the current perception of the website and its mobile version. Its conclusions reveal areas for improvement for the [www.bnf.fr](http://www.bnf.fr) site, but also the loyal and enthusiastic relationship that those surveyed have with the BnF. There was very positive feedback on the mobile version, which offers flexibility, technical reliability and time savings.

### REVISION OF THE WEBSITE: A CROSS-DISCIPLINARY PROJECT

A complete revision of the website was launched in autumn 2015, with, as its aim, the uploading of an initial version at the end of 2016. Mobilising, in a cross-disciplinary way, a large number of stakeholders and allowing the BnF to reconsider its relationship with its users, the project implies major challenges: of a technical (change of CMS – adoption of Drupal – and migration of existing content), methodological (selection of the Scrum methodology known as “agile”, see part 6 section D, already used at the BnF) and editorial (the readability, clarity and highlighting of services and resources) nature.

**A complete revision of the website was launched in autumn 2015.**

The personal space of the mobile website and the homepage of [www.bnf.fr](http://www.bnf.fr)



## 03.B

# Gallica content curation targeting different user types

In the survey into the uses of Gallica conducted in 2011, one user stated: “It would be useful to help (new) users in the same way as librarians guide a search in a physical library.” In order to meet this expectation, the BnF produced 306 web pages of user-oriented digital content grouped together into 25 corpora, meant as a pedagogical access method to Gallica’s collections.

### MORE VISIBLE DIGITAL CONTENT MANAGEMENT FOR USERS

Designed with a content management system (Drupal), the user-oriented content pages, online since 2014, are especially highlighted in the new version of [Gallica](#) (2015): promoted on the homepage, they are accessible from any page of the website via the “Collections” button. They provide a number of access routes – by theme or location in addition to item. New “object” and “video” categories have been added. From 2016, the existing templates will be improved in line with contributors’ requests, with historical timelines, interactive maps etc. The BnF and its partners engaged in Gallica will also study their contribution to pages presenting user-oriented and managed content.

### THE GALLICANAUTES, AN ACTIVE COMMUNITY

Over the longer term, the promotion of content from the Gallica website itself could rely on an enlarged scope of contributors: beyond staff from the BnF and its institutional partners, *Gallicanauts* could themselves help to enrich and create contextualised content, following specific processes and using tools to be introduced. The addition of links to documents from Gallica in Wikipedia presents a challenge in terms of increased visibility.

## Les Essentiels de la littérature (The Essentials of literature)

As part of its educational service, the BnF is developing [Les Essentiels de la littérature](#), a specific website aiming to offer a wide audience (secondary school level) simple access to essential works of literature. These are available on Gallica, via a structured interface with modules on authors, works and themes.

A prototype on Montesquieu went online in October 2014, with two work modules, dedicated to *The Persian Letters* and *The Spirit of Laws*, plus a module on slavery, giving good insight into this digital, user-oriented, content management project. The project continued in 2015, with the uploading of the 18<sup>th</sup> century, which will be followed by that of the 19<sup>th</sup> century and the start of the 20<sup>th</sup> century.



# 03.C

## A digital communication ecosystem

Thanks to its “portal” website offering numerous services and sub-sites for professionals and the general public, the BnF has become a major digital stakeholder in France and beyond. It is developing and presenting this strong identity via diverse social media channels. Blogs, research logs on collaborative platforms, newsletters, video channels and social networking accounts all form interlinking bridges between the BnF and its users.

### THE KEY ROLE OF SOCIAL NETWORKING

In this ecosystem, digital social networking plays a key role. [Facebook](#), [Twitter](#), but also [Pinterest](#), [Instagram](#), [YouTube](#), [Dailymotion](#), [iTunesUniversity](#) all host “official” accounts for the BnF. These general communication channels, which showcase the BnF’s essence and programme as well as its collections and heritage challenges, contribute to a dynamic image of the Library, connected to the world’s life and news.

Other social networking accounts focus on the richness of collections and the diversity of services via dedicated accounts (accounts linked to the digital library Gallica, [Ad vivum](#) blog for the prints and photographs department for example) which encourage discussions relating to the activities of a department and the deepening of links with communities of researchers and specialists.

These general communication channels contribute to a dynamic image of the Library, connected to the world’s life and news.

### THE THREE GOALS OF A LONG-TERM STRATEGY

To establish and affirm this new virtual and living BnF, three avenues emerge.

The first is to strengthen its online presence via an ambitious, more coordinated, digital communication, in tune with new habits. Several measures contribute to this: two staff training courses, “Communicating on social networking” and “Writing for the web”, a social media charter, and the implementation of tools both for defining editorial lines and planning publications.

The second avenue relies on durably engaging internet communities and intensifying the feeling of belonging for the BnF’s users (invitations to users e.g. *Gallicanautes* to become ambassadors for the BnF, looking for links and internet communities further afield or those interested in unusual topics).

The third avenue is to expand the BnF’s e-reputation monitoring and management activities. Taking a step back, analysing performance, understanding readers’ expectations, managing crises – these are the challenges of an expanded reputation monitoring, organised via the BnF’s online presence outside of its own channels. The BnF can rely on a considerable asset in this area – its public credibility regarding heritage promotion.

### The Gallicanautes

Internet users who use Gallica and participate actively in the circulation of its documents on the web (the *Gallicanautes*) have a place within the digital library, which dedicates a specific page in its new version, “[Du côté des Gallicanautes](#)” (Alongside the *Gallicanautes*), to them. From the specific reuse to the recreational diversion of the most varied collections (medieval manuscripts, papers from the 19<sup>th</sup> century etc.), reuses by *Gallicanautes* form a rich and varied palette: vintage colouring, cutting out and sticking together globes, illumination diversion, creating animated GIFs, trying out (old) cooking recipes, illustrations of monuments visited during a trip etc.

One *Gallicanaute*, Thomas Henry, who is passionate about the history of records and author of the blog *Ceints de bakélite*, has launched the “Disquaires de Paris” (Record shop owners of Paris) project and created a website dedicated to the history of Paris

record shop owners where he offers an interactive map of Paris, providing access to digitised resources, often from Gallica. Another, Johanna Daniel, known as “Peccadille”, created the *Orion en aéroplane* blog in 2012 to “share the wonderful things that her studies in art history led her to discover”. Her blog suggests using the Gallica tablet and mobile app to track the history of Paris monuments.



## 03.D

# Online research tools

### SINDBAD: PROVIDING REUSABLE RESPONSES

In order to efficiently process users' numerous queries and information needs, the BnF has created a dedicated service – [SINDBAD](#) (online and telephone reference service). Sent via instant messenger, telephone or webform, users' questions receive a rapid response there, live or via email.

Over 10,000 questions are processed every year. 1,800 responses provided to date have been archived in the "Knowledge base", hosted on the QuestionPoint site (OCLC). It is planned to incorporate these questions and answers into the pages of the [www.bnf.fr](#) website in order to allow for their crawling by search engines. Internet users accessing the database could, moreover, find an answer to their questions without leaving [www.bnf.fr](#).

### THE "CHERCHER & TROUVER" (SEEK & FIND) WEBSITE

This research assistance measure has been complemented, since January 2014, with the "[Chercher & trouver](#)" website, with twenty short research memos on how to find a newspaper article, a film's copyright owner, travellers' writings etc. Each memo specifies the sources to refer to and how to use them. New memos will enrich the website by covering more areas.

### RESEARCH ASSISTANCE PORTALS

At the end of 2012, four [portals](#) produced with the LibGuides software went online, as an extension of the four thematic resource centres in the reference library: Sustainable development, Europe, French-speaking countries and Public debates. In 2014, regular communication on these portals on [www.bnf.fr](#), the Facebook page of the BnF and the Twitter account @laBnF resulted in a 36% traffic growth.

Regarding the portals' content, information packs on specific topics are to be developed. For example, the "Debate" category of the Europe portal – one of the most consulted – was extended in 2014–2015 with the following information packs: "Turkey in the European Union?", "Does the EU lack democracy?", "TTIP, a trap or an opportunity for Europe?" etc. Each is completed with suggestions for further reading, websites and videos, allowing readers to grasp the different points of view on these issues.

Since 2015, the BnF has also provided a resource and information website on the book professions, the [Métiers du livre portal](#).

Example  
of a thematic portal:  
the Europe portal



Within these portals, suggestions for further reading, websites and videos allow readers to understand the different points of view on current issues.



# 03.E

## Circulation of digital collections for commercial purposes

With 13,000 orders per year, the on-demand digitisation service ensures the fee-based reproduction of approximately 25,000 digital documents – 1.2 million pages reproduced – for 6,500 French and foreign customers.

This commercial service complements the document supply service by allowing readers to access works not yet digitised. It also promotes collections from the BnF via the sale of digital images to professionals (publishers, museums etc.). 1,500 commercial projects giving rise to a usage fee are therefore undertaken every year, on the basis of collections from the BnF: publications, audio-visual productions, exhibitions, websites, mobile applications etc. Altogether, these activities generate, on average, 1.4 million Euros annual income for the BnF, with a growth target of 200,000 Euros per year by 2018. Sales are completed via the pictures collection ([Banque d'images](#)) – a collection of reproductions of out of copyright iconographic documents ordered by customers (approximately 200,000 images).

### ADAPTING THE RANGE OF SERVICES

In an extremely competitive digital content market, the implementation of an ambitious commercial strategy implies a range of services regularly updated to meet customers' expectations.

This imperative must lead to the constant development of the reproduction offer, with an improvement in the standards and quality of digitisation, a diversification of the products offered and the implementation of an adequate range of fares.

The development of new sales and content promotion tools is also necessary: online sale of reproductions, a new BnF pictures collection, the simplification of the customer path, content access and curation etc. The impact of these improvements on customers and their habits will be measured by sales and customer satisfaction indicators.

### PROSPECTS LINKED TO VISUAL RECOGNITION TECHNIQUES

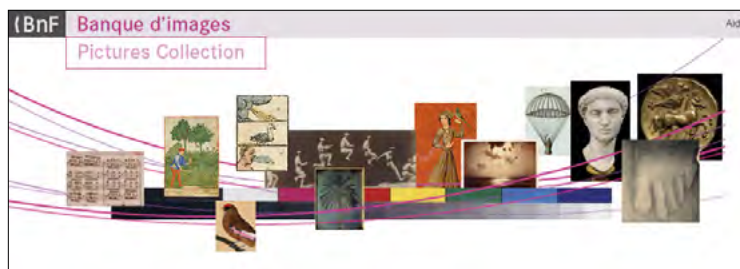
Moreover, to improve the cataloguing and online visibility of its digital iconographic heritage, the BnF is conducting, with the ETIS laboratory from the CNRS, a scientific project entitled ASAP (Semi-automatic annotation of heritage images), the objective of which is to offer new assistance solutions to describe digitised images. Relying on the latest visual recognition by content similarity algorithms, the work will deal with the design of automatic indexing functions calculated with the assistance of already annotated images from the BnF.

**The possibility to widely disseminate images and metadata should lead to building up a distribution network via foreign image agencies.**

### Pictures collection, version 2

In order to support income development, a new pictures collection, currently being designed, will present, in one single place, a large part of the iconographic documents from the BnF and will integrate various copyrighted collections currently under-exploited.

The integrated purchasing and downloading options will be able to automate the sale and manage the usage fees. The possibility to widely disseminate images and metadata should lead to building up a distribution network via foreign image agencies.



## 03.F

# Scientific publications

The BnF supports France's national Open Access policy regarding sharing, disseminating and preserving research results. This requirement was especially met by the release of an online bibliography of the BnF professionals who are conducting research activities: the [BnF Scientific Production](#) site. The full text of articles is circulated on the [open archives portal of the BnF](#) in HAL (Hyper articles online). At the same time, the BnF endeavours to upload and update its [National, International and Research Actions](#) (ANIR) database, registering its research programmes and initiatives.

### SCIENTIFIC EXCELLENCE

Moreover, in order to develop the collections, the BnF publishes *catalogues raisonnés* and scientific works designed for researchers and library professionals.

These reference publications are a reflection of the BnF's scientific excellence. An active e-publishing policy is built around them: following the launch of its "[OpenEdition Books](#)" portal, the BnF intends to increase the circulation of its scientific productions via this platform, which contained 27 works in March 2016.

### A DIVERSITY OF PUBLICATIONS

Designed to highlight collections and publicise the work of BnF researchers and librarians, the [Revue de la Bibliothèque nationale de France](#) (BnF journal) is available in full on the Cairn.info portal, with a 36 month moving wall. [La Revue des livres pour enfants](#) (Children's book journal) explores the themes dealt with in youth literature and is evidence of the critical activity and monitoring by professionals from the CNLJ (French National Centre for children's literature). It is available online with a 24 month moving wall. Moreover, the BnF ensures the promotion of its scientific events within [Calenda](#), the French-speaking scientific calendar for the arts, humanities and social sciences, as well as the publishing of academic blogs on the [Hypothèses.org](#) platform from the Centre for open e-publishing (Cléo).

Direct scientific communication tools are fully exploited by the BnF, which has gone as far as creating a LinkedIn group ("BnF – cercle des chercheurs", BnF – researchers circle), dedicated to research fellows and aimed at supporting online the BnF's thriving research community.

### DEDICATED WEBSITES

Some programmes undertaken within the BnF result in the production of websites. That's notably how [Mandragore](#), the Manuscripts department and Arsenal Library's iconographic database offering over 170,000 illumination analysis records often accompanied by a digitised image, was

created, as well as [Reliures de la Bibliothèque nationale de France](#), which presents French decorated book bindings from the 16<sup>th</sup> to the start of the 19<sup>th</sup> century, held in the Rare Books Reserve. The [Bibliographie des éditions parisiennes du xv<sup>e</sup> siècle](#) (bibliography of 16<sup>th</sup> century works published in Paris) is another of these sites showcasing the BnF's desire to confirm its active role in the dissemination of knowledge.

**An active e-publications policy is able to publicise the work of librarians and researchers from the BnF.**



# 03.G

## Profiles and expectations of Gallica users

Faced with a plethora of digital practices, media and online environments, creating a global overview of online usage and monitoring its development is a real challenge. Collecting information on users of the Gallica interface is part of this vast mission to better understand usage.

### ANALYSING USAGE

The extent of documentary usage known as “traditional” was confirmed by an initial large-scale survey undertaken in 2011 with 3,800 Gallicanautes: most of them used Gallica in the context of specific, professional or amateur research, while usage simply out of curiosity or for entertainment purposes remained exceptional. These results were complemented by qualitative surveys to reconstruct professional and amateur researchers’ navigational patterns around a theme or a corpus: online manuscripts, science and technology collections, heritage from the Great War.

ÉFÈLE Livres Efele  
@LivresEfele  
Nouvelle interface @GallicaBnF : Superbe ! Défilement vertical, double page, j'adore. Merci !

Sergio Belluz  
@SergioBelluz  
Gallica, c'est comme un grenier de grand-mère: on y trouve que des merveilles (ici un journal intime de George Sand).

Sean Fraser  
@TheatreSean  
gallicalabs.bnf.fr is @GallicaBnF's website's new extraordinary user interface; Remarkable consolidation, e.g.: [gallicalabs.bnf.fr/services/engin...](http://gallicalabs.bnf.fr/services/engin...)

RacontemoiHhistoire  
@RMLHhistoire  
Bon, je suis sur [gallicalabs.bnf.fr](http://gallicalabs.bnf.fr) je vous tiens au courant quand j'en sors, en 2023.

Hannah Williams  
@hannahwill  
Ooh Gallica has a new iPad app [itunes.apple.com/us/app/gallica...](https://itunes.apple.com/us/app/gallica...) - this might be the final motivation I need to get an iPad.

### The profile of Gallicanautes

According to the 2011 study, *Gallicanautes* mainly connect from within France, even if many use Gallica in Italy, the USA, Canada, Belgium, Spain and even Brazil. More intensive readers and highly educated than the average French person, they are also older than those who use the BnF's reading rooms and more likely to possess an e-book reader or tablet.

Although 95% of users state they are satisfied with Gallica, they still have high expectations, raised by a competitive environment (Google Books etc.): growth of digitised collections, easier browsing, search engine improvements and advanced content appropriation tools.

### MEASURING DEVELOPMENTS

A major study was launched in 2016 in order to measure developments, since the 2011 study, in Gallica's user profile and usage, and assess whether the significant changes that the digital library has undergone over the last years have allowed it to diversify its users and develop their practices. It also aims to understand current navigational patterns (between different resources, different media, and different interfaces).

### THE BIBLI-LAB

The BnF is now entirely legitimate in positioning itself as a specialised and expert observer of the usage of digitised heritage: a position that is as much institutional as scientific, strengthened by the partnership agreement signed in January 2013 with the Télécom ParisTech School. This five-year partnership has already led to the creation of Bibli-Lab, an online heritage uses research laboratory for libraries. It consists of specific agreements and projects funded in the context of calls for national and international projects. Thanks to an initial research programme, the theoretical and methodological challenges

raised by the analysis of Gallica usage were explored in all its digital library dimensions (interface, repository of documents and associated services). Recommendations were formulated to invite the assessment of usage by collection type, in relation to the BnF's strategic challenges regarding digitisation and the promotion of digital collections.

A second programme together with the BDIC concerns online content dissemination of WW1 digitised material as part of the research project "*Les passés dans le présent*" (The past in the present) (see part 6 section G).

**The Bibli-Lab, an online heritage uses research laboratory for libraries, is positioning itself as a specialised observer of usage.**



# 03.H

## The digital cultural policy

Already well embedded in the BnF's practices, the digital cultural policy is ripe for development. It combines the use of digital tools both on-site (multimedia audio-guides, for example) and online (virtual exhibitions etc.).

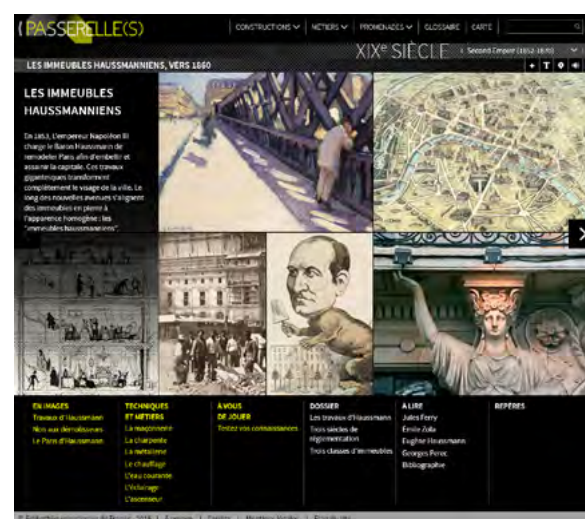
### DISSEMINATION

The aim is to speed up the dissemination of content and knowledge within different internet user communities, from the most general to the most expert. All the conferences, training courses and public interventions recordings are available on the BnF website and other connected websites and portals (Radio France, iTunes U, the sites of institutions and universities with which the BnF collaborates). Online exhibitions and other digital albums are also available on associated portals, like the French Ministry of Culture's art history portal and the French National Education Ministry's *Eduthèque* portal. Researcher blogs, for example on the *hypotheses.org* platform, which help to provide insights into collections, are regularly harvested and referenced by search engines from the scientific community, thereby increasing their visibility.

### PRODUCTION

The BnF has launched, in collaboration with Orange, a collection of enriched digital books on great works of French literature available as applications for tablets. *Candide*, *Dom Juan* and *Au bonheur des dames* are all independent editorial productions which contribute to the creation of a digital publishing policy, notably including a commercial section (via downloads and products). Attentive to the development of knowledge access methods and wishing to adapt to user habits and in particular respond to growing expectations in terms of self-learning, the BnF also plans to develop a massive open online course (MOOC).

The aim is to speed up the circulation of content and knowledge within different internet user communities, from the most general to the most expert.



## The different forms of user-oriented content curation

User-oriented content curation takes varied forms, addressing user communities of which the BnF must still improve its knowledge, in order to better adapt its services:

- **Adapting content for training purposes:** intends to guide readers in catalogues and collections. Especially addresses students not sufficiently trained in research;
- **Adapting content for introductory purposes:** linked to building and exhibition visits. Aims to explain e.g. what a national library is or what an exhibition is about. The source of growing interest, it opens the prospect of digital modelling (interactive terminals etc.);
- **Adapting content for educational purposes:** dedicated to teaching and schools, this

is encountering growing success. Its digital aspect is key and under development;

- **Adapting content for entertainment purposes:** using fun as an access method is suitable for children and families. It encourages discovery of the BnF and will be developed;
- **Adapting content for awareness raising:** for users who would not dare or dream of entering a cultural site. Includes the *Passerelles* project – a website dedicated to apprentices in the building industry. It offers an interactive exploration of various professions, notably via some 70 buildings representative of the history and techniques of their period.

# 03.1

## Digital accessibility

In the same way as it welcomes users onto its physical premises, the BnF sees the accessibility of its digital services as a priority – both its websites and online services, and the content provided. This includes a greater focus on the development of interfaces addressing people with all types of disabilities (visual, auditory, motor, cognitive etc.) and the creation of dedicated collections.

### TRAINING COMPUTING TEAMS

The BnF is developing, as a priority, the skills of those teams responsible for creating a new version of Gallica and revising the general catalogue, before extending this awareness raising to all teams in charge of the development of online services. It endeavours to follow the recommendations of the French General Accessibility Baseline for Administrations (RGAA version 2.2.1), to which the Gallica mobile website already adheres.

### DEVELOPING THE ACCESSIBILITY OF COLLECTIONS

The second major challenge for the BnF is making more documents accessible to people with disabilities. The exploitation of the potential of the EPUB format – which will be developed in the collections via the legal deposit of e-books – is heading in this direction (see part 6 section A and part 1 section H). The 3,200 items from Gallica processed in this format were also systematically produced in the DAISY accessible format. Due to its role in applying the disability exception of the French intellectual property code, via the PLATON platform (see part 5 section D), the BnF is capable of becoming, in partnership with the approved associations in charge of the adaptation and circulation of texts, one of the major stakeholders of e-book accessibility, both by contributing to the better cataloguing of adapted works and by working with publishers to include e-books on the PLATON platform.

### ON-SITE: MEETING SPECIAL NEEDS

Digital technologies greatly increase the potential provision for people with special needs of accessible tools and facilities in the reading rooms and exhibitions: video magnifiers, reading machines, applications accessible to people with a visual or auditory disability etc. A communication system via video-interpretation in sign language and instantaneous transcription of speech has been in place since the end of 2015, for communication on-site and remotely with deaf and hearing-impaired users.

**The BnF is capable of becoming a leading stakeholder in e-book accessibility.**



Video magnifier

# 03.J

## Dissemination of Gallica resources and API strategy

Gallica's ability to disseminate its resources by facilitating their search engine optimisation and online circulation is crucial to increasing the website's visibility. Three major activities contribute to this: the development of sharing tools, allowing the user to appropriate content and disseminate it easily; the introduction of measures to improve search engine optimisation and resource ranking from Gallica; and the implementation of initiatives to establish a Gallica presence on third party websites (partners and external websites).

### DEVELOPING SHARING TOOLS AND REINFORCING SEARCH ENGINE OPTIMISATION

In line with these three activities, numerous tools have been implemented: permalinks, widgets, social network sharing etc. These are subject to continuous improvement. Natural search engine optimisation for Gallica resources has also been launched, via the optimisation of web pages and the multiplication of links to the website (from the [data.bnf.fr](http://data.bnf.fr) pages for example). Moreover, Gallica has developed a dissemination policy through APIs, at first internally and for partners. Thanks to these interfaces, data use is optimised on the Gallica mobile site and on smartphone and tablet apps.

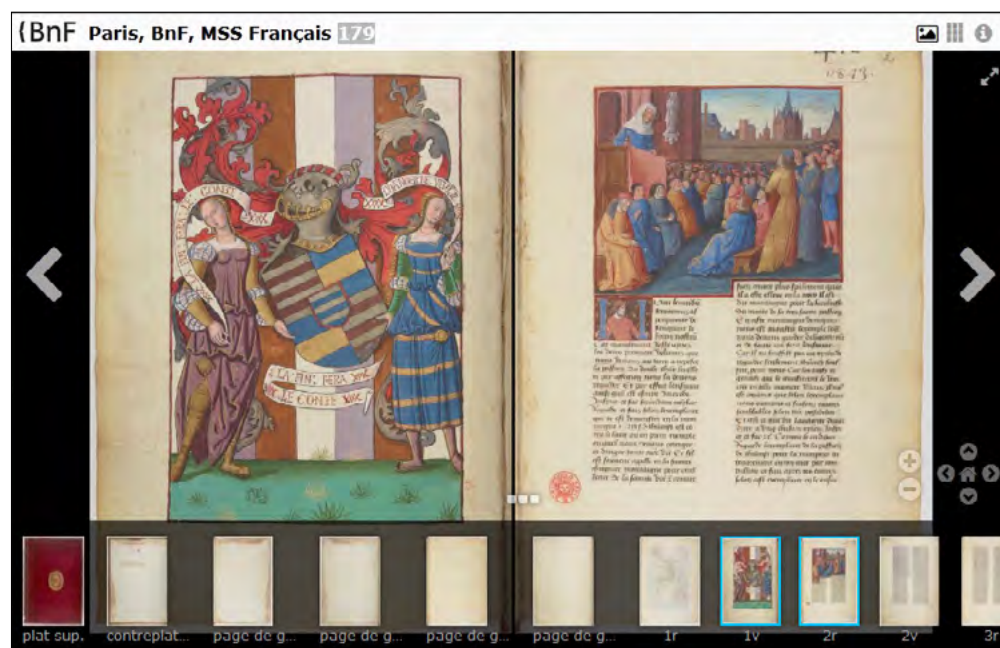
### INCREASING VISIBILITY

Metadata from Gallica resources is also disseminated in a standard and reusable way through the OAI-Num repository, as part of the BnF's open data strategy (see part 2 section A). The work to provide partners with instant, standardised access to digitised resources has been bolstered by the commissioning of the API International Image Interoperability Framework (IIIF), a protocol designed by the University of Stanford to allow for the consultation of images and parts

of images on remote sites. The BnF is a founding member of the consortium supporting this API. This kind of tool allows external developers to appropriate Gallica's resources via user-friendly technical interfaces. Directly circulated on partner sites, these resources have since seen their audience increase.

### The metadata from Gallica resources is part of the BnF's open data strategy.

Study of a manuscript  
from Gallica via the API  
IIIF on the [Bibliissima](#)  
prototype





# 03.K

## Developing the general public's use of Gallica

BnF sees the general public dimension of the digital library as a priority. It seeks to engage people beyond students, researchers, amateurs and enthusiasts who make up Gallica's traditional users. The development of leisure and discovery usage is a lever which will allow Gallica to become popular among internet users.

### ALL FOR THE BUZZ

A wider audience can be engaged by diversifying the way content is managed and increasing the number of distribution chains. The contextualisation of content in the "Collections" heading of Gallica, along with publications on other platforms, such as [Les Essentiels de la littérature](#) (The Essentials of literature), are part of this dynamic. Gallica's community managers use collaborative methods to encourage circulation via the social networks, by highlighting discoveries and organising interactive recreational activities. Educational workshops also promote these curated resources.

### ENGAGING USERS

Among the projects being studied, the "Gallica Studio" project will offer a new approach directly engaging users to co-create new content based on Gallica resources, which are therefore re-appropriated and reused. Via an interface distinct from that of the digital library, internet users will be able to import documents from Gallica in order to correct, develop, geo-locate and annotate them, and even carry out creative projects. The extended documents may then be transferred back into Gallica.

### CREATING A NEW SERVICE OFFERING

By enriching Gallica with recent publications likely to interest new readers, the digital lending project is part of a new service offering for a wider audience.

The *ABC de Babar* by Jean de Brunhoff on Gallica



## A delight for kids and adults

Drawing on all periods and forms, there are many books for children and young people on Gallica: picture books, alphabet books, fairy tales, illustrated classics, including popular prints such as *Images d'Epinal*, fun publications like colouring books, and educational ones.

Various documents are suitable for family use like the reading of tales, the bedtime book or the (re)discovery of universal texts from mythology, or French classics by Perrault, the Countess of Ségur, Jules Verne etc.

Such uses require kid-friendly tools (touch screens, dedicated Vtech or Kurio type tablets etc.) and consistent content. Offering a Gallica world for children through a specific web interface and a mobile application will allow everyone to discover a service suited to the different age ranges (0-3 years, 3-7 years, 7-12 years), developed by children's culture experts from the CNLJ (French National Centre for children's literature). Content could be provided in serial format, similar to a subscription to a children's magazine.

# 03.L

## Artistic and cultural education

### A NATIONWIDE SERVICE

Since 2015, the BnF has extended its artistic and cultural education (EAC) activities across the country. With its agreed regional partners, a new coordinated and structured training service in book heritage has been implemented, this being an EAC area of excellence for the BnF. It covers digital resources, book and newspaper history, media and information literacy, image technology and children's literature, in order to enrich France's cultural offering and contribute to cultural and artistic education, while strengthening the BnF's ties with the regions: two major objectives of its 2014-2016 performance contract.

### EXHIBITIONS AND TRAINING DAYS

This activity takes the form of an exhibition on book heritage and the BnF, to be held in a variety of institutions.

A regional training day is also available to the different stakeholders involved: teachers, school librarians, librarians, curators (museums, archives),

organisers of before -and after- school clubs, elected representatives, directors of cultural affairs, booksellers; inspectors from the General Inspectorate for Libraries; correspondents from Delegations for arts and culture (DAAC – *délégations académiques aux arts et à la culture*) from local educational authorities, school inspectors; local inspectors, regional educational inspectors (secondary schools and colleges) etc.

The complementarity of an approach based on both digital resources and physical media is at the centre of this training. In particular it consists of:

#### - Diversifying the heritage approaches:

the heritage documents from a Public library with a regional mission (BMVR) are presented, highlighting the link between the source document and the digital one, and professions and techniques (meeting with professionals, relationships with scientific, technical and industrial culture).

- **Promoting the BnF's offer:** this training is also an opportunity to engage ever wider audiences with the BnF's digital resources (Gallica; data.bnf.fr;

CCFr; the classes.bnf.fr portal etc.).

An educational kit (samples of different materials, USB key, tutorial etc.) gives a concrete idea of the educational workshops created by the BnF.

#### - Unifying stakeholders from the same region:

presentation of the outcome of a survey by the book and reading department of the French Ministry of Culture into the implementation of artistic and cultural education in libraries, as well as examples of how discovering and learning opportunities in culture and art may be provided to pupils within a region.

**Enriching France's cultural offering and contributing to cultural and artistic education are two key objectives of the 2014-2016 performance contract.**



# 03.M

## Digital lending

The free circulation of information online has profoundly changed the way people read: access to books is no longer limited by physical constraints (reading rooms of libraries) or time constraints (opening hours). Library users increasingly demand reading services that are available any time, online and on the go, via their own equipment, portable computers, tablets or mobile phones, wherever they are. The digital transition of the publishing industry is therefore accompanied by an adaptation in the library lending service.

### BETWEEN THE BnF'S MISSION AND A CHALLENGE FOR SOCIETY

Although the BnF has never been a lending library, its main mission being the preservation and conservation of a collection for heritage purposes, Gallica actually plays a similar role to a lending library for titles in the public domain or with negotiated rights. However, access to copyrighted works remains limited to the BnF's physical premises. Regarding purchased e-resources, remote access has been available since 2009 for readers with an annual research card.

### A NEW OPPORTUNITY

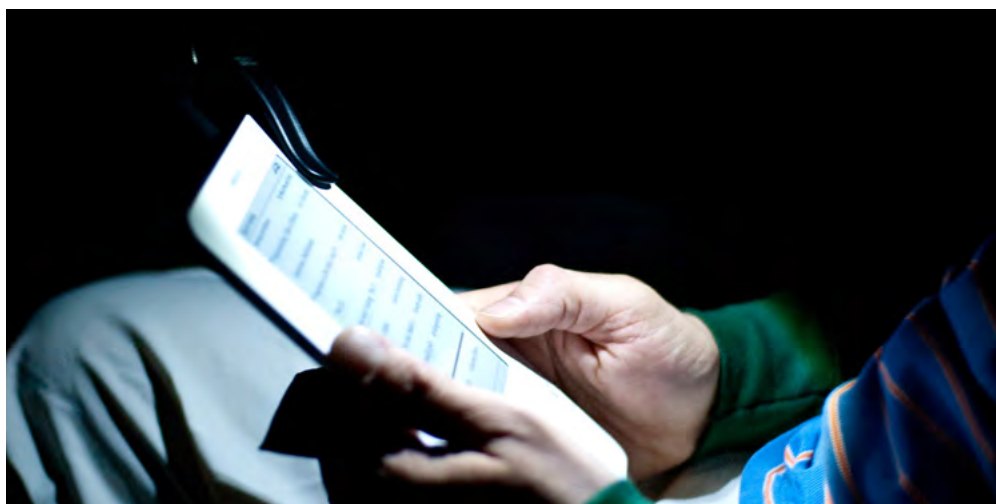
In 2014-2015, the BnF defined a new policy for increasing its audience, both on-site and online. The aim to engage a wider audience was naturally echoed in a digital lending project, available to users across the entire country. The digital lending service designed by the BnF could, in the future, be open to any reader with an annual subscription to the BnF (François-Mitterrand reference library and research library, Richelieu and other sites).

### A DEDICATED POLICY

The BnF can only design the content selection policy for its digital lending service by seeking complementarity with public and university libraries that already provide lending services and remote access to their digital resources. This strategy would be built as a continuation of that for reference library acquisitions and in accordance with its vocation as a heritage library: encyclopaedic and general reference works that remain useful way beyond the length of their availability in book stores. In the area of literature, the service could centre around stock works from contemporary publishers and classic works.

### A PARADIGM SHIFT

Increasing the BnF's audience involves a paradigm shift in the consideration of user needs and expectations. Digital lending could contribute to changing the image of the Library, still sometimes viewed as a "last recourse" library only addressing a limited group of privileged researchers, by proposing a high-quality offering content-wise, with a simple and efficient user interface. In addition, gathering anonymised usage data through the digital lending system would allow for a better understanding of users' habits, tastes and expectations, and, ultimately, for a collection development policy and service tailored to their needs.



**The digital transition of the publishing industry is accompanied by an adaptation in the library lending service.**



# 03.N

## Improved access to digitised newspapers

On Gallica, the “[Collections](#)” page providing access to the main daily titles is the most used of all the curated entry points. However, regular feedback of users’ expectations on Gallica’s newspaper pages led to the identification of several improvement opportunities.

### PRECISELY DESCRIBING THE CONTENT

The [Europeana Newspapers](#) project (2012-2014) offered valuable scope for study and experimentation (see also part 6 section B). Thanks to the full-text search with article segmentation, newspapers are now more precisely described and benefit from a content classification at page and article level (adverts, small ads, tables, illustrations, legends). Another type of advanced processing, the named-entity recognition (NER), consists of locating and categorising entities such as people, places and organisations. Uploaded first of all on a dedicated portal by TEL (The European Library), these enriched resources are designed to eventually feed into a dedicated thematic collection of the Europeana portal.

With 18 million pages processed (including 3 million from the BnF), Europeana Newspapers was the largest aggregation project for Europeana in terms of the amount of data processed, metadata converted and images incorporated.

### IMPROVING THE SEARCH INTERFACES

The improvement of search and textual navigation parameters would constitute another development. The search results report developed for Gallica is a step in this direction: this service, made available to researchers, facilitates the analysis of a specific full-text search by assisting the user in the formulation of their query, then providing them with the results in an easy to use form (cumulative statistical report, references and links, printable extracts from the text).

### CREATING NEW FEATURES

It would be relevant to draw inspiration from records created for each title in data.bnf.fr in the development of Gallica: nominal index of the

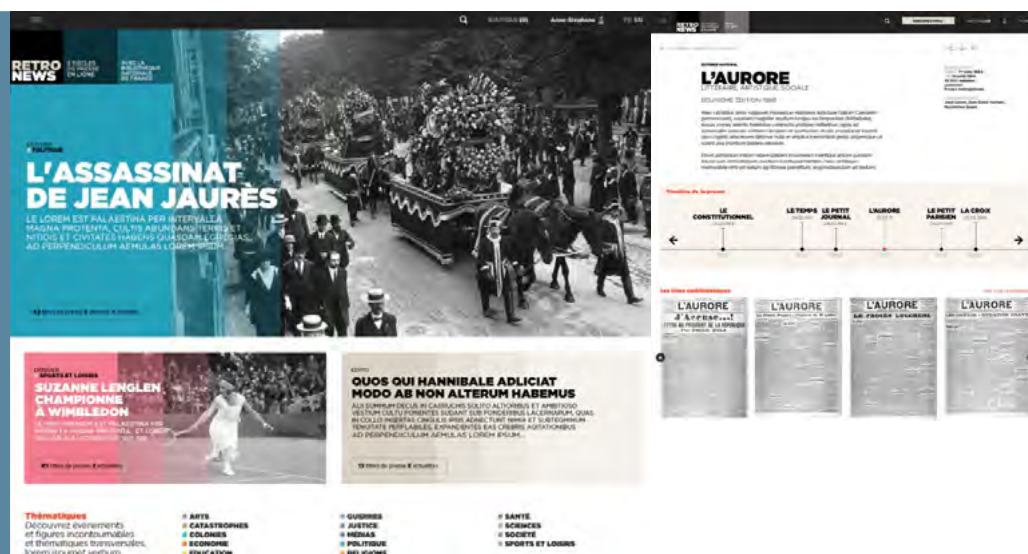
collection of titles available, introduction of a “newsstand” feature etc. The gaps in the digitised collections (absence from the paper collection, missing or defective digitised files) also need to be addressed. Interoperability developments could contribute to filling them by encouraging links to partner sites, whose collections complement those of the BnF, without the user leaving Gallica.

**Europeana Newspapers represents the largest aggregation project for Europeana in terms of the amount of data processed, metadata converted and images incorporated.**

## RetroNews, a dedicated portal

In partnership with Immanens and in collaboration with Syllabs, BnF-Partenariats (see part 1 section D) has designed an innovative service around digitised newspapers: [RetroNews](#). 3.5 million newly-digitised pages should be added to the 5.5 million newspaper pages already available on Gallica.

The paid, premium service will provide additional features for sharing and downloading, semantic search tools and curated content.



# 03.0

## Wi-Fi access

From 2016, BnF readers will be able to access the internet from all BnF sites via the Wi-Fi network, in accordance with the health precautions required in the use of this technology.

### FACILITATING MOBILITY

This service follows BnF developments since 2008 to offer readers cable internet access (with over 800 user seats equipped in 2010 in the research library and nearly 500 user seats equipped at the end of 2011 in the reference library, on the François-Mitterrand site). From 2016, it will enable BnF readers to work directly with their own equipment – whether PC, tablet or mobile phone – and to become “mobile” readers.

**In respect of health standards and in order to minimise the electromagnetic field, the maximum emission power of the Wi-Fi access points is limited.**

### ... OUT OF CONCERN FOR HEALTH AND SAFETY

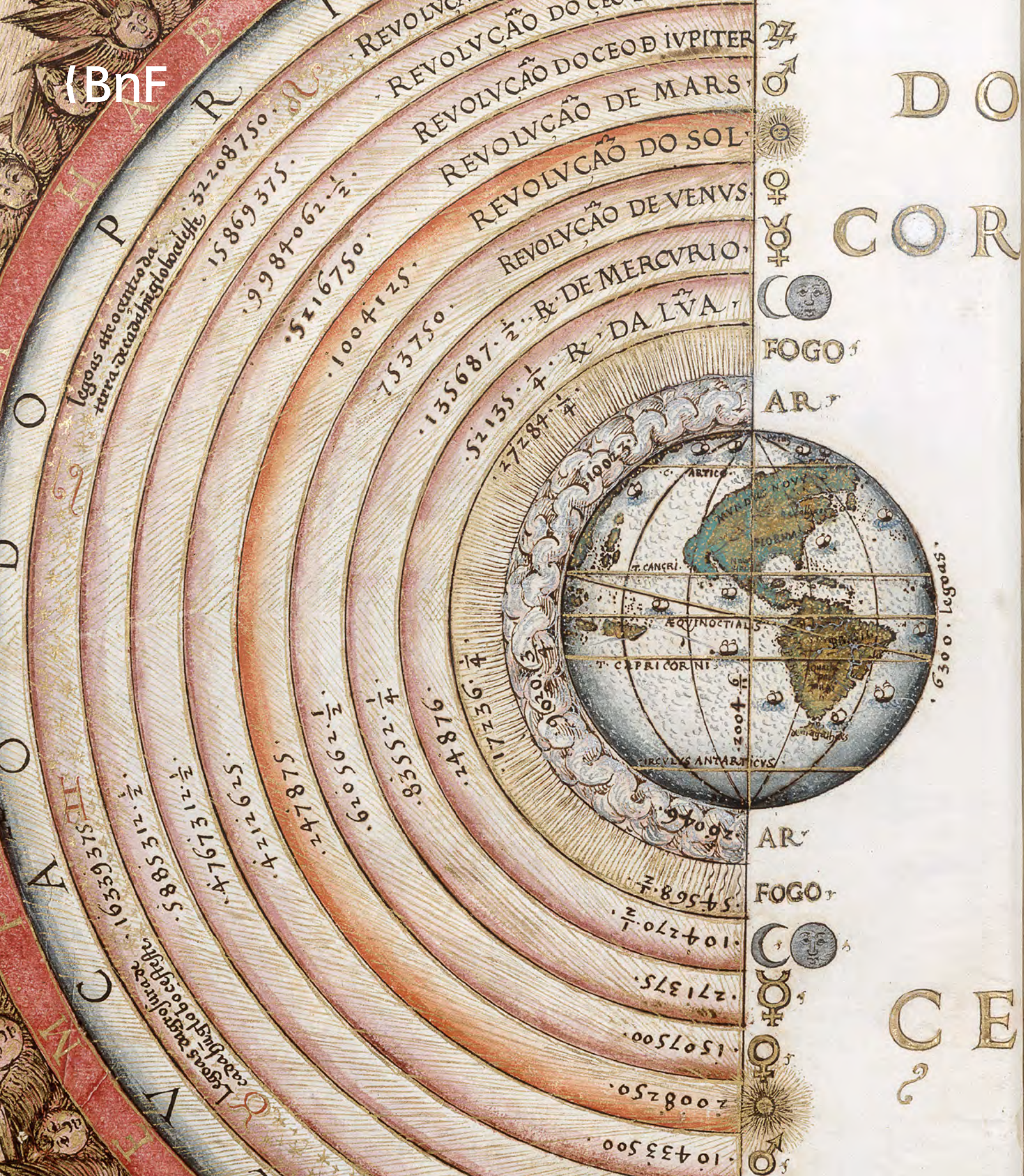
Wi-Fi will be available in all BnF public spaces (François-Mitterrand, Richelieu, Arsenal, Louvois, Opéra, Avignon), in the reading rooms, exhibitions, halls and auditoriums, with the exception of Wi-Fi free areas for people with electromagnetic hypersensitivity.

In respect of health standards and in order to minimise the electromagnetic field, the maximum emission power of the Wi-Fi access points is limited to 30 milliwatts (0 to 30 milliwatts, or 30% of the nominal wattage of antennas at 2.4 gigahertz and 15% at 5 gigahertz) and the distance between access point and user, or access point and BnF staff member, is at least three metres for a user seat and five meters for a reference desk. Clear signage will be put in place to indicate areas with and without Wi-Fi.





(BnF)







# 04

FEEDING  
INTO A NETWORK  
OF NATIONAL AND  
INTERNATIONAL  
PARTNERS



---

145

conventions agreed in 2015 with partner members  
of the National Cooperation Network

---

6 million

Euros granted in financial support to 176 digitisation partners  
between 2009 and 2014

---

30 million

records in the CCFr

---

+ 480,000

CCFr visits in 2015

---

3,366

collections referenced in the RNBFD

---

21

international digital cooperation projects

The BnF is taking advantage of the digital era to make its exceptional heritage an instrument for dialogue between regions and between cultures. The unique expertise that it has been able to develop in the production, collection, processing and conservation of digital collections allows it to play a leading role in international scientific and technical activity, by developing partnerships with other cultural organisations throughout the world, and particularly in French-speaking areas. Thanks to a dynamic national cooperation policy, the BnF is helping to unite digital collections across the country and to lead digitisation and cataloguing programmes (retrospective conversions, cataloguing, collection inventory etc.) across a vast network of partner institutions. The BnF is also conducting an ambitious research policy, contributing to its national, European and international status.

#### MAIN TOPICS

- 04.A\_ The national shared digitisation programmes
- 04.B\_ The French Union Catalogue (CCFr)
- 04.C\_ Third party archiving
- 04.D\_ International digital cooperation
- 04.E\_ Europeana, the European cultural heritage digital library

#### RESOURCES, TOOLS AND METHODS

- 04.F\_ Gallica vanilla
- 04.G\_ Involvement in international bodies
- 04.H\_ Online information for librarians and professionals

#### DEVELOPMENT OPPORTUNITIES

- 04.I\_ Nationwide access to web archives
- 04.J\_ The dissemination of documents digitised by other heritage institutions



## A national network of partners

In order to build a national digital heritage, the BnF has committed, since 2009, to cooperate on digital projects with several libraries and public institutions (local authorities, universities, research institutes) and private stakeholders (foundations, associations etc.). This approach, building on the [Digital roadmap for libraries](#) published in March 2010, places French libraries in a collective digital collection development project, via thematic and regional interest programmes.

The cultural and scientific challenges of this collective enterprise are many:

- offering the public easy access to national, regional and local heritage;
- making French language and culture more visible on the internet, by gathering efficiently a critical mass of data;
- encouraging the dissemination of digital resources and their appropriation by the public;
- making overlooked sources available to researchers.

The BnF has gradually widened the scope of this digital cooperation by offering partner institutions the option of putting online, via Gallica, their own digital heritage resources. The development of the Gallica vanilla service to create digital libraries for partners aims to share the investment granted for the development of the digital library. This service provides partner institutions with a digital library under their own colours, using Gallica as the interface and integrating their own digitised documents.

Gallica therefore acts as the hub of a digital partner ecosystem: 305 organisations with different statutes, sizes and very varied locations thus make hundreds of thousands of documents available on Gallica.

**The BnF has committed, since 2009, to cooperate on digital projects with several libraries and public institutions and private stakeholders.**

Finally, the BnF is the key stakeholder in the implementation of the Action Plan for Written Heritage launched in 2005 by the French Ministry of Culture and Communication. The BnF is hence continuing its policy of support for the digitisation of catalogues for collections that are either ancient, special or of local interest, thus contributing to the continued enrichment of the French Union Catalogue.

**réseau BnF**  
**coopération**



# The BnF at the heart of an international ecosystem

Foreign heritage collections are present in significant amounts in the BnF collections. These collections contribute, in a decisive way, to its universalism and constitute a powerful driver for international relationships. Yet they also give the Library a completely unique responsibility for the description, conservation, circulation and promotion of this universal heritage. They are now the basis of a cooperation founded on the idea of sharing these exceptional resources.

Since 2005, the BnF has therefore undertaken an ambitious digital cooperation programme with national and research libraries across the world. These projects grant access to a shared heritage through joint actions for promoting these resources, gathered online in thematic collections. Collections spread out across different partner institutions and countries are united again in their digital form and shine a different light on French history and heritage. Although several of these programmes have an essentially bilateral dimension ([France-Japan](#), France-Poland, [France-Brazil](#), [France-China](#), France-Russia), some will cover wider

**The BnF exercises a very specific responsibility for the description, conservation, circulation and promotion of a universal heritage.**

geographical areas with a strong historic relationship with France (libraries of the Levant, Indochina). Skill-based patronage with foreign institutions allows, moreover, for the description and digitisation of prestigious collections conserved by the BnF:

the International Dunhuang Project, which brings together several foreign collections from the Dunhuang site (China); the digitisation and e-publishing of old Chinese prints conserved in the major foreign libraries undertaken by the University of Shandong (China); the digitisation of the Ottoman newspapers in allophone languages led by the French Institute of Anatolian Studies and the SALT Research Turkish Institute.

Finally, major developments and global challenges of the digital transition also occur on an international level. Due to its recognised expertise, the BnF plays a leading role within those bodies where international innovation dynamics converge: [IFLA](#) (International Federation of Library Associations), [IIPC](#) (International Internet Preservation Consortium) and the [Europeana](#) foundation, among others.

## Relations with the research world

Areas of excellence for the BnF with regard to digital activities – preservation, digitisation, digital library, web archiving, semantic web, data curation etc. – have the purpose of serving a wide spectrum of professional stakeholders, who are all potential ambassadors for the BnF's expertise: students, researchers and documentation and information professionals or, more widely, all research and development professionals in the areas of services and industry.

Taking advantage of its partnership with five top laboratories (LabEx Arts-H2H, PATRIMA, CAP, PP, OBVIL), three top facilities (ÉquipEx Biblissima, Patrimex and Ortolang) and the [CNRS](#), the BnF has found its place in a national research landscape that is undergoing a complete re-organisation. Scientific

collaborations are being carried out and novel fields of study explored, creating over the last few years new opportunities to promote the significant digitised collections.

**The BnF has found its place in a national research landscape that is undergoing a complete re-organisation.**

Several research and development projects have therefore been initiated in partnership with multi-disciplinary teams. A digital cooperation model with the research and academic world was designed, leveraging the BnF's expertise in terms of new uses (digital reading habits, social networking, crowdsourcing etc.) and big data. These programmes deal with challenges such as semantic indexing,

data visualisation, audio-visual processing, automatic selective indexing and computer aided indexing (CAI). They aim to lay the foundations of a new shared culture: the digital humanities.





# 04.A

## The national shared digitisation programmes

The national digitisation partnership strategy has been formed according to a content selection policy guided by structuring principles: digitisation of printed works in French and in the languages of France within the public domain, or for which rights have been negotiated. The aim pursued is the creation of thematic or regional interest corpora, designed to be very visible online.

### SHARED CORPORA

French libraries are invited to participate in several shared digitisation programmes around thematic collections (legal sciences, art history, the First World War, children's heritage literature), collections of regional interest (history of the regions, local and regional learned societies) and more focused corpora, valuable on a documentary level (publications from national academies and societies, trench newspapers, clandestine newspapers of World War II Resistance fighters, immigration newspapers). The BnF also leads shared digitisation

programmes around non-printed resources (portolans, manuscripts), with the aim of virtually recreating dispersed collections.

### FINANCIAL INCENTIVES

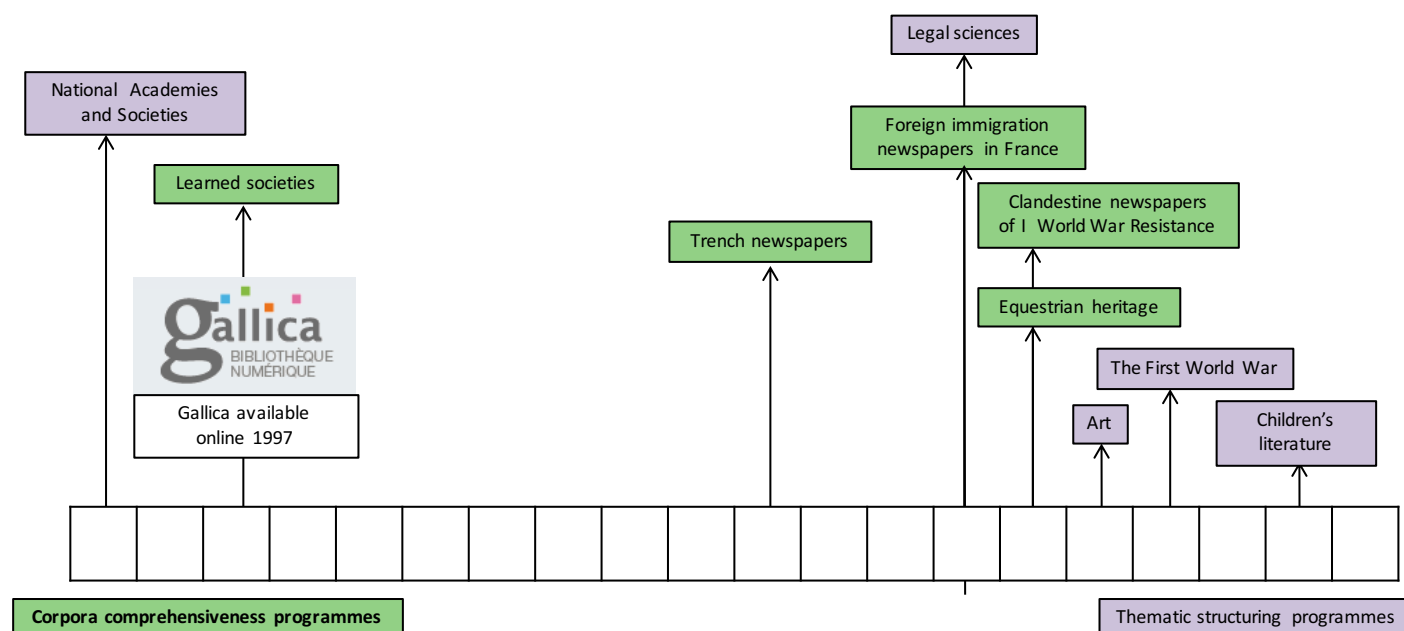
To widen the scope of this partnership work and ensure its success, the BnF has introduced a system of financial incentives – grants for the digitisation or integration on the BnF's printed material digitisation contract, 30% of which is dedicated to collections from French libraries.

The effectiveness of this measure is shown by the results obtained: between 2009 and 2014, 17.38 million printed pages were digitised by 176 partners, funded with 6.1 million Euros – work that will continue and increase over the 2015-2019 period. The development of the national digital context and of the work of the French Ministries of Culture and Communication, on the one hand, and Higher Education and Research on the other,

nevertheless requires optimal articulation of the financing measures to be sought.

Digitisation assistance is accompanied by increased demand for improved access to digitised resources (OCR, structuring, indexing) and shared promotion. In order to grant the greatest visibility possible to resources, special attention is given to optimising the circulation of digital documents.

**Digitisation assistance is accompanied by an increased demand for improved access to digitised resources and shared promotion.**





# 04.B

## The French Union Catalogue (CCFr)

As part of its cooperation policy, the BnF produces the French Union Catalogue (CCFr), which provides access to 30 million references held by over 3,500 libraries. By providing a single point of access to the main French catalogues, it increases the visibility of these libraries' resources and is a valuable working tool for researchers and book professionals.

### THREE AUTHORITATIVE COLLECTIONS

The CCFr provides three collections:

- the **Base patrimoine** (heritage database) describes and locates 5.7 million heritage documents held by 182 French libraries. This collection is constantly increasing due to retrospective conversions financially supported by the BnF, and to voluntary contributions and voluntary transfers.
- the **National Directory of Libraries and Collections (RNBFD)** contains information on over 5,000 libraries and over 3,300 collections, with links to digital libraries.
- Finally, the CCFr serves as a **national union catalogue of manuscripts** by providing a single point of access to the General catalogue of manuscripts in French public libraries (CGM), BnF archives and manuscripts (BAM) and Calames – the catalogue of manuscripts and archives in academic and research libraries.

**The CCFr is constantly increasing the amount and variety of data accessible via its [ccfr.bnf.fr](http://ccfr.bnf.fr) portal.**

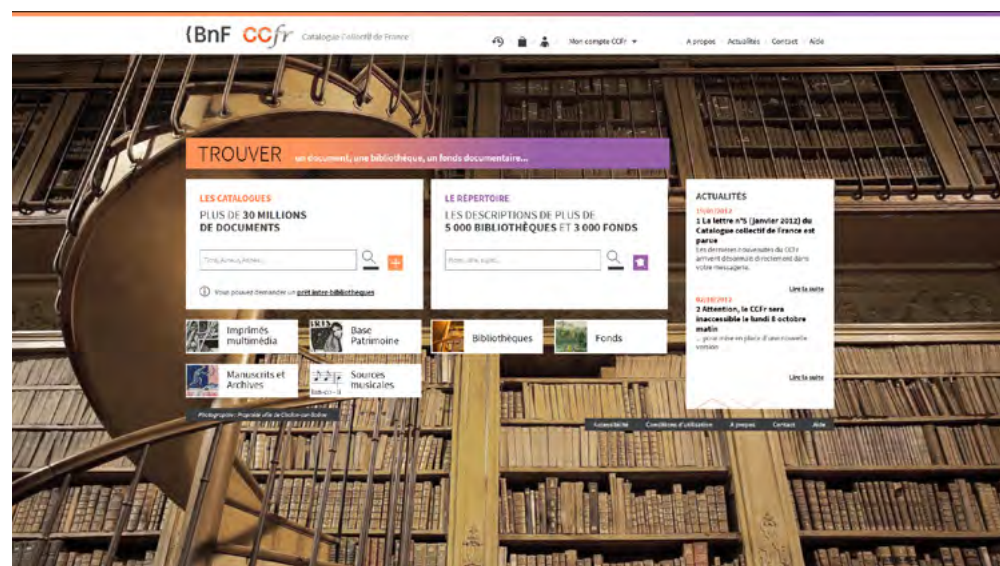
### A CONSTANTLY EVOLVING CATALOGUE

The CCFr is constantly increasing the amount and variety of data accessible via its [ccfr.bnf.fr](http://ccfr.bnf.fr) portal. It develops shared tools, allowing its partners to participate in the improvement and enrichment of resources (participative updating of collection and institution records in the RNBFD, a shared cataloguing tool in EAD (Encoded Archival Description) format in order to update the CGM), and reuse available data (exporting to bibliographic management tools and recovering records, permalinks, link resolvers etc.). Another notable development of the CCFr consists of multiplying the links to digital documents. Focused initially on manuscript databases,

this approach will enable access from the catalogues to a growing volume of digital resources. Next steps are the harvesting by search engines of all hosted data and the integration of data sets into specialised databases, like the Heritage Printed Books database of the Consortium of European research libraries (CERL).

### INCREASING VISIBILITY OF THE CCFr

In order for the institutions represented by the CCFr to benefit from it, it is essential to increase its visibility on the web. The new website's updated, user-friendly features (2016) should significantly contribute to a broader dissemination of the data it provides.



CCFr new interface (2016)



# 04.C

## Third party archiving

Since the creation of SPAR (see part 2 section B), questions have been raised regarding the system’s accessibility to other organisations confronted with the same conservation challenges and wishing to benefit from the BnF’s technologies and expertise. In this particularly dynamic area, which requires the regular updating of technologies, pooling is a solution for organisations which cannot shoulder the costs alone.

### BNF DIGITAL ARCHIVING

To respond to these challenges, the BnF now offers a [digital archiving service](#) for third parties with the same safety and sustainability guarantees as its own heritage collections, at a fair and moderate price.

### STRUCTURING INITIATIVES

In order to offer a quality of service comparable to that of the private sector, the BnF has undertaken several initiatives. Certification by the SIAF (Interministerial Department of the French National Archives) was obtained in 2013, allowing the BnF to position itself as a trusted operator. MEDONA (Data exchange Model for archiving), which has become standard NF Z44-022, is to be implemented in order to standardise the messages between electronic archive flow managers (transferring departments, archive departments etc.). Finally, the BnF has developed the *Espace Coopération* extranet (see part 5 section C) in order to facilitate transfer operations and data access for the beneficiaries of third party archiving.

The Centre Pompidou has been archiving its data in SPAR since 2013 (10 TB archived in 2015). The Picasso and Quai Branly museums have also chosen the BnF to ensure the archiving of their digital collections (first transfers in 2015).

While positioning itself as a trusted operator, the BnF is facilitating transfer operations and data access.

## Rates

Rates are established according to the amount of storage, the upload method of the data to be archived, the number of copies and the type of storage media, as well as the length of the contract.

Archiving on tapes offers a low-cost solution, but it is slow. Data stored on discs is available more quickly.

Income increased to 40,000 Euros in 2015.

Example of rates for a five-year contract, with minimal archiving on two tapes:

Bnf Digital archiving		
General rate, per terabyte per year in Euros excl. taxes, based over 5 years		
Volume of data to be archived over two tapes	extranet	custom
1 TB	€2,545	€3,990
2 to 5 TB	€1,960	€2,680
6 to 9 TB	€1,740	€2,160
10 to 29 TB	€1,590	€1,840
30 to 49 TB	€1,420	€1,530



# 04.D

## International digital cooperation

### ENCOMPASSING NEW GEOGRAPHICAL AREAS

The bilateral digitisation programmes encourage the creation of thematic collections based on documents spread throughout different countries and institutions. They therefore contribute to rebuilding linked historical and cultural heritages.

Since 2005, the BnF has been initiating various digitisation cooperation programmes, in the context of partnerships such as: [France in America](#), [France in Brazil / A França no Brasil](#), [France-Japan](#), “Sinica”, France-Poland, [Ottoman newspapers in French](#) etc.

The BnF also contributes to multilateral collaborative programmes ([World Digital Library](#), [the International Dunhuang Project](#)) as well as to multilateral programmes undertaken by national institutions (early-printed books from China and Taiwan). It is regularly asked to participate in new projects (the University of Shandong, Indian manuscripts, the Qatar Digital Library).

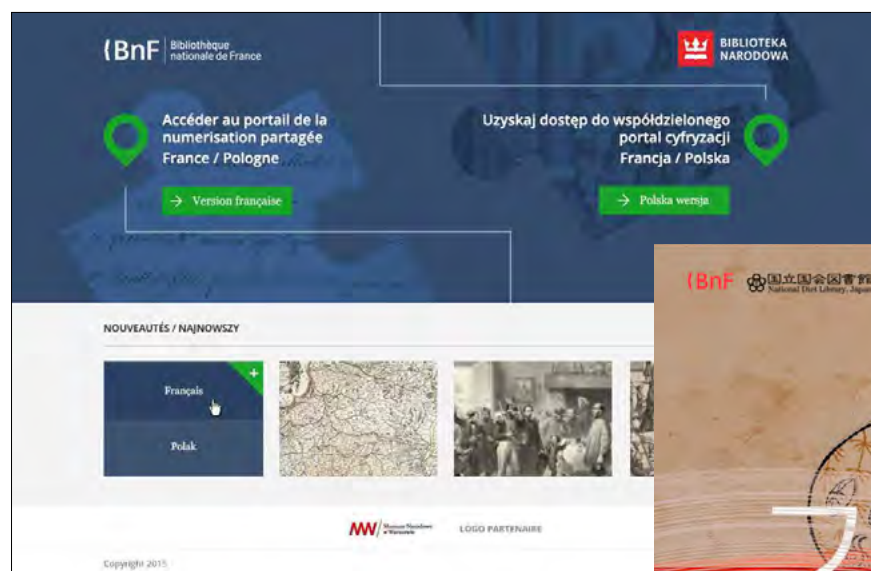
Some of these programmes aim to cover geographical areas with strong historical links to France, as demonstrated by the support provided for creating a digital library of French libraries

from the Levant (French Biblical and Archaeological School of Jerusalem, Dominican Institute for Oriental Studies in Cairo, Saint-Joseph Library in Lebanon etc.) or the development of a portal dedicated to French Indochina.

The content produced through these digitisation programmes is made sustainable by its integration into Gallica. The digital library's technological infrastructure allows for the gradual pooling of functional building blocks (content curation in Drupal, Gallica vanilla).

### BOOSTING THE STATUS OF THE FRENCH LANGUAGE

These thematic pathways enrich the BnF's cultural heritage offering for French and foreign researchers and enthusiasts, as part of the Library's user-oriented digital content management approach. They also reflect a resolute policy in support of French as a language of culture.



Examples of bilateral portals from the “Shared heritage” collection: France-Poland, France-Japan



# 04.E

## Europeana, the European cultural heritage digital library

Launched in November 2008, with the support of the European Commission, [Europeana](#) aggregates metadata on 50 million digitised objects (texts, images, audio, videos) from 2,000 European institutions (libraries, archives, museums etc.).

### A KEY PLAYER

The BnF's status as a founding member of Europeana, and the election in 2011 of BnF president Bruno Racine as head of the Europeana Foundation (renewed in 2013 and 2015) enable the BnF to play a major role in defining the new 2015-2020 strategy. Europeana seeks to provide a window into European cultural heritage and develop its portal as a data distribution platform used by the creative industries, data providers and ordinary citizens.

### SHARING THE BnF'S KNOW-HOW

The BnF strongly supports this development by providing Europeana with its expertise in terms of managing data and allowing the consortium to benefit from technological advances developed in the context of data.bnf.fr. It has long been the number one provider to Europeana, thanks to the sharing of metadata from all documents available on Gallica. It is also aggregator to Europeana for Gallica's French partner libraries, thereby encouraging the European dissemination of data produced in France.

### PROVIDING ACCESS TO MAJOR EUROPEAN WORKS

The BnF greatly contributes to enriching the portal's curated content through [Europeana Collections 1914-1918](#), dedicated to the First World War, and [Europeana Sounds](#), dedicated to Europe's audio heritage. The creation of a contextualised collection presenting the major works of European culture should, for its part, rely on the expertise developed by the BnF for *Les Essentiels de la littérature* (see part 3 section B).

### The BnF plays an aggregator role for French libraries to Europeana.



européana  
think culture

européana  
newspapers



européana  
sounds



# 04.F

## Gallica vanilla

The Gallica vanilla system (*Gallica marque blanche*) was born from a twofold desire: to pool the developments achieved through Gallica and enable the BnF's partners to benefit from them; and to complete the national collections by integrating documents digitised by other French libraries.

### A CUSTOMISED DIGITAL LIBRARY

Each Gallica vanilla partner has its own digital library. Except for the homepage, which is specific to each site, the query and document consultation process is identical and uses the same tools and technology as Gallica. The pages are personalised with the colours and logo of the partner, who may also configure the activation and deactivation of the various features offered.

This system will also enable the BnF to improve the service provided on Gallica by offering an enriched digital collection, also accessible on Gallica intra muros. Every document present in a Gallica

vanilla library has a record in the BnF's catalogues. Moreover, the BnF ensures the maintenance and hosting of each vanilla site, even if it has its own URL, independent of the bnf.fr domain.

### A STREAMLINED APPROACH

In 2015, the BnF began the streamlining and simplification of Gallica vanilla, in order to enable as many interested institutions as possible to benefit from this service. Further to this approach and to the technical and functional infrastructure of Gallica's new version, new projects can be created in 2016 within a simplified context.

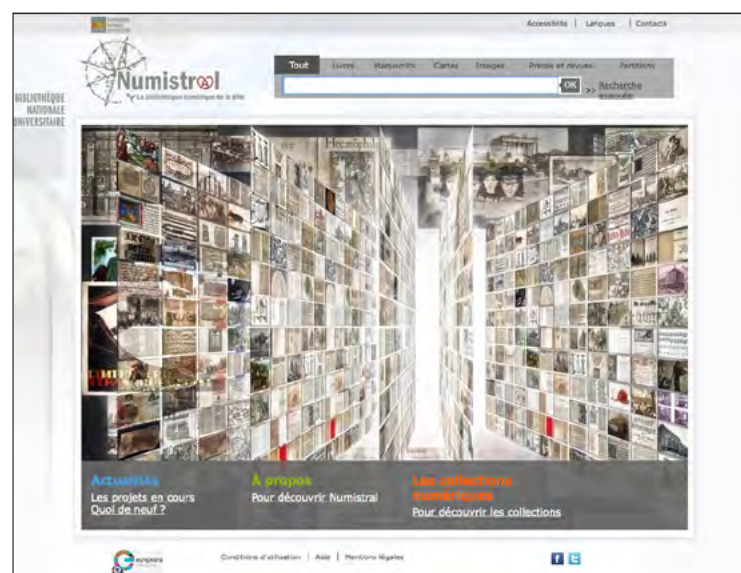
### SHARED BENEFITS

The Gallica vanilla system is a formidable opportunity for the BnF to develop its digital collection, while allowing other public stakeholders to benefit from

several years of IT investment and technological expertise. The development of future projects in Gallica vanilla will favour partners with an interesting documentary offering and a suitable technical infrastructure. Particular attention will also be paid to the diversity and type of partnerships.

An economic model has been defined, to ensure a balance between the investments allocated by the BnF and partners' contributions.

**Two Gallica vanilla websites:**  
[Numistral](#) (The National and University Library of Strasbourg, 2013) and [La Grande Collecte](#) (partnership with the Interministerial Department of the French National Archives, 2014)



# 04.G

## Involvement in international bodies

The BnF is highly involved in the major international heritage bodies, contributing in a significant way to the innovation and development of libraries, especially in the digital domain.

### AN ACTIVE INTERNATIONAL STAKEHOLDER

The BnF has played a major role in the governance of the International Internet Preservation Consortium ([IIPC](#)) since it was founded in 2003 and actively participates in that of the Virtual International Authority File ([VIAF](#)). Strongly involved with the International Federation of Library Associations and Institutions ([IFLA](#)) and the Association of European Research Libraries ([LIBER](#)), the BnF's expertise is especially called upon for issues related to web legal deposit, bibliographic standardisation, data exchange and even sustainable preservation. The BnF is part of the [IASA](#) (International Association of Sound and Audiovisual Archives) technical committee, which sets conservation format recommendations for audiovisual materials. As president of the Conference of European National Libraries ([CENL](#)), the BnF has worked

a great deal on the definition of new services that academic libraries could provide to researchers (data processing and mining).

### CONTRIBUTING TO STANDARDISATION

Finally, the BnF collaborates in decisive standardisation processes, notably within the International Organisation for Standardisation ([ISO](#)), such as the storage and preservation format standardisation of web archives, WARC (ISO 28500: 2009), or a technical report on the statistics and quality indicators of web archives (ISO/TR 14873).

**The BnF's expertise is particularly called upon for issues related to web legal deposit, bibliographic standardisation, data exchange and even sustainable preservation.**



## The French-Speaking Digital Network

[RFN](#) (the French-Speaking Digital Network) is a consortium of 25 French-speaking documentary institutions. Its mission is to ensure the preservation and accessibility of digital content in French. Digital cooperation is one of its main tools, along with expertise and training. To reach its target, the consortium has developed a new strategic plan around two key objectives – to support and

accelerate digitisation and specific digital library creation programmes; and to develop the portal using curated content with high added value. The BnF significantly contributes to this scheme by supervising the technical development and enrichment of the portal. Moreover, it is committed to developing, from 2016 onwards, the RFN's collective digital French-speaking Library via Gallica vanilla.



# 04.H

## Online information for librarians and professionals

Online since 2002, the “[Professionals](#)” section of the website is openly accessible from the [www.bnf.fr](#) institutional website but targeted at professional users, thereby allowing the BnF to play its role as leader of national and international networks. The section is intended for librarians and book and conservation professionals, but also for other communities like archives, the web and digital industries. It is used by experts as an information source, but also by students and practitioners for their professional development or to respond to specific needs.

In the current context of open data, digital transition and skills pooling, the section acts as a repository of resources, useful for better understanding the BnF “from the inside” and for helping professionals keep track of developments within their field.

Through the “Digital innovation” section of the [www.bnf.fr](#) website, the BnF shares its experience of handling major digital changes.

### REFERENCE INFORMATION

The BnF shares its librarianship expertise in a variety of areas. The “Cataloguing and indexing” section refers to cataloguing standards, formats and their evolution. The “Preservation” section contains valuable information on emergency plans to safeguard collections in the event of an accident. The “National cooperation” section gives information on cooperation policy and programmes.

### FOCUS ON “DIGITAL INNOVATION”

The BnF shares its experience of handling major digital changes in the “[Digital innovation](#)” section. The different themes in this section (digital legal deposit, Gallica, digitisation, digital preservation, records management and linked open data) are organised according to the same model: a general presentation of the challenges, a “tool box” with reference documents and formats used by the BnF and a “Cooperation and implementation” section, presenting the BnF’s R&D projects and key services such as SPAR, the digital repository, and [data.bnf.fr](#).

### NUMEROUS SERVICES

The professional pages also give practical information on the services provided by the BnF in the context of its oldest missions, like legal deposit and data dissemination. Placed under the open licence of the French State (*Licence publique de l’État*) since 2014 (see part 2 section A), bibliographic and authority data from the BnF is now freely reusable. In response to growing demand, the “BnF data services” section provides practical files and technical information on download methods.

### RECORD OF PAST AND FUTURE PROFESSIONAL MEETINGS

Finally, this section promotes professional news from the BnF via news flashes and a calendar. The proceedings, communications and video recordings of the main conferences organised by the BnF are accessible there, organised by theme and year. All proceedings from the national cooperation conferences can be found here, as well as recordings from the joint annual conference with AFNOR on information standardisation.

## Children’s literature: a new website for professionals

In 2016, [cnlj.bnf.fr](#) will become the renewed site for activities, training and resources from the French National Centre for children’s literature (CNLJ), for book promoters, researchers and professionals in children’s and youth literature. The site will continue to focus on general news from the sector and will provide a large amount of online content including specialised journals (*La Revue des livres pour enfants*, *Takam Tikou*), bibliographies etc. These resources are also designed to be accessible to the general public.



# 04.1

## Nationwide access to web archives

The issue of collecting web archives raises that of their visibility. It is important that the collections are better publicised, ensuring that they are referenced online and that focused crawls are highlighted on data.bnf.fr.



### STRENGTHENING THE NETWORK

Libraries in charge of legal deposit by the printer in the region now also provide access to web archives. This network ensures good interconnection over the national territory. The 2014 decree defining the list of bodies authorised to implement on-site consultation of communication services to the public online, collected in terms of the legal deposit, authorises the BnF to deploy this access. A specific training programme has been provided to partners, along with suitable communication media. The libraries of Montpellier, Nancy, Strasbourg, Lyon, Limoges, Toulouse and Bordeaux benefited from this in 2014-2015.

**In each library in charge of legal deposit by the printer, access to web archives occurs in the same way as at the BnF. It is therefore reserved for accredited researchers.**



### SECURING ACCESS TO DATA

In accordance with the legal provisions, each library in charge of legal deposit by the printer must reserve access to the web archives for accredited researchers.

To guarantee security access to these collections, the BnF uses virtualisation – access occurs remotely, directly on the BnF servers, without the need to create a local copy of the data.

### HARVESTING THE LOCAL INTERNET

This access plan completes an active cooperation policy with the regions around the theme of web archiving (see part 1 section A). Since 2004, a shared selection of websites to be harvested has been organised during electoral campaigns (21 participating libraries for the 2015 regional elections). The National and University Library of Strasbourg was the first to participate, in a sustainable way, with the creation of a local and regional interest web collection: “Alsatiques du web”. Remote access to all these archives makes available within the region, for accredited readers, a key local heritage resource (also including the regional daily newspapers harvested in digital form).

Training in web archive consultation at the Nancy multimedia Library

# 04.J

## The dissemination of documents digitised by other heritage institutions

Among the processes that contribute to enriching Gallica, the ingest of resources digitised by other heritage institutions is an efficient way to quickly increase the digital collections.

### TWO INGEST METHODS

Gallica provides two ingest methods for external digital resources:

- **Interoperability through Metadata Exchange, via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH):** in November 2015, over 70 digital libraries were referenced in this way on Gallica;
- **The integration via deposit of digital files:** this solution, trialled since 2010, is for partners without a digital library or that wish to use Gallica as their own digital library. Initially requiring individual document ingest, the system was subject to a mass integration experiment within the Numistral project, undertaken with Strasbourg's National University Library.

Although interoperability proved particularly easy to implement, the integration into Gallica of files digitised by French libraries remains a complex process, the technical criteria for managing these files being selective, in order to ensure the best circulation conditions on Gallica and the best conservation conditions in SPAR. So that the files can be integrated on mass scale, special attention must be paid to the accuracy of the metadata, their standardisation and their degree of precision.

**The integration of files into Gallica has a lot of potential, providing partners with a solution for the conservation and circulation of their digital collections.**

### CIRCULATING AND MAKING AN UNEXPLOITED HERITAGE SUSTAINABLE

The integration of files into Gallica has a lot of potential, providing partners with a solution to a problem identified several years ago – the conservation and circulation of their digital collections. Although funding programmes for digitisation has allowed a certain number of institutions to reproduce their collections, they do not all necessarily have a digital library or a distribution interface, and many of these collections constitute a “dormant heritage” which should, at last, be circulated and promoted.

This valuable content is presented on the pages dedicated to partners on Gallica. It can be viewed and used with the help of all the features of the digital library, like the 9 million printed pages from one hundred partner libraries, which were digitised by the BnF and are currently circulated on Gallica.



Gallica page dedicated to the French National Assembly resources



{BnF







# 05

SIMPLIFYING  
TOOLS AND  
PROCESSES FOR  
THE BNF AND  
ITS PARTNERS

---

## Exponential progress

---

**20.000**

documents digitised in 1998

**350.000**

documents on Gallica in 2007

**3.5 million**

documents on Gallica in 2015

---

Between 2005 and 2015,  
an increase from

**16 TB**

to over

**668 TB**

of web archives

---

The uptake of digital activities in the first decade of the 21<sup>st</sup> century has been accompanied, at the BnF, by a huge change in scale. With mass digitisation, the number of digitised documents increased tenfold between 2007 and 2015.

At the same time, the web archives, which represented just 16 terabytes of data in 2005, have also undergone a sharp increase (180 terabytes at the end of 2009 and 668 terabytes at the end of 2015).

Exponential growth of the same nature is expected for born-digital documents to be ingested via legal deposit, including thousands of e-books and 178,000 music albums now mainly sold in digital form.

This considerable increase in the objects produced and collected, due to the effect of digital technology, has had several implications, including the need to set up a suitable system for the ingest of a large amount of documents, automated monitoring for physical and digital document flows and quality control processes. It has also ●●●

---



- 
- become essential to answer requests from partners, whether institutional or from the business world, who expect fluidity and responsiveness from the BnF to absorb their data flows, to connect their own production or monitoring tools to those of the Library and to facilitate automated interactions between the systems. Moreover, the BnF's adaptation to changes introduced by digital technologies questions the legal framework, both in terms of ingest (for instance e-books) and processing (data and metadata mining), as well as in terms of accessibility (consultation on-site or remotely, in a secure environment).

Faced with these challenges and in a context of limited financial and human resources, the BnF has created a set of tools to adapt and simplify its internal workflows, as well as its exchanges with third parties (publishers, libraries, and service providers).

---

#### MAIN TOPICS

- 05.A\_** The digitisation workflow: processes and system
- 05.B\_** The legal deposit extranet
- 05.C\_** The *Espace Coopération* extranet
- 05.D\_** PLATON: Access for people with disabilities

#### RESOURCES, TOOLS AND METHODS

- 05.E\_** The document management system
- 05.F\_** Web archiving tools
- 05.G\_** Quality management

#### DEVELOPMENT OPPORTUNITIES

- 05.H\_** Digital collection management tools
- 05.I\_** Ingest of purchased and donated digital documents (ADDN)

# The digitisation of administrative processes

The BnF has anticipated, since the middle of the 1980s, the impact of digital technologies on the majority of library processes, starting with cataloguing which has been computerised. When the move from individual management of internal administrative documents to a more collective organisation was required, in the middle of the last decade, to address the growing number of cross-disciplinary projects combining various internal competencies, the BnF committed to the dematerialisation of processes considered mature: call for tenders, relationships with certain service providers, staff identity and equipment management, revenue and expenditure workflows etc. The BnF was therefore the first public administrative institution in France to exclusively manage digital documents within its financial management workflow, thus saving one and a half tons of paper each year and shortening payment time frames by six days on average per dossier.

These projects sought to reconcile the strategic interests of the BnF with the proper functioning of individual day-to-day activities:

**The BnF is the first public administrative institution in France to exclusively manage digital documents within its financial management workflow.**

1) By facilitating the creation and transmission of documents as well as all the confirmation stages; 2) By ensuring the administrative and legal validity of documents and data, in order to preserve the rights of people and the BnF while ensuring administrative transparency; 3) By managing the life cycle of the documents and data therefore created, with accurately defined periods of time before elimination or preservation.

The management of office documents is highly representative of this development. Since 2004, a records management programme on a considerable scale affected all of the BnF's staff and activities. It drew the framework of a new human and technological organisation, allowing for the management of born-digital documents from their creation – email

included –, and able to ingest digital documents on their arrival at the BnF. This specific production workflow creates a new collection within the BnF's digital heritage – its public archives, eventually preserved in SPAR via the “Administrative archives” channel.

# The extranets

Numerous processes, whether they come from cooperation, the legal deposit or other activities, involve stakeholders outside of the BnF e.g. publishers and legal depositors, libraries and partner institutions, and even service providers within the context of public contracts. The BnF has therefore developed expertise in the creation of reserved access platforms dedicated to exchanges with these third parties – the extranets.

The [Extranet website of the legal deposit](#) provides an online declaration service designed to facilitate the legal deposit process for publishers and document producers. Bibliographic data is created by the depositors and, most often, accompanied by cover images. This data is transferred to the BnF's information system with an electronic receipt sent instantaneously in return. If the publisher agrees, the data is published in the form of an advert on the [Nouveautés Éditeurs](#) site.

Another platform, the [Espace Coopération](#), was created in 2011 as a working tool to support BnF partners in their day-to-day administrative processes (agreements, signatures etc.), for displaying their collections on Gallica and for subscribing to the third party archiving service. The *Espace Coopération* also provides a discussion forum and a news feed.

Finally, other extranets dedicated to specific functions have been implemented, one example being the [PLATON](#) platform, used for the transfer of digital files between publishers and approved associations in charge of adapting works for disabled users, under the exception planned for in the French intellectual property code.

**The extranets play an essential role in the BnF's digital transition, by simplifying and digitising the processes.**

These extranets play an essential role in the BnF's digital transition. By simplifying and digitising the processes, they have opened the door to a profound transformation of the BnF's working methods, for instance with the automatic integration of metadata from publishers into the catalogue. By strengthening the BnF's image of modernity and trust, they enable new items to enter the collections, such as e-books and retrospective digitisations from partner libraries. Finally, automation sometimes proves to be synonymous with savings. Collecting the regional daily newspapers by machine and thus transferring display media to a digital environment has reduced the yearly budget for microfilm conservation by 300,000 Euros.

These trends should lead to a convergence of the interfaces and tools for BnF partners. Eventually, a publisher who provides a digital file for the legal deposit, accompanied by its metadata, will access a referral system, allowing him to insert his document onto the PLATON platform in the prospect of an adapted publication, to authorise the circulation of a summary and cover in *Nouveautés Éditeurs* and in the catalogue, and to be assured of its sustainable preservation in SPAR. He may even authorise partial consultation of his content on Gallica, in order to increase its visibility. In this regard, the BnF seeks to pursue experimentation with e-distributors, to provide visibility on Gallica for recent e-books. Publishers could be given the option of authorising the display of their e-books on Gallica under the same conditions as those digitised under the ReLIRE project, (open access to the first 15% of the item with a link to the bookstore platforms), in addition to being displayed on Gallica intra muros.



# Digital collection management tools

For each new digital channel, an optimal coverage of the workflow is created, from document ingest to display, including conservation and cataloguing.

Some of these tools, like those used to manage web archives from crawling to reading room access, were designed and created in a collaborative way by international associations of library professionals who all face the same challenges. Others, like CEDN (digital objects ingest system), are built specifically to meet the needs of the BnF, but are also based on existing components of the Library's information system.

One of the major issues encountered when managing digital collections is the loss of a concrete relationship with the docu-

ment for the librarian handling it. It is no longer a tangible medium that can be touched, moved and stored – the tasks associated

with managing digital collections are often limited to monitoring data flows via statistics or log files. Some of these activities are now done by computer scientists or production engineers and no longer performed by librarians or stack attendants. Creating tools which will, in the future, allow librarians to keep full control over their collections is both a challenge and a requirement. Projects like SIPIL (Steering IT system) or the future digital collection publication tool will bring together the necessary indicators to conduct day-to-day digital

activities, progressively leading to the emergence of a new digital librarian tool box.

**Creating tools  
which will,  
in the future, allow  
librarians to keep  
full control over  
their collections  
is both a challenge  
and a requirement.**

# 05.A

## The digitisation workflow: processes and system

The management of document flows to be digitised relies on something called an “ingest system”, which combines conservation and stock management applications (initially developed for the processing of printed collections: de-acidification, binding etc.) with new applications designed to manage the exchanges between the BnF and its digitisation service providers and to allow visual tests before uploading.

### THE MAIN STAGES

An automatic testing of documents is implemented upon ingest. By confirming the format and consistency of the metadata produced, it is able to identify the processing required to upload the documents onto Gallica. At the level of this “ingest system”, digitised copies of the documents are catalogued, and images undergo various transformations to allow for zooming, thumbnails etc. The archiving in SPAR is also programmed here, following a staggered timetable.

### A SCALABLE SYSTEM

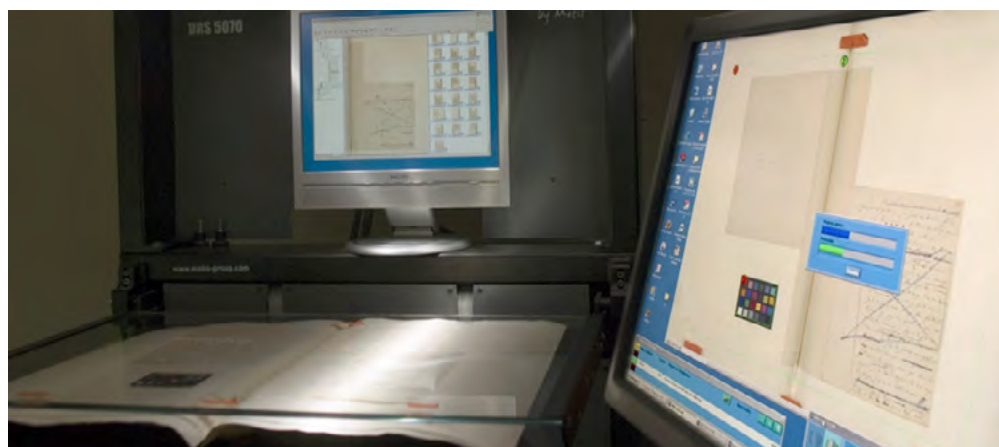
Since the first contract in 2007, the areas and purposes of digitisation have developed considerably. New types of document were targeted (newspapers, special printed books from *Réserve* and Arsenal, manuscripts, prints, maps, sheet music, objects etc.), additional processing was implemented (addition of OCR, EPUB on an already digitised document), the “emergency on-demand digitisation” channel was introduced, and various technical developments were integrated (increasing the resolution from 300 to 400 DPI, OCR processing, audio and video material).

### REDUCING THE TECHNOLOGICAL BARRIERS

These changes aim to make the system more flexible, as it has become too rigid at a time when the BnF is extending its partnerships and projects. It is also essential to reduce the technological barrier to entering digital markets, which can lead to an

excessively complex workflow and create a roadblock for certain service providers. The technical redesign of the “digitisation system” application, started in 2015, represents a decisive strategic challenge for the BnF. The Library must seize this opportunity to assess the increased need for flexibility, which emerged nine years ago, and to adapt the monitoring and automation tools that play an ever more crucial role in managing the mass of digitised documents.

**The technical redesign of the “digitisation system” application, started in 2015, represents a decisive strategic challenge for the BnF.**



Digitisation  
of a manuscript

# 05.B

## The legal deposit extranet

The [legal deposit extranet](#) website provides an online declaration service designed to facilitate legal deposit for publishers and document producers. It affects all types of documents subject to legal deposit with the BnF.

In 2015, the legal deposit extranet was used for 75% of map and plan deposits and 52% of printed book deposits.

### A DIGITISED DECLARATION SERVICE

Bibliographic data is entered by the depositors, usually accompanied by cover images. This data is recovered by the BnF's information system and electronic receipts sent instantaneously in return. If the publisher agrees, the data is published on the [Nouveautés Éditeurs](#) site. Since 2014, automatic metadata transfer, from base to base, has been implemented for the main publishing groups. Descriptive data is sent each week via automated flow by the distributors, who send the printed book packages at the same time.

### THE ADVANTAGES

Reducing document recording and receipt delivery timeframes makes the extranet fully adapted to the needs of publishers, while providing enriched records and ISNI-formatted author identifiers (see part 2 section F). It promotes close collaboration between the BnF and publishers in the area of metadata, especially regarding ONIX – an XML format used by publishing professionals all over the world and one that the BnF has been able to implement for this purpose. In return, the BnF catalogue is enriched with new data that has high added value.

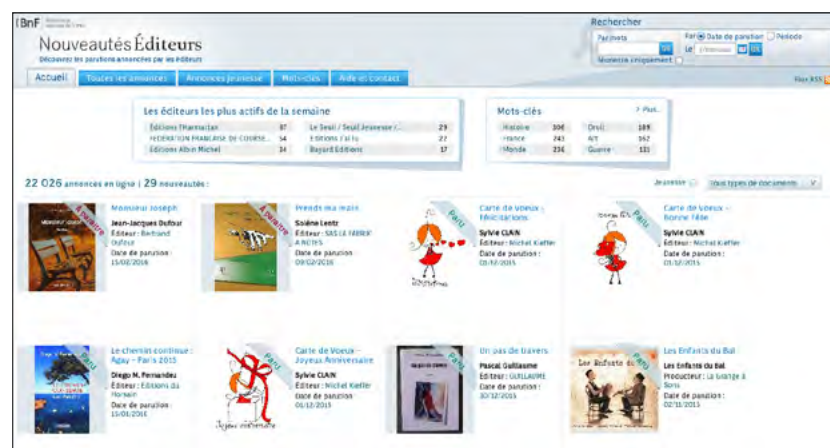
### EXPANSION TO E-BOOKS

The features of the legal deposit extranet will be extended to receive the legal deposit of e-books. The e-distributors will provide e-books in the EPUB or PDF format as regular data streams, without protection measures (DRM), accompanied by their metadata in the ONIX format. After automatic file testing, the legal deposit extranet will ensure ingest of the metadata, the publication of descriptive records, if required, on *Nouveautés Éditeurs*, the recording of the documents' deposit and the creation of records in the BnF's general catalogue (see part 1 section H).

The legal deposit extranet will evolve to allow for the legal deposit of digitised sound material as regular data streams. The automated ingest process for digital files and their metadata should enable the BnF to have over 170,000 music albums entered into its collections, which are only published online and which currently bypass the audio-visual legal deposit processes. The deposit of digital files individually will also be considered: depositors will be able to attach their files to declarations entered online by themselves.

**In 2015, the legal deposit extranet was used for 75% of map and plan deposits and 52% of printed book deposits.**

**The *Nouveautés Éditeurs* website, supplied by the legal deposit extranet**





## 05.C The *Espace Coopération* extranet

The network of BnF partners (libraries, publishers, online distributors) has expanded over the years across the country. It has also become more complex, with the creation of partnerships with multiple cooperation purposes. A communication, document exchange and partner collection management tool on Gallica was essential for the BnF, but also for all of its partners, in order to easier and more simply unite digital collections across the country.

### A WORKING TOOL

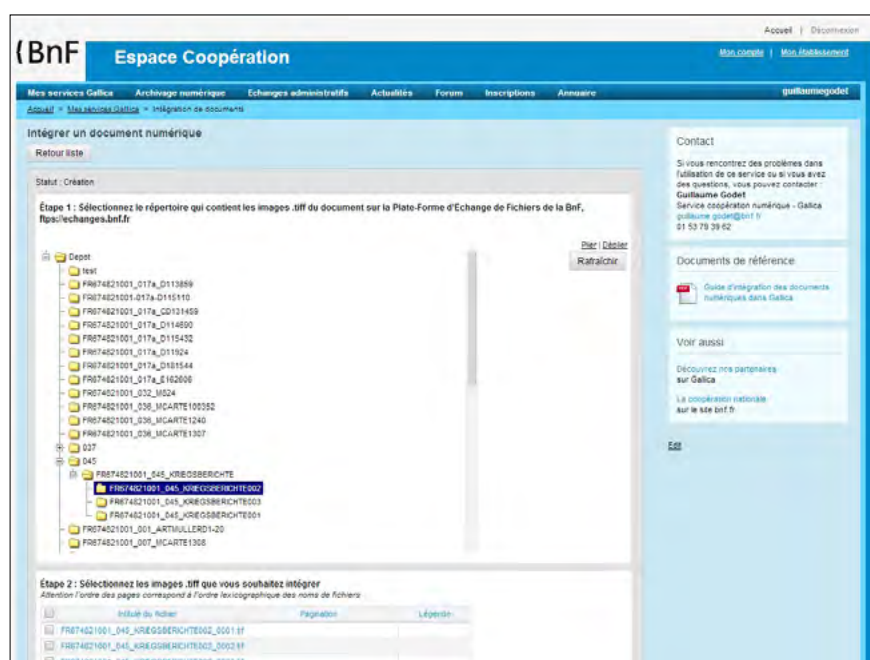
Launched in 2011, the [Espace Coopération](#) is a working tool which assists the BnF's partners on a day-to-day basis, as well as the customers of third party archiving, simplifying their tasks and allowing them to be informed and even discuss on the forum. The "Mes services Gallica" (My Gallica services) section enables the configuration of partner

OAI-PMH repositories for referencing documents on Gallica, the monitoring of documents sent to the BnF within the digitisation of printed works contract, and the integration of digitised documents by partners. The third party archiving section includes the monitoring of collections conserved by the BnF as part of its "BnF digital archiving" service, as well as search and retrieval functions for the archived content.

**One of the major developments of the *Espace Coopération* is to simplify and expand the dissemination on Gallica of BnF partners' digitised documents.**

### DEVELOPMENT OPPORTUNITIES

One of the major developments of the *Espace Coopération*, currently underway and to come, is to make it one of the main entry points for the integration of digitised documents by partners, either for circulation purposes on Gallica (see part 4 section J) or with a view to creating a Gallica vanilla website (see part 4 section F). In the years to come, the *Espace Coopération* must develop new services in accordance with the BnF's cooperation policy and the Gallica vanilla project, in order to allow partners to manage, as well as possible, the selection, production, monitoring, circulation and archiving of their digital collections, as well as their administrative exchanges with the BnF.



Integration of partner's digital files with a view to their consultation on Gallica and on their vanilla website

# 05.D

## PLATON: Access for people with disabilities

The French law of 2006 relating to copyright and related rights in the information society created an exception for people with a disability. The reproduction of works in a form adapted to disabled users can be undertaken freely, and without financial compensation, by associations and organisations jointly approved by the French Minister responsible for culture and the French Minister responsible for disabled people.

### A DEDICATED PLATFORM

In order to exercise its role as a trusted third party in the application of the disability exception planned for under the French intellectual property code, the BnF has developed the [PLATON](#) platform (digital work transfer platform) which allows approved bodies to carry out requests and publishers to deposit their files. These bodies may then access the files and register their adaptations, in partnership with the Adapted Publishing Database (BDEA – *Banque de données de l'édition adaptée*), administered by the French National Institute of Young Blind People (INJA – *Institut national des jeunes aveugles*), using an interoperability solution developed for this purpose.

### A VERY POSITIVE IMPACT

Since its launch in 2010, PLATON has led to the significant development of accessible content for disabled people, mainly those with a visual impairment. More than 40 approved bodies and 700 publishers use this service, completing



**Since its launch in 2010, PLATON has led to the significant development of accessible content for disabled people.**

an average of 6,000 to 8,000 adaptations per year. Subscription to a large selection of journals and periodicals has been possible on PLATON since 2013 via two newspaper aggregators. Also, the “*Rentrée littéraire accessible*” (Autumn literary season accessible to people with disabilities) has been held since September 2013. Thanks to the support of the CNL and exemplary collaboration between stakeholders, the event’s 2015 edition made 330 new literary works available to people with disabilities – 42% more than in 2014.

### A secure process

The approved bodies have the option of requesting, in the 10 years following the legal deposit of a printed work and provided its legal deposit occurred after 2006, that digital files used for printing these works be made available to them by the publishers via the PLATON platform.

The BnF acts here as a trusted third party. Once the adaptation is carried out (adaptation in large-type, Braille, adapted digital publishing etc.), the source files are destroyed by the adapting bodies, but kept indefinitely by the BnF.

Other developments should further increase the effect of PLATON in the future, especially the deposit of adapted files, coupled with the easier communication of these files to the approved bodies. In the near future, school textbook files should be subject to a systematic deposit, allowing disabled students to obtain copies as soon as possible. Finally, the BnF plans to offer publishers, on a voluntary basis and over time, a correlation between the legal deposit of e-books and the provision of files on PLATON.

# 05.E

## The document management system

---

The BnF started to use born-digital documents with the development of office automation by French administration. However, by 2004 it became necessary to move from the individual management of documents to a more collective organisation, to cope with the increase in cross-disciplinary projects combining a variety of skills.

### FIRST STEPS

The records management programme provided the framework for a human and technological organisation, enabling the management of born-digital documents from their creation, including email. In 2006, a documentary production system was launched, encompassing all stages from document creation through revision, version management and validation to life cycle completion. The data forms the public archives produced by the BnF, in part conserved in SPAR via the “Administrative Archives” channel.

### COMBINING STRATEGIC AND OPERATIONAL CONCERNS

An effective system needs to combine the strategic interests of the BnF with the proper functioning of individual day-to-day activities. The Library’s knowledge assets and memory are thus preserved and transmitted. Knowledge, experience and information are held here. Properly described and correctly classified, the documents created in electronic form are easily exploitable (according to the access rights granted). Their administrative

and legal validity is, moreover, better established, for the benefit of people and administrative transparency.

### ASSISTING CHANGE

An updated document management system, currently being deployed, is improving interface user-friendliness and intuitiveness and provides user options for independent learning. The system’s fundamental features remain, for their part, unchanged: a classification scheme, a list of typologies associated with a template repository, a unified documentary catalogue, an automatically attributed, permanent and unique identifier for each document or file, including versioning. In order to simplify and accelerate user tasks, all metadata that can possibly be supplied automatically are so supplied, allowing editors to concentrate on the essential – their current activity.

The implementation of such a system required significant training and adaptation by the BnF’s 2,400 staff. They needed to learn how to choose the correct categories in the classification scheme, what naming rules to use, how to send the information internally or outside of the BnF. This assistance for change had a very positive outcome, with the vast majority of staff supporting such an evolution and RM (records management) teams getting involved in the BnF’s strategic projects. It is now necessary to further improve tool user-friendliness, continue the DMS implementation within professional processes and develop advice and support for teams.

---

**Knowledge, experience  
and information are  
held here. Properly described  
and correctly classified,  
the documents created  
in electronic form  
are easily exploitable.**

---



# 05.F

## Web archiving tools

The implementation, between 2006 and 2013, of a comprehensive web archiving system meant developing or integrating a panel of specific tools covering the different stages of the process. Created following a modular approach and compliant with international standards, this workflow relies as much as possible on open source software, most often developed within the International Internet Preservation Consortium (IIPC).

### SELECTING

*BnF Collecte du web* or BCWeb (Building Collections on the Web), the curator tool developed by the BnF to select websites, has evolved into an open software project involving multiple stakeholders. The information contained in BCWeb is reused for cataloguing purposes, especially via data.bnf.fr.

### HARVESTING

The BnF uses the Heritrix web crawler, itself run within NetarchiveSuite – a complete web archiving software package which schedules harvests, monitors their progress, and ensures data quality control. The technological update of these tools, with a view to improving the management of audio-visual distribution platforms, social networks and even applications, is undertaken collectively by the members of the IIPC.

### CATALOGUING AND PROVIDING ACCESS

The BnF also uses shareable tools (SOLR software for full text indexing, WAT format – developed by the Internet Archive – for content analysis) and is also involved in the Open Wayback project used to provide access to the archives in the reading rooms.

### PRESERVING

The SPAR system (see part 2 section B) aims to receive all data from the web legal deposit. At the same time, preservation strategies are being considered by the IIPC preservation working group, which launched a web documentation operation in order to reconstruct in the future the working environment of an internet user in 2000, 2005 or 2010.



The “BnF Collecte du web” tool, for the selection of websites



Access to the web archives in a reading room

## 05.G Quality management

Quality control of the digital collections is vital to balance between the nature of the resources produced or harvested and the expectations of professionals and the public regarding these resources. Expressed in the form of statistical indicators, quality becomes one of the management tools of the BnF digital strategy, pursued with special attention as it represents one of the singularities of a national library.

## ADAPTED CRITERIA ACCORDING TO CONTEXT

Defining and reaching quality objectives is not simple, as quality is not an absolute value. For example, in terms of preservation, the quality of a digitised document is evaluated for its similarity to the source document. In terms of consultation, it is the opposite: an edited document can appear as better quality than a document strictly adhering to the defects of the original.

Moreover, quality is sometimes governed by technical or contractual constraints. In the case of web archiving, the coverage of a website is not the same for a surface crawling (homepage only) as it is for an in-depth crawling (all pages and media). Assessing the quality of the collection in this case means making sure that it covers the desired forms and areas rather than verifying that the entire website has been covered.

## THE SCALABILITY CHALLENGE

As regards digitisation contracts, the large amount of processed material and the implementation of real-time production within digitisation contracts between the BnF and service providers have made individual testing of the quality of documents impossible. Other methods had to be introduced, such as more sustained test and ramp up stages, auditing of service provider workflows and sample-based controls.

## THE TEST OF TIME

A defect may be detected years after digitisation. Moreover, expectations and methods are still evolving. An old digitisation, judged at the time to be good quality, may appear unsatisfactory later, justifying a new digitisation operation (replacing the black and white with shades of grey or colour, more advanced OCR processing).

## THE CHALLENGE OF CROWDSOURCED PRODUCTION METHODS

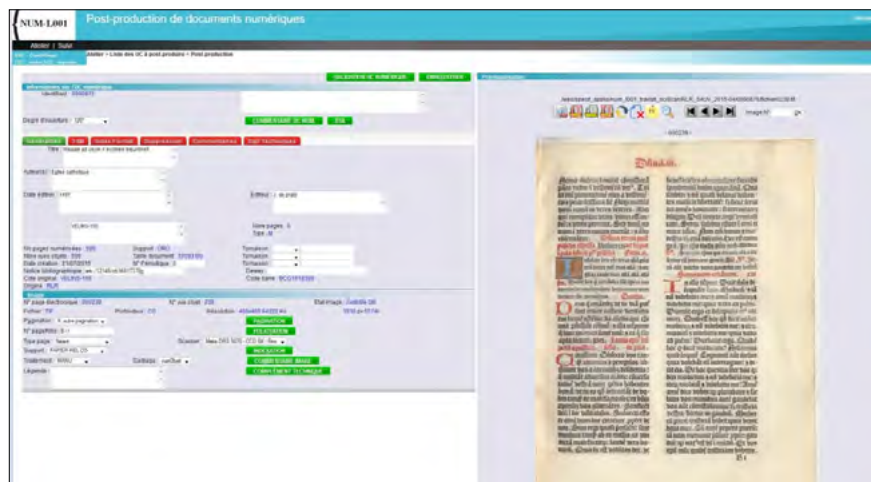
Furthermore, the quality level set today by the BnF for its own digitised collections can create an obstacle to the development of alternative or even crowdsourced production methods. Sharing the production work requires accepting that different quality levels will co-exist in national collections, referenced as such.

What is true for collections is also true for metadata. Planning for the crowdsourced production of records (with publishers, the OCLC network, other stakeholders like the ABES etc.) implies adapting quality expectations or accepting the co-existence of different quality levels between datasets.

Finally, the notion of quality must not be limited simply to managing collections. Measuring the quality of the service offered to end users, in terms of content, response time or even service continuity, proves just as crucial.

**Sharing the production work requires accepting that different quality levels will co-exist in national collections.**

The quality control tool  
of the digitisation system (early printed  
book from the *Réserve des livres rares*  
(Rare Book Reserve), Vélins 155)



# 05.H

## Digital collection management tools

Digital collection management tools in all their diversity cannot be summarised by a single application covering all needs. On the contrary, several tools need to be combined, each dealing with a specific activity: monitoring contracts, conservation, access etc. Some channels like web archiving and the audio-visual system (“SA”, see part 2 section E) have their own range of tools to select, collect, process and circulate the corresponding collections.

### BENEFITING FROM INDICATORS

When it comes to digitisation, production monitoring is assisted by the query tool SIPIL (Steering IT System), designed to collect and query a large amount of data within large sets, as well as document by document. Anomalies and rejections can therefore be identified and then processed. Moreover SIPIL mines the data to produce valuable indicators (weekly and monthly). These indicators serve, firstly, to appraise the performance and stability of digitisation operations. Yet the monitoring needs have gradually evolved, moving the digitisation contracts off centre to enable a better understanding of the digital collections – assessing processing needs, better selecting the collections to digitise or process again, deleting double-ups, identifying those present on Gallica or absent from SPAR.

### ACTING ON THE DATA

Fully describing the digitised collections requires SIPIL to meet ever greater needs. The challenge is to bring together and standardise data from different silos – data from the selection process before digitisation, from the digital repositories and from the catalogues.

However, SIPIL is a statistical tool – it is not able to act on the data or on the digital collections to process them. This kind of operation requires other tools. One good example is the publication tool, which must be able to direct the digital documents towards the different access environments (Gallica, Gallica intra muros, vanilla etc.). Digital collection management tools must also allow for management that is “macro” (applied to large masses) and “micro” (such as the precise monitoring of a document through the digitisation, ingest,

archiving and access workflows). One of the challenges is to connect the physical collections with the digital ones, in order to consider them as a whole, both from the point of view of conservation and access. It is important to ensure that these tools become more specific, intuitive and user-friendly for the staff using them day-to-day. The creation and continuous improvement of its tools must remain a priority for the BnF.

**These tools must allow for operational management that is both “macro” and “micro”.**

### Digital document publication tool

The development of this tool began in 2015, to enable a more precise management of the circulation of content from the digital library.

Each document will be directed towards the appropriate consultation environment: an out of copyright document designed for consultation online will appear on Gallica, while

an e-book entered via legal deposit will have its consultation limited to the research library. In certain cases, access to the collection must also be managed according to the BnF's quality policy e.g. a digital document may be removed from consultation when a better quality version is available – a function that is comparable to weeding.



# 05.1

## Ingest of purchased and donated digital documents (ADDN)

Ingesting documents in a digital form via other means than that of the legal deposit presents the BnF with a new opportunity to reconsider its methods of conservation, cataloguing and communication. It also presents a challenge, as the number of expected documents amounts to several tens of thousands.

### HERITAGE DOCUMENTS

Introduced via donation or acquisition, these digital documents are of a heritage nature (archives, editorial niches) and are to be conserved indefinitely. Coming from the French territories without being covered by legal deposit, they do not, however, exist in a physical medium and thus illustrate the complete switch of certain industries to digital.

In 2004, the donation by writer Pierre Guyotat of his records and electronic correspondence (emails and SMS) to the Manuscripts department was one of the triggers for a debate on the need to adapt to the new practices of depositors. The consideration of digital documents in the ingest channels thus became a vital challenge for the BnF.

### A NEW WORKFLOW

The main difficulties of the project lie in the variety of existing formats and the diversity of document types collected (books, articles, newspapers, photographs, maps, sheet music, private archives etc.). Managing access rights to documents in a wide variety of situations requires a very sophisticated management of each item.

To this end, a new workflow, "Ingesting purchased and donated digital documents" (ADDN), covering all the processing stages, was implemented. The project relies on existing software components and ones currently being developed as part of ongoing initiatives e.g. e-book legal deposit, Gallica intra muros and even SPAR.



Brigitte Pougeoise,  
"Le conte d'hiver",  
Compagnie Arketal,  
Chartreuse du  
Val-de-Bénédiction,  
Villeneuve-lès-Avignon,  
2012

**These documents do not exist in a physical medium and thus illustrate the complete switch of certain industries to digital.**

### An example of a pilot collection: performing arts

ADDN will initially allow for the processing of born-digital photographs, acquired from performing arts photographers. Among these is the Joël Verhoustraeten corpus on street performances (2,750 files in the JPEG format) and the collection of Brigitte Pougeoise,

a puppet show photographer (1,070 files in the JPEG and TIFF formats). These born-digital documents are in line with the photograph acquisition policy which aims to preserve the memory of performing arts.







(BnF





# 06

ENGAGING  
IN SUSTAINABLE  
INNOVATION

---

72

training courses in the digital area in 2013

---

611

people trained in 2013

---

16 million

Euros dedicated to software development  
in the digital sector  
between 2011 and 2014

Until the middle of the last decade, digital activities represented an area of exploration and experimentation for the BnF. This period has seen the birth and growth of major projects like Gallica, SPAR and web archiving.

The Library then went through a digital implementation stage, which gradually affected the organisation of all its working units and job positions. The BnF, like all protagonists in the cultural sphere, has undergone its digital transition.

This profound change, as yet incomplete, would not have been possible if the BnF had not mobilised its resources and skills to create a momentum for change and maintain it over several years.

#### MAIN TOPICS

- 06.A\_ New e-book formats: EPUB, DAISY
- 06.B\_ Research and development in the field of OCR processing

#### RESOURCES, TOOLS AND METHODS

- 06.C\_ Human resources
- 06.D\_ IT investment
- 06.E\_ Internal communications and assistance for change
- 06.F\_ Professional development

#### DEVELOPMENT OPPORTUNITIES

- 06.G\_ Text and data mining
- 06.H\_ 3D digitisation
- 06.I\_ A cooperation project with the French Atomic Energy Commission



# Experimenting and innovating

Digitisation produces primary material for scientific research. The development of innovative and higher-performance tools, in fields like optical character recognition for example, contributes to the uptake of research, notably in the area of human and social sciences. Highlighted during their digitisation, comprehensive corpora become as many areas for study. Furthermore, the mass of digital data available has now reached such a volume that computerised processing becomes the only way to understand it in its entirety. The relationship with documents has been thoroughly transformed: text, data and URL mining leads to the emergence of new corpora and new research methods. By investing in this area of study, the BnF intends to participate in the development of the digital humanities in France.

The BnF is participating in different projects involving innovative digital information production systems. Several strategic directions are favoured, from analysing visual media with the ETIS laboratory (Cergy University) to named-entity recognition (people, places, dates, concepts etc.) in order to create links between entities found in a document and controlled vocabularies maintained by the BnF and others. The cataloguing and research processes are developing, as in the case of CartoMundi, with both a union catalogue and a visualisation tool for 2,000 French and foreign sets of maps. The improvement of OCR includes the crowdsourced production of content (the OZALID research project led by Orange with the Correct experimental platform). The modelling and reuse of data in the semantic web and the analysis of corpora archived in the context of the web legal deposit are also part of this strategy.

**Text, data and URL mining leads to the emergence of new corpora and new research methods. By investing in this area of study, the BnF intends to participate in the development of digital humanities in France.**

Research and development (R&D) projects must be distinguished from innovation projects. Whereas innovation leads to an operational product (data.bnf.fr, for example), R&D aims to experiment with temporary platforms (prototypes, proofs of concept). This is nevertheless true research work, most often guided by academics (laboratories, universities etc.), combining the skills of experts and operational teams. These projects receive funding for this purpose, in the context of open calls for projects (French national research agency, European projects etc.). Although these prototypes are not designed to be reinserted, as they are, in the BnF's information system, they create a real increase in skills and help to confirm the Library's strategic development hypotheses. R&D also allows for platforms to be tried out on a smaller scale, to assess their success and relevance before committing to longer-term investments.

The creation of digital corpora and the digitisation of heritage for research purposes cannot be conducted regardless of usage. When digital practices are thriving, it becomes a real challenge to capture and analyse usage and the appropriation methods of Gallica's digital collections by Gallicanautes. The partnership with Télécom ParisTech regarding the Online Heritage Uses Research Laboratory for Libraries (Bibli-Lab) was created to this end (see part 3 section G). Without over-anticipating the outcome of this work, it may lead in the near future to a better understanding of how digital collections are empowered by data mining and big data analysis tools.

# The means for change

The skills acquired and the experiments undertaken in the context of research translate into a development of the BnF's culture in its day-to-day activities. By adopting digital technologies, the BnF has acted on the need to constantly adapt its services to emerging uses, as well as to the new possibilities offered by technology.

New governance tools are designed to define and maintain change, avoiding the disruptive effect sometimes caused by the constant movement that characterises the digital transition. They include committees (digitisation, conservation, steering European projects etc.), reference documents (collection development policy, conservation policy, technical reference documents for staff, partners and service providers etc.), management and monitoring tools (SIPIL – Library management information system, see part 5 section H), assessment and observation measures of practices and uses (ORH-ION observatory around the development of internal professional practices, Bibli-Lab on internet users, see part 3 section G).

Always being on the move requires considerable investment, notably in IT. The skills of librarians and computer engineers are deeply interlinked, leading to redefined common working methods. Agile methodology for software development was therefore a key factor in digital innovation, resulting in user-oriented design, exchange of expertise between knowledge of collections and technological skills, and increased flexibility in software development planning.

**The BnF  
has joined innovation  
laboratories in order to  
monitor digital  
transition examples  
outside  
of the library sphere  
and explore internal  
pools of expertise  
and imagination.**

The human factor also plays a decisive role in the digital transition. The predictive management of workforce, jobs and skills (GPEEC) is used to analyse the development, over time, of the BnF's human resources, taking into account the evolution of skills and job positions in a digital environment. Many initiatives support the transition – the attribution of dedicated resources to new projects, awareness raising operations for a wider audience beyond staff directly involved in digital operations etc. Innovative methods are also chosen to spearhead this support – training has been reinvented, in workshops where experts share their knowledge with their colleagues, while internal communications are transitioning towards corporate social networking.

In order to further sustain this momentum, in 2015 the BnF joined the Ideas Laboratory, an open innovation body of the French Alternative Energies and Atomic Energy Commission. This is a place to share thoughts and practices according to a rationale that is both pragmatic and exploratory. The installation at the BnF of a Parisian branch of this laboratory is an opportunity to monitor digital transition examples outside of the library sphere and to explore the pools of expertise and imagination from all the strata of the BnF that are still overlooked. What is happening here is a real change of identity for the Library. At the approach of the quarter-century, the BnF needs to reinvent how it implements its missions in a world where access to culture and knowledge is undergoing a total revolution.

# 06.A

## New e-book formats: EPUB, DAISY

EPUB has set itself up as the reference standard for e-book production and distribution. This universal format responds to the demand for mobility, portability and accessibility in accordance with the development of contemporary cultural practices. In its last version (EPUB 3), this format satisfies all reading requirements (regular, for those who are visually impaired or blind, for those with dyslexia etc.) and for all types of content (non-fiction, literature, youth, school, university).

### ADHERING TO THE E-BOOK ECOSYSTEM

Since 2013, the production of e-books in EPUB has been an integral part of Gallica's circulation strategy. However, at a time when a vast digital ecosystem has been put in place, the BnF is ensuring that its free EPUB production for Gallica is not perceived as competition to the commercial offering. This is why the development of this service is done at a constant but measured pace: 3,200 EPUBs will be produced over the 2014–2017 period. Rare authors, critiques and commentaries on works from the public domain, scientific studies and even thematic corpora are favoured.

### AN ACCESSIBLE FORMAT

The EPUB format is not the only technology adopted by the BnF to circulate its digitised collections more widely. By also producing files in the DAISY format, designed for access by people with a disability, the BnF encourages access to all written heritage works, while at the same time further developing its technical mastery of new formats.

**The production of electronic publications in EPUB has been an integral part of Gallica's circulation strategy.**

An e-book in EPUB  
format on Gallica



## EPUBs from the BnF

These EPUBs are a literal transposition of the original item in a digital form, without transformation or adaptation. However, other projects undertaken by the BnF's partners are designed to produce EPUB e-books in an editorial way (correction of original text, new cover, addition of critical apparatus etc.). E-books sold in the context of the ReLIRE project are produced by FeniXX, while those

which will join the [bnfcollection.com](http://bnfcollection.com) service of the BnF-Partenariats subsidiary are produced by Ligarán, a specialised partner in the area of publishing. Another partner, the OBVIL Labex produces, as part of its research activities, EPUBs of classic literary texts, including critical scientific apparatus, designed to be circulated under a free licence and integrated into Gallica.



# 06.B

## Research and development in the field of OCR processing

Over the last few years, the BnF has invested heavily in the field of applied research in optical character recognition (OCR). The Library intends to provide users with digitised documents in text mode and of the best possible quality – something that can only be achieved by acquiring the appropriate expertise and tools, designed in the context of R&D work.

### NEW QUALITY CONTROL PROCEDURES

The BnF participated, during the 2011-2014 period, in numerous research projects and launched a project as part of its triennial plan in association with the University of Rouen (*Contrôle qualité des données OCR* – Quality control of OCR data). The procedures developed were adapted to the processing of considerable volumes of digital files via OCR processing software, in order to be applied to data delivered by the digitisation service providers.

### STANDARDS CONTROL

As the BnF is dependent on technical formats that are constantly evolving (METS, ALTO), the active presence of its representatives on the editorial boards of the formats in question is necessary to ensure the richness of the information carried.

### NEW USES

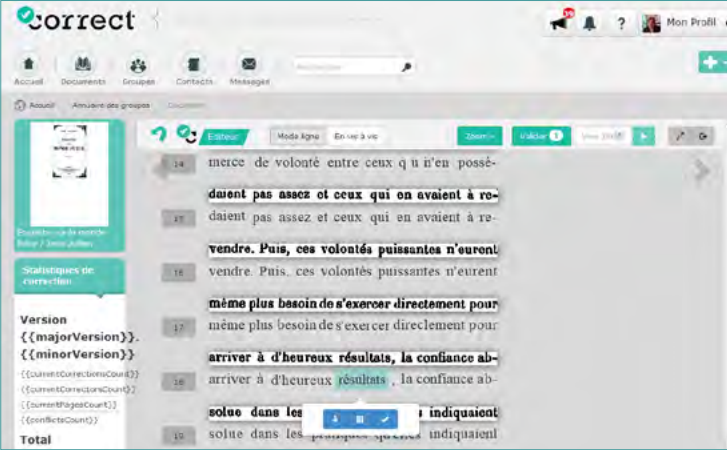
OCR processing must also be considered with regard to user needs, which develop with cultural practices and which OCR processing contracts must anticipate. The [Europeana Newspapers](#) (2012-2014) European project (see part 3 section N), designed to digitise and promote the newspaper collections, allows for experimentation of the quality assessment of OCR processing, recognition of the logical structure of part works (headings, titration, layout of the article) and search options in full-text.

### A EUROPEAN CENTRE OF COMPETENCE

As a member of the [IMPACT](#) Centre of Competence (an NGO promoting the digitisation of heritage texts), the BnF will be fully involved in the creation of new European projects (notably in the context of the Horizon 2020 call for projects), focused on OCR technologies considered in their wider scope: text recognition, but also automatic correction and enrichment of information carried by the text (recognition of named entities, extraction of concepts etc.) Over the long term, this work will result in important developments, such as the integration of an automatic correction tool

in the digitised newspapers processing workflow and the development of a tool for the retrospective processing of newspaper collections.

**The Library intends to provide users with digitised documents in text mode and of the best possible quality.**



The screenshot shows the CORRECT web interface. At the top, there's a navigation bar with icons for Accueil, Documents, Groupes, Contacts, and Messagerie. Below this, a sidebar on the left displays 'Statistiques de correction' and 'Version' information. The main area shows a document with several lines of text, some of which are highlighted in red to indicate errors or areas needing correction. A table at the bottom summarizes the correction statistics.

Version	Statistiques de correction
{{majorVersion}}	14
{{minorVersion}}	15
{{commentCorrectionsCount}}	16
{{commentCorrectionsCount2}}	17
{{commentThesesCount}}	18
{{commentThesesCount2}}	19
Total	20

**The CORRECT platform**

As part of the FUI OZALID research project for the design of the CORRECT platform (Correction and crowdsourced enrichment of text), the BnF experimented with crowdsourcing by suggesting to its users that they correct the quality of the OCR of documents digitised on Gallica. This R&D project has provided the opportunity to explore the possibilities for enriching documents in order to produce new versions (electronic publications, adapted editions for blind people etc.).

The OCR correction interface of CORRECT

# 06.C

## Human resources

While committed to maintaining the expertise of its staff in other areas (collections, art professions etc.), the BnF wishes to also support them sustainably in the digital transition. Regarding the predictive management of workforce, jobs and skills (GPEEC), digital technologies are considered as a factor for development with a repercussion on jobs (emergence of new profiles, development of activities and processes etc.).

### **JOBS AND SKILLS REFERENCE DOCUMENT**

At the BnF, GPEEC relies on a certain number of tools, at the forefront of which is the jobs and skills reference document, updated between 2010 and 2012, to take into account the impact of digital technologies. Various working groups have analysed, by job category, the changes (observed or anticipated) caused by mass digitisation and by new activities (web archiving, Gallica content curation etc.).

### **MEASURING THE IMPACT OF DIGITAL TECHNOLOGIES**

However, the main challenge in terms of managing human resources is to implement indicators able to measure the qualitative and quantitative effect of digital technologies on jobs. Although activities linked to digitisation can easily be identified, it is more difficult to identify activities completely or partially directed towards digital projects, as their diversity is so great.

The only indicator that objectively assesses the growth in the number of staff involved in digital activities and related tasks is the number of training courses given and taken by staff in this area. A 2013 survey showed that between 2008 and 2013, the number of digital training courses increased from 21 to 72. Up to 2007, less than a hundred staff members were trained each year in this area in the context of core professional activities (services to the public, processing of collections). Excluding the two peaks of over 600 in 2008 and 2013, this number then stabilised to around 400 per year, or approximately 15% of Library staff.

### **ORH-ION: an observatory for sharing experiences and ideas**

Created within the BnF in 2008, the Observatory for work organisation and human resources under the operational effect of digital technologies (ORH-ION) regularly brings together, around thematic seminars, stakeholders of the BnF's digital programmes and activities. Free of any hierarchical or operational framework, ORH-ION plays a catalytic role in the emergence of a shared digital culture.

Feedback on digital content curation and the management of born-digital collections are among the fundamental questions raised regarding the development of professions and tasks, which have been debated in this context since 2010. The group has also conducted several sector enquiries – on the role of cataloguing, storage in the digital age and even the development of the photographic profession.

### DEDICATED RESOURCES

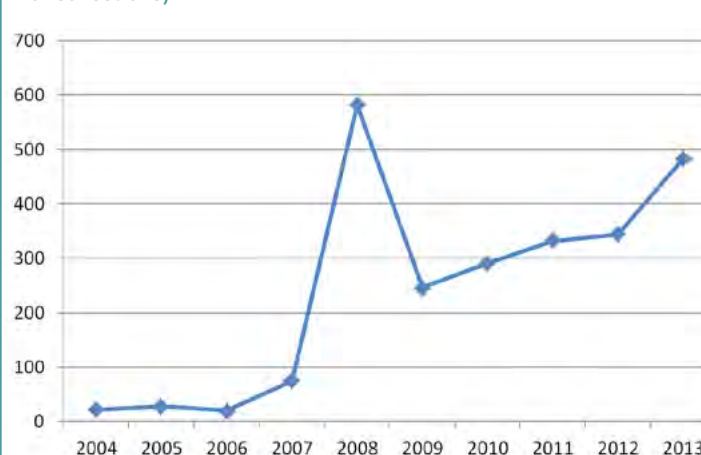
Certain projects were, however, subject to a deployment of dedicated human resources, in particular during the implementation of new processes. The ReLIRE project (see part 1 section G) therefore benefits from dedicated human resource funding from the CNL, to the tune of approximately 400,000 Euros per year, or six FTEs. Meanwhile, implementing the legal deposit of e-books requires the redeployment of seven FTEs previously dedicated to redistribution of the second copy from the legal deposit for printed material, deposits having been reduced to a single copy by a 2015 reform.

**The main challenge in terms of managing human resources is to implement indicators able to measure the qualitative and quantitative effect of digital technologies on jobs.**

2008-2013 development in the number of training courses in the digital area offered to the BnF staff



2004-2013 development in the number of staff members trained in the digital area: core profession (public services, processing of collections)





# 06.D

## IT investment

The BnF's capacity for innovation depends, on the whole, on its aptitude to invent and develop applications and software that can take on new functions and volumes greatly exceeding the performance of software packages on the market.

### OPEN SOURCE COMPONENTS

The investment allocated for the creation of these tools, often innovative from a functional and technical point of view, is sometimes shared with communities through open source components, not just for financial reasons, but also out of a concern for sustainability, respect for standards, and the drive for change and openness (this is the case in particular for software supporting the web legal deposit: see part 5 section F). Part of this software, developed using budgets allocated to the BnF by the CNL, contributes to a national investment also benefitting to the BnF's partners (French, French-speaking and European institutions; publishers; publication stakeholders etc.).

### A SOFTWARE DEVELOPMENT PLAN

The BnF has set up an annual "software development plan", listing all software needs from all areas combined, with details of the financial and human resources necessary over the next year plus financial projections for the following two years. Once these investments are approved, teams from the IT department take responsibility for each project and all of its components (quality, cost, time frame, risks).

### PROJECT PORTFOLIOS

Portfolio managers are in charge of a set of related projects. They support teams, play an advisory and expertise role, ensure technological and prospective

monitoring and participate in the life of projects in their portfolio (analysis of events, decision making, and hierarchy information). This structure, combined with the use of agile methodology, favours innovation and user-oriented results. It guarantees constant visibility throughout the process through continuous discussions between the stakeholders (IT engineers, service providers, users, partners and decision makers). It allows for a review of the priorities during the year and for the integration of new requirements. The investment in software development for digital projects between 2011 and 2014 represented over 16 million Euros, over 6 million of which were provided by the CNL. Over this period, it mobilised 33 to 36 FTEs within the IT department.

**The investment in software development for digital projects between 2011 and 2014 represented over 16 million Euros.**

### The "Scrum" agile methodology

Used at the BnF since 2010, this method has played a key role in the dynamic of software development. Experimental to begin with and then spread in 2013 to the different IT development projects from the digital and library community areas, it fosters communication between teams and mutual

understanding of needs and constraints. Allowing for a more flexible development planning, a better suitability of the product to users' needs and ability to anticipate development in usage, this methodology has also created a new role – the product owner – in charge of combining and prioritising needs.

	2011	2012	2013	2014	Total
Bnf	€2,808,000	€2,483,000	€2,517,000	€2,144,000	€9,952,000
CNL	€1,077,000	€865,000	€1,940,000	€2,167,000	€6,049,000
Other	-€	-€	€17,000	€30,000	€47,000
<b>Total</b>	<b>€3,887,011</b>	<b>€3,350,012</b>	<b>€4,476,013</b>	<b>€4,343,014</b>	<b>€16,048,000</b>
<b>FTE IT</b>	<b>33.4</b>	<b>33.1</b>	<b>34.3</b>	<b>35.6</b>	<b>-</b>

Summary table of digital-related investments in the software development plan (2011-2014)

# 06.E

## Internal communications and assistance for change

---

Digital technologies feed all professions at the BnF. To face this structural development, support for change is required. Internal communications are an important part of this support.

### THE “DIGITAL MEETINGS”

In 2013, awareness raising sessions on the challenges of this digital era were introduced at the BnF at the rate of two per year. Entitled the “*Rendez-vous du numérique*” (Digital meetings) these sessions are designed for all staff. They aim to create a shared digital culture and to promote the training courses offered by the Library. The *Trajectoire* internal journal also encourages digital cultural integration by focusing on a project, team, department, method or innovation. The “*Midi de l’info*” information meetings (uploaded to the intranet in an enriched version, like all the information sessions) present projects and achievements plus major areas of the digital strategy. This drive for information and learning is nourished by experience sharing with other institutions, like the crowdsourcing method of the Royal Library

of Denmark and even the digital transition of the Agence France-Presse.

### AN ORGANISATIONAL SOCIAL NETWORK

Beyond these educational and informational initiatives, the BnF keeps evolving along with new staff practices and the tools they use to share their expertise and experience. The implementation of a corporate social networking tool would contribute to this, by establishing a community of interest, allowing for the management and monitoring of projects, the co-creation of documents, the creation of a knowledge corpus, the listing and seeking of skills, the emergence of ideas etc. Such a collaborative instrument breaks away from traditional organisational schemes, so its implementation requires the development of new management practices. Its success implies a consensus around shared “rules of the game” and management involvement in networking. Information channels will be made more fluid and dynamic, interdisciplinarity and efficiency will be favoured, crowdsourced innovation developed and, finally, the motivation of “stakeholder-staff” bolstered.

**A collaborative instrument breaks away from traditional organisational schemes, so its implementation requires the development of new management practices.**

# 06.F Professional development

Once the digital revolution had been concretely integrated into its missions, the BnF saw new skills emerge, often developed informally and experimentally by experts. The Library gradually became aware of this collective increase in knowledge and expertise and decided to support it through training.

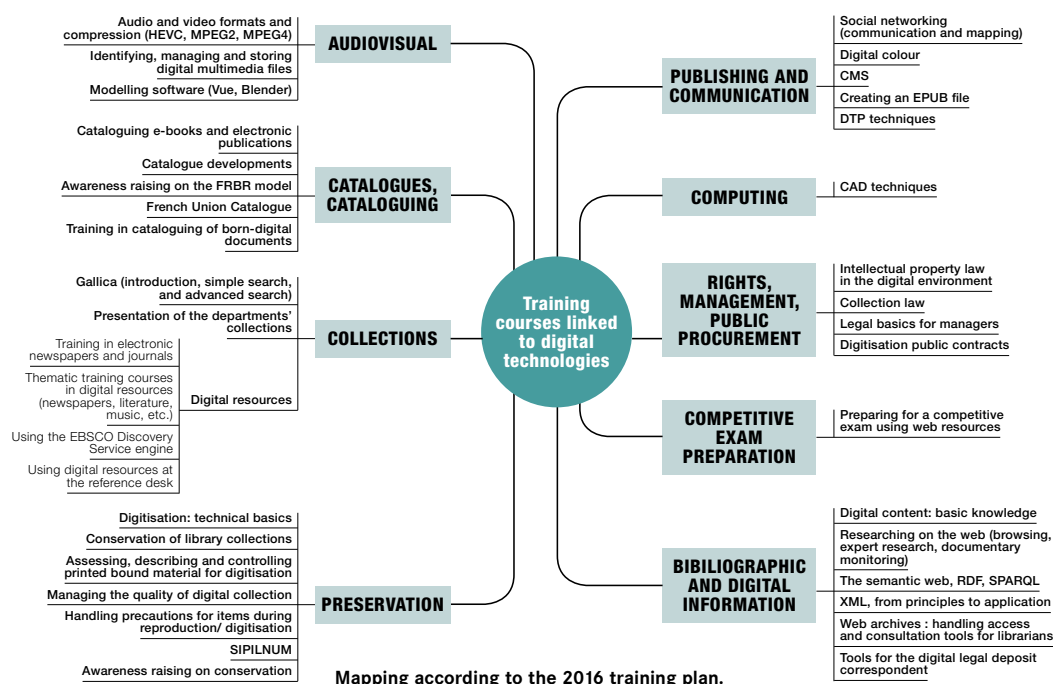
## THE TRAINING PLAN

The Library's training plan gradually integrated digital technologies under different methods: instructorship, mentoring, information, and training. These activities cover several areas of the BnF's training plan: digital collections and resources, preservation, publishing, bibliographic and digital information etc. Expectations in terms of management, law and even public contracts were expressed, since librarians are not the only ones affected by the development of digital technologies.

## INNOVATIVE PRACTICES

The acquisition of new skills cannot occur just within training courses, as the most innovative topics are not always covered by established training bodies. Other forms of support, such as instructorship, participating in satellite workshops at international specialised conferences and even sharing technology monitoring must now be used to develop knowledge and expertise in the most emergent areas.

**The BnF encourages diverse forms of support, such as instructorship, participating in satellite workshops at international specialised conferences and even sharing monitoring.**



Mapping according to the 2016 training plan.

## An example of internal training

BnF departments have proven highly creative in guaranteeing the transmission of expertise between staff members involved in digital projects and those who only perceive them remotely or partially. In order to support the digital innovation projects that it manages, the department of the legal deposit launched, at the end of 2013, an internal digital training curriculum. This initiative aimed to foster the acquisition of a shared digital culture

by all department staff. In the simple and interactive form of a workshop, the training course promoted discussions among attendees, especially when the issues dealt with concerned situations they experienced day-to-day. The course opened with two plenary sessions and continued with optional workshops followed by 48% of department staff. This initiative also reinforced interdisciplinarity within the department itself.



# 06.G

## Text and data mining

To ensure access to as many collections as possible, the BnF must offer more than simple access to digitised content: it must allow researchers to use analysis tools which are able to go beyond what the human eye can read, to establish new knowledge. Text and data mining (or TDM) offers the opportunity to see new avenues emerge: a document will no longer be consulted just for its content, but also for its position in a corpus. Its metadata will also be questioned, in addition to its internal and external occurrences.

### PROJECTS IN PARTNERSHIP

The data produced and conserved by the BnF (whether bibliographic and authority records placed under the open licence of the French State (*Licence publique de l'État*), digitised and OCR processed corpora from Gallica, or web archives (see part 2 section A)) open up immense possibilities in this regard. Two projects were launched at the same time: the first, conducted with the Grenoble Political Science PACTE Laboratory, consisted of analysing web archives corpora linked to the elections of 2012 and the Olympic Games of 2012; the second was carried out with Télécom ParisTech and the French

Library of international contemporary documentation (BDIC), in the context of the “Les Passés dans le Présent” Labex (see boxed text). The purpose of these two initiatives is to extract metadata from web archives allowing for corpora to be described, analysed, characterised and mapped.

Data that the BnF creates and collects during its activities is added to that of a heritage nature – consultation logs of Gallica resources, tracks of reader’s on-site activities, administrative data used in the context of the Legal Deposit Observatory etc. There are as many sources of information for which big data technology could, after anonymisation, greatly increase the impact.

### A NEW SERVICE FOR RESEARCHERS

The BnF aims to develop this activity in future, by offering its researchers a secure infrastructure that allows them, according to the data they are interested in and in particular the legal status of the collections they are studying, to build their research corpora and explore them, either with their own tools or with those of the BnF.

**The big data technologies could, after anonymisation, increase the impact of the data created and collected by the BnF.**

### “The online future of digitised heritage from the Great War”

Done in collaboration with the Library of international contemporary documentation (BDIC) and the Institute for Political Social Sciences (ISP), this programme was established for the 2013-2016 period in the context of the “Les passés dans le présent” Labex (“Pasts in the present” top laboratory). Its purpose is to better understand the way in which a digitised corpus – that from the First World War – circulates in the

web, encouraging discussions and multiple appropriations. One of the research steps relied on a corpora coming from the web archives, harvested for this purpose. TDM tools were used, on the one hand to map the links between sites and on the other to analyse the content of the 14-18 forum – one of the key amateur sites dedicated to The Great War.

# 06.H

## 3D Digitisation

Developed in 2015, the new version of Gallica integrates the [3D visualisation](#) of globes kept in the maps and plans department. This new kind of digital object is the result of a skill-based patronage agreement for the digitisation and 3D reproduction of a set of 55 terrestrial and celestial globes, concluded in July 2015 by the BnF and Dai Nippon Printing (DNP). DNP provided the BnF with its 3D digitisation technology, perfected for the occasion. With this achievement, the digitisation processes implemented by the BnF literally took on a new dimension.

### 3D DIGITAL REPRODUCTIONS

The maps and plans department keeps approximately 120 ancient terrestrial and celestial globes, including several major world heritage

artefacts. This exceptional collection is a priceless testimony to the development of geographical knowledge and the understanding of the known universe in the Christian West and Arab world between the 11<sup>th</sup> and late 19<sup>th</sup> century.

This partnership between the BnF and DNP aimed to carry out 3D digital reproductions of the spheres. Like the original items, these copies can be rotated in all directions. Their high definition textures allow a clear and seamless zoom over any surface.

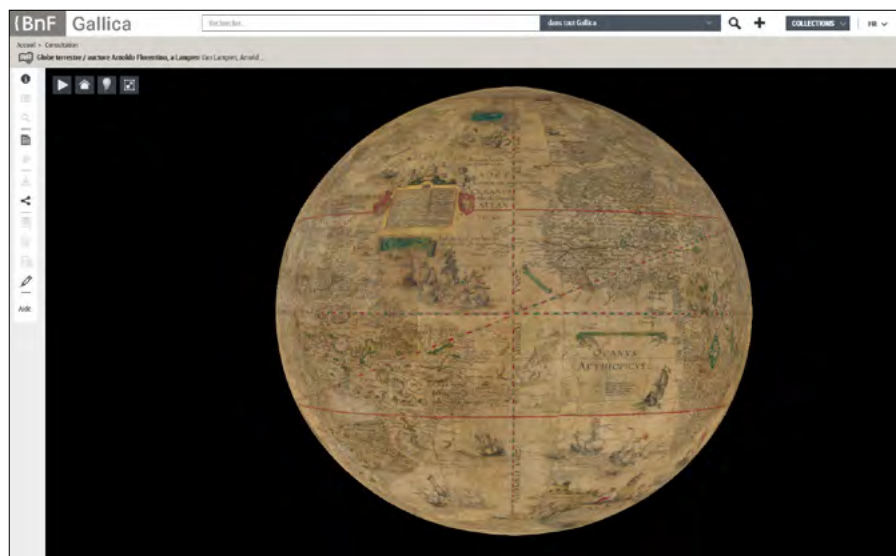
### A REMARKABLE COLLABORATION

After the 3D modelling of digital data, carried out in Tokyo by the Japanese company, the BnF archived the high definition masters in SPAR and circulated a version with a smaller file size on Gallica. Thanks to this remarkable Franco-Japanese collaboration,

these groundbreaking objects can now be accessed in the new version of the digital library's viewer. After 2016, an online virtual exhibition, supported by the Total Foundation, will provide a commented discovery of these globes to a wider audience. Innovative displays have also been developed for this project, to be used in DNP (Tokyo) and BnF (Paris) exhibition spaces.

**The 3D digitised globes can be accessed in the new version of the digital library's viewer.**

Photogrammetric shot of a terrestrial globe in the studios of the maps and plans department.



# 06.

## A cooperation project with the French Atomic Energy Commission

At the end of 2014, the BnF initiated a prospective approach in order to sketch out the outlines of a national library for the France of 2025. Started during a management seminar led by the [CEA TECH](#) team, this reflection intends to consider the scientific, cultural, educational and societal role of the BnF in ten years' time. Which users will it reach? Which services must be offered and for which research, culture and leisure requirements and practices? How will the BnF's missions be carried out in a digitally transformed environment and who will its partners be?

**The partnership ecosystem must be redesigned between operators and partners with which the BnF cooperates or could cooperate.**

Specifically addressing the scope and developments of digital assets, the BnF's innovative approach will be guided by three major issues:

### **A NATIONAL LIBRARY... FOR WHAT PURPOSE AND FOR WHOM?**

Physical and virtual library spaces must adapt to new uses and a new knowledge context.

Within a growing context of digital practices and mobile uses, new tools and digital services must be invented.

User experience must be set at the core of the professional culture and reconsidered in all its dimensions.

### **A NATIONAL LIBRARY... WITH WHOM?**

The partnership ecosystem must be redesigned between operators and partners with which the BnF cooperates and could cooperate, for the production of bibliographic data, the creation and provision of resources, cultural production, heritage

conservation etc. In a context where the production and circulation of knowledge (research, teaching etc.) is profoundly changing in terms of methods, what place and which services should be considered for preservation institutions like the BnF with significant volumes of data and "certified" metadata? New digital working environments, spaces and networks, enabling data mining or the co-creation of proofs of concept may appear, thanks to opportunities like the installation of 1,000 innovative start-ups, from 2016, in the Halle Freyssinet (near the François-Mitterrand site).

### **A NATIONAL LIBRARY... HOW?**

The approaches implemented for conducting and assisting change (cultural integration in digital developments, renewed relationships with users etc.) must themselves take on new forms, in order to set a sustainable innovation culture within the Library. A strong collective involvement of third parties (staff, users, stakeholders, partners etc.) may appear via crowdsourced innovation and the development of specific expertise in creative methods.

The BnF's collaboration with the CEA's [Ideas Laboratory](#) will lead to its hosting of a branch of the laboratory. This creative space will be open to members of the Ideas Laboratory, partners of the BnF, plus staff and users wishing to participate in the design of the BnF of the future and to experiment with the most innovative projects.

## The challenge of organised technology monitoring

The technology monitoring practised at the BnF consists of collecting and possibly recirculating information judged useful or strategic. This new skill was developed in less than a decade using web 2.0 tools and social networking. It has been included since 2012 in the jobs and skills reference document. To optimise circulation and management of the technology monitoring results, a mapping of this activity and its circulation methods should be undertaken, to survey needs and satisfaction levels; to identify causes of an absence of technology

monitoring and possible prioritisation; to list internal skills in terms of technology monitoring training. On the basis of such a synthesis, the complementary nature of the varieties of professional monitoring taking place in France could be considered, given the 2010 *Digital roadmap for libraries*. Organisational, training, circulation, evaluation and archiving scenarios for technology monitoring could be developed within the BnF, hand-in-hand with the innovation activities conducted by the Ideas Laboratory.



**Photo credits:** Eric Sempé / BnF (Cover, p. 9, 21, 42-43, 82), Paolo Verzzone-VU' / BnF (p. 14-15, 84-85), David Paul Carr / BnF (p. 22, 26, 49, 62, 65, 116), Luc Verrier / BnF (p. 23), Alain Goustard / BnF (p. 30-31, 73), Patrick Bramoullé / BnF (p. 37), Thierry Ardouin / Tendance Floue / BnF (p. 39), Pascal Lafay / BnF (p. 46), Jennifer Boche (p. 59), David Goehring / Flickr (p. 63), Catherine Floc'hlay (p. 82), Béatrice Lucchese / BnF (p. 92), Brigitte Pougeoise / Fonds Brigitte Pougeoise / BnF, (p. 100), DR, Michèle Laurent / BnF (p. 101), BnF (p. 102-103), DR.

**Consulting and creation:** Kstories / [www.kstories.paris](http://www.kstories.paris)





**François-Mitterrand**

Quai François Mauriac, Paris 13<sup>e</sup> \_ 33 |0|1 53 79 53 79

**Richelieu**

5, rue Vivienne et 2, rue Louvois, Paris 2<sup>e</sup> \_ 33 |0|1 53 79 53 79

**Bibliothèque de l'Arsenal**

1, rue de Sully, Paris 4<sup>e</sup> \_ 33 |0|1 53 79 39 39

**Bibliothèque-musée de l'Opéra**

Place de l'Opéra, Paris 9<sup>e</sup> \_ 33 |0|1 53 79 37 40

**Maison Jean Vilar**

8, rue de Mons, 84000 Avignon \_ 33 |0|4 90 86 59 64

**Centre technique (CTBnF)**

14, avenue Gutenberg, 77607 Bussy-Saint-Georges \_ 33 |0|1 53 79 38 44

**Centre de conservation Joël Le Theule**

Le Château, 72300 Sablé-sur-Sarthe \_ 33 |0|2 43 95 19 92

[www.bnf.fr](http://www.bnf.fr)