

CHARTRE DOCUMENTAIRE DE LA NUMÉRISATION DES COLLECTIONS DE LA BIBLIOTHÈQUE NATIONALE DE FRANCE



{ BnF | Bibliothèque
nationale de France

Sommaire

Introduction 4-5

Partie 1 6-8

Les grands principes de la numérisation

Partie 2 10-11

Le champ d'application de la numérisation

Partie 3 12-13

Les méthodes de numérisation

Partie 4 14-17

Les critères de sélection

Partie 5 18-19

Les processus de décision et de validation

Partie 6 20-21

Être acteur de la politique documentaire de numérisation

Annexes 22-29

- Bref historique de la politique de numérisation

- État actuel de Gallica par type de document

- Répartition actuelle de Gallica par disciplines

- Répartition actuelle de Gallica par siècle

- Quelques informations sur les publics et la consultation de Gallica

- Brève synthèse sur la coopération

- Principaux éléments techniques

- Quelques éléments financiers

- Bibliographie

Bref historique

La numérisation des collections de la BnF a commencé dès le début des années 90, sous l'impulsion de l'EPBF (établissement constructeur de la Bibliothèque de France). Gallica, la bibliothèque numérique de la BnF, a été lancée en 1997. Depuis cette période, la numérisation a connu des évolutions majeures.

- ▶ Lancée dans la perspective de créer une bibliothèque numérique de « l'honnête homme » (à l'époque l'objectif était de 100 000 ouvrages), la bibliothèque numérique est passée à la numérisation de masse, appliquée d'abord aux imprimés (en 2007 sur la base de 100 000 documents par an sur crédits extérieurs du CNL) puis aux collections spécialisées (en 2009 sur les fonds propres de la Bibliothèque).
- ▶ La bibliothèque numérique a d'abord été conçue pour conserver et diffuser les collections de la BnF. Dans les mêmes années (2009-2010), la BnF a fait de la numérisation l'axe central de sa coopération en France comme avec l'étranger. Gallica est ainsi devenue une sorte de « bibliothèque numérique de France » même si elle n'en porte pas officiellement le nom, dont une grande partie est ensuite mise à disposition dans Europeana¹.
- ▶ Au fil des années, les techniques de numérisation se sont considérablement améliorées (niveau de résolution notamment, reconnaissance optique de caractères, normalisation...) et la BnF a intégré ces progrès, d'où, aujourd'hui, une hétérogénéité des qualités de numérisation qui peut être sensible pour les utilisateurs².

L'offre aujourd'hui

Aujourd'hui, Gallica est une double bibliothèque numérique :

- ▶ Gallica, version internet diffuse les documents du domaine public, et ceux qui ont fait l'objet de négociations avec les ayants-droit.
- ▶ Gallica intramuros, consultable uniquement dans les sites de la BnF, propose la bibliothèque Gallica accessible sur internet à laquelle s'ajoute l'ensemble des documents sous droits, non diffusables sur internet pour des raisons relatives à la propriété intellectuelle, les documents réalisés dans le cadre des partenariats public-privé, et dont la diffusion est restreinte pour quelques années, des acquisitions numériques, ...

À ce jour Gallica propose (chiffres arrondis) 346 000 livres (dont certains en plusieurs volumes), 1 700 000 fascicules de périodiques (pour 12 000 titres dont 4 600 titres de presse), 1 100 000 images, 100 000 cartes, 40 000 partitions, 60 000 manuscrits, 40 000 enregistrements sonores, 178 000 objets. Pour les imprimés, plus de 410 000 documents proviennent de collections de plus de 300 partenaires dont plus de cinquante relèvent du Ministère de l'Enseignement supérieur³. Gallica intramuros propose plus de 4 millions de documents, dont 452 000 livres (dont certains en plusieurs volumes), 120 000 cartes, 81 000 manuscrits, 1 300 000 images, 2 000 000 fascicules de périodiques, 54 000 partitions et 35 000 enregistrements sonores. En outre, 585 000 documents proviennent de bibliothèques partenaires⁴.

Pour les imprimés uniquement et en ce qui concerne les collections BnF seulement, l'on considère que 34 % de l'offre relèvent de la littérature et de l'art, 30 % des sciences humaines (histoire, philosophie, religion, ...), 19 % du droit, de l'économie et des sciences politiques, et 17 % des sciences et des techniques, correspondant grosso modo au volume des collections des différents départements concernés.

Pour les collections des partenaires, les disciplines les plus fortes sont l'histoire (31 %), les sciences et techniques (30 %), les sciences juridiques (19 %), l'art (10 %)...

L'ensemble des documents présents sur Gallica se ventile par siècle comme suit : 34 % pour le xx^e siècle, 37 % pour le xix^e siècle, 14 % pour le xviii^e siècle, 5 % pour le xvii^e siècle, 3 % pour le xvi^e siècle et 7 % pour les documents des autres époques.

Par ailleurs, la BnF a entrepris à partir de l'an 2000 un plan de numérisation dit « plan de sauvegarde » de ses collections patrimoniales sonores, audiovisuelles et multimédias particulièrement fragiles. Ces collections étant majoritairement sous droits, sont peu consultables dans Gallica. Elles sont principalement accessibles dans les salles de lecture de la BnF. L'offre actuelle est de 920 000 supports (dont 566 000 supports sonores, 196 supports vidéo et 36 000 supports électroniques (CD-Rom, multimédias).

¹Lancé fin 2008, Europeana est le portail numérique du patrimoine européen. Il donne accès à plus de 48 millions d'objets numériques issus de plus de 3300 institutions (bibliothèques, archives, musées et fonds audiovisuels, ...). Les collections de la BnF y sont intégralement versées. Ce sera le cas progressivement pour les collections des bibliothèques partenaires à partir de 2017.

²Voir annexe A pour un historique plus développé.

³L'ensemble de la numérisation coopérative depuis 2009 représente 17 millions de pages avec un soutien financier de la BnF à hauteur de 6 millions d'euros.

⁴Sauf mention particulière, tous les chiffres donnés dans ce document ont comme date de référence décembre 2016.

Les usages

Gallica connaît une consultation annuelle d'environ 16 millions de visites, soit environ 44 000 par jour, avec des pics qui peuvent atteindre 50 000 visites par jour.

Quelques informations issues de la dernière enquête (2016) auprès des « gallicanautes » (7 600 répondants) :

L'âge moyen des répondants est de 54 ans. S'ils résident majoritairement en France, 28 % se connectent de l'étranger. Usagers assidus (un répondant sur deux se connecte au moins une fois par semaine), ils sont plus diplômés que la moyenne des Français, et ont déjà fréquenté les espaces physiques de la BnF pour 38 % d'entre eux. Évolution notable des usages depuis la précédente enquête (2011), ils ont tendance à privilégier de plus en plus fréquemment la consultation attentive en ligne sur tout autre usage des documents. Les domaines les plus consultés restent l'histoire, suivie par la littérature, puis l'art, avec un bon niveau de notoriété et d'usage de la diversité des différents types de documents accessibles (à l'exception des objets). Les consultations de Gallica sont majoritairement le fait de projets de recherche précis, le plus fréquemment menés à titre personnel. Pour l'ensemble des gallicanautes, les motivations de consultation (études, recherches professionnelles, personnelles, ...) sont cependant de plus en plus mêlées : seulement 28 % ne déclarent qu'un unique motif de visite.

Ce document présente la politique de numérisation de la BnF⁵. Il s'accompagne d'un document quinquennal de perspectives (2017-2021). En complément, la BnF publie annuellement le bilan et le programme de sa numérisation.

⁵Pour la période précédente la BnF avait élaboré trois documents, diffusés en leur temps sur le site web de la Bibliothèque : « La charte de numérisation de Gallica » (2003-04, document interne), « La politique documentaire de la numérisation de masse à la BnF, années 2007-09 » (document interne) et « La charte documentaire de Gallica 1997-2007 ».

LES GRANDS PRINCIPES DE LA NUMÉRISATION

- Un rôle central dans l'offre numérique
- Une visée culturelle et de recherche au service de tous les publics
- Un cadre juridique rigoureux
- Une politique résolument coopérative
- Des financements diversifiés
- Des technologies éprouvées et évolutives

Un rôle central dans l'offre numérique

La numérisation des collections de la BnF s'inscrit dans le contexte de la mutation des savoirs et de la création vers le numérique.

La numérisation rétrospective assure la conversion de tout ou partie du patrimoine du passé dans le monde numérique. De leur côté, les productions éditoriales contemporaines (dont la collecte est la mission de la BnF) sont massivement en transition vers la production numérique

- ▶ livres, revues, audiovisuel, manuscrits, documents graphiques, partitions, ... L'enjeu global est la conservation et la diffusion du patrimoine du passé et de l'avenir sous une forme numérique unifiée, pour en permettre une consultation et un partage simplifié.

Plus globalement encore, cette offre documentaire prend place dans une offre numérique globale de la Bibliothèque qui concerne aussi son offre culturelle (expositions virtuelles, conférences en ligne, applications multimédia...), et l'ensemble de ses services (accueil, inscription, reproduction, communication, médiation, ...) ⁶.

Une visée culturelle et de recherche au service de tous les publics

La numérisation des collections de la BnF a une visée à la fois de conservation et de diffusion du patrimoine à l'échelle mondiale. La mission première de la Bibliothèque est de conserver et de communiquer ses collections à des fins de recherche, académique, professionnelle, ou personnelle. La numérisation s'inscrit cependant au service de la politique des publics de la BnF qui n'exclut aucun public de son champ d'action ⁷. Un des critères de sélection des collections à numériser est la prise en compte de l'intérêt que les collections peuvent susciter auprès du plus grand nombre possible de segments de publics : actifs, enseignants et médiateurs, scolaires et étudiants, jeunes et enfants...

La conservation de la numérisation rétrospective comme des documents dits nativement numériques (sans publication sur un support physique) est également une des premières préoccupations de la Bibliothèque.

Un cadre juridique rigoureux

La numérisation des collections de la BnF respecte en tous points les dispositions relatives à la propriété intellectuelle et met l'accent sur le domaine public (documents dits *libres de droits* ⁸). Toute numérisation de documents sous droits par la Bibliothèque à des fins de diffusion dans Gallica a comme préalable un accord avec les auteurs, leurs ayants droit ou selon les cas, les autres responsabilités associées (éditeurs, diffuseurs, ...) ⁹.

Une politique résolument coopérative

La numérisation est la principale mutation du monde culturel de la fin du xx^e siècle : elle concerne archives, musées, bibliothèques, industries culturelles. La BnF tient compte des réalisations des établissements culturels et en particulier des bibliothèques et s'attache en lien étroit avec le ministère de la Culture et de la communication à développer une politique résolument coopérative autour de la numérisation, sur le plan national comme international, et en particulier européen. Gallica peut être considérée aujourd'hui comme une « bibliothèque numérique de France », grâce à la contribution de plus de 300 partenaires.

La coopération prend principalement trois formes :

- ▶ la numérisation partagée (BnF et partenaires se répartissent les corpus à numériser pour les produire ensemble et éviter les doublons)
- ▶ la BnF complète des numérisations faites par les partenaires, ou inversement
- ▶ sans s'appliquer à la numérisation elle-même, la coopération peut concerner la diffusion. Les partenaires ont le choix entre l'intégration de documents numérisés dans Gallica (métadonnées et fichiers rejoignent les serveurs de la BnF), ou le « moissonnage » ¹⁰

Des financements diversifiés

Les crédits affectés à la numérisation des collections sont issus du budget de la Bibliothèque, qui lui est attribué

⁶ Les personnes intéressées à restituer la numérisation dans l'ensemble de la stratégie numérique de la BnF peuvent consulter utilement le *Schéma numérique de la BnF* (2015), publié sur le site de la BnF.

⁷ Pour en savoir plus sur la politique des publics de la BnF, voir le rapport annuel 2015 : (http://webapp.bnf.fr/rapport/tf4_reconquete_publics.html)

⁸ Le droit d'auteur, régi par le code de la propriété intellectuelle, protège tous les types d'œuvres (œuvres littéraires, créations musicales, graphiques et plastiques, logiciels, créations de mode, etc). Les droits dits « patrimoniaux » qui lui sont liés durent 70 ans après la mort de l'auteur (ou du dernier vivant des collaborateurs dans le cas d'une œuvre collaborative). Ce délai est calculé à partir de la date de divulgation si l'œuvre appartient à une personne morale (société, association). Des dispositions spécifiques sont appliquées aux auteurs et artistes morts pour la France. Par ailleurs, les artistes-interprètes, les producteurs de vidéogrammes et de phonogrammes, et les entreprises de communication audiovisuelle jouissent de droits voisins du droit d'auteur. Sources : le site de l'INPI (Institut national de propriété industrielle, <http://www.inpi.fr/fr/l-inpi/la-proprietie-industrielle/comment-protoger-vos-creations/le-droit-d-auteur.html?S=&print=1>) et les fiches techniques du Ministère de la Culture et de la Communication (<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Fiches-techniques/Fiches-techniques-sur-les-droits-d-auteur-et-les-droits-voisins>) [consultés le 3 décembre 2015]

⁹ En cas de non respect involontaire de droits, la BnF procède dans les meilleurs délais aux corrections nécessaires.

¹⁰ On entend par moissonnage la collecte de métadonnées contenues dans les entrepôts OAI des bibliothèques partenaires pour les intégrer dans les index de Gallica. Dans ce cas, les fichiers ne sont pas intégrés dans Gallica. Reposant sur des normes simples à implémenter et des logiciels open source peu complexes, standard conforme aux standards du web, OAI-PMH définit notamment un protocole de transfert et un format de métadonnées commun (le Dublin Core). Des données hétérogènes et de sources multiples peuvent alors être interrogées via une interface de recherche unique (un exemple de « service ». En l'occurrence ici : Gallica). Les documents ne sont pas dupliqués et leur localisation d'origine n'est pas modifiée et ils restent consultables uniquement sur le site d'origine. Les mises à jour ne se font pas en temps réel mais lors du passage du moissonneur.

par l'État, de dotations spécifiques de sources diverses, par exemple le Centre National du Livre qui finance la plus grande partie de la numérisation des documents imprimés, de crédits de l'Enseignement supérieur (par exemple l'Agence Nationale de la Recherche, des Labex ou Equipex), dans le cadre de projets de recherche universitaire (par exemple le partenariat avec Paris-Sorbonne). Les financements peuvent également provenir de partenariats financiers et scientifiques avec des institutions publiques patrimoniales ou de recherche étrangères, désireuses de reconstituer virtuellement un patrimoine national dispersé. Souhaitant démultiplier les efforts, la BnF fait également appel au mécénat d'entreprises, de fondations ou de particuliers. Elle a également créé une filiale de droit privé, *BnF Partenariats*, en charge de développer des projets de partenariats public / privé de numérisation – notamment en utilisant les sources de financement des « investissements d'avenir » et des apports de sociétés françaises ou étrangères¹¹.

Des technologies éprouvées et évolutives

Si la BnF mène, souvent en partenariat, des activités de recherche et de développement sur la numérisation (par exemple la numérisation en trois dimensions, les techniques de reconnaissance optique de caractères¹², ...), la numérisation de ses collections ne s'en appuie pas moins sur des technologies et des protocoles éprouvés, applicables à grande échelle, normalisés pour la plupart, de sorte à permettre le travail en réseau entre établissements¹³.

Les évolutions technologiques et la diversité des supports d'origine expliquent les qualités inégales de la numérisation dans Gallica. La numérisation des microformes réalisées à des fins de conservation avant l'émergence des techniques numériques, même si elle pose parfois des problèmes de qualité, a toujours été indispensable – et le reste – pour parvenir à la numérisation de collections de très grande ampleur. Elle est en effet beaucoup moins coûteuse que la numérisation des documents originaux. Elle offre également la garantie que la reproduction du document a déjà été jugée nécessaire précédemment pour des raisons de conservation ou de communication. La ré-océrisation, voire la re-numérisation des documents est une contrainte pour toute entreprise de numérisation documentaire commencée au début des années 90, à une époque où la résolution numérique standard était celle de l'image vidéo.

¹¹Voir annexe J les éléments financiers.

¹²La technique d'OCR (*optical character recognition*) permet de situer et de reconnaître les chaînes de caractères dans une image et donc de faire la conversion des mots qui peuvent ensuite être utilisés pour faire une recherche plein texte. Cette conversion est assurée automatiquement par un logiciel. Les caractères sont identifiés à partir de formes mémorisées par le logiciel et de termes déjà connus car présents dans le dictionnaire utilisé par l'outil. Le traitement est composé de différentes étapes : prétraitements (redressement de la page, suppression des bords) ; binarisation (conversion en noir et blanc) ; segmentation (extraction des zones de la page, découpage en lignes, mots, caractères) ; reconnaissance des caractères ; reconnaissance des mots (avec des ressources linguistiques).

¹³Voir annexe I sur les principaux éléments techniques.

LE CHAMP D'APPLICATION DE LA NUMÉRISATION

Le patrimoine national...
... imprimé, musical, graphique et audiovisuel...
... majoritairement dans le domaine public...
... dans un esprit encyclopédique et d'objectivité
... en prenant en compte le paysage numérique
international et national et les évolutions technologiques



La politique de numérisation de la BnF concerne le patrimoine national imprimé, musical, graphique et audiovisuel, très majoritairement dans le domaine public, dans un esprit encyclopédique et d'objectivité, en prenant en compte le paysage numérique international et national et les évolutions technologiques.

Le patrimoine national...

Parce qu'elle est nationale, et à l'instar des bibliothèques nationales de nombreux pays, la BnF a choisi de donner priorité au patrimoine français ou, plus largement, francophone. La langue des documents très largement majoritaire est le français, dans son évolution historique (latin notamment, ancien français, ...) et sa diversité territoriale (langues régionales, par exemple).

Globalement, les priorités sont :

- ▶ le patrimoine national édité en France ou à l'étranger¹⁴
- ▶ les documents en langues étrangères ou issues des pays étrangers qui peuvent se trouver liés à l'histoire de la France – par exemple la presse de la période coloniale ou la presse d'immigration publiée en France...
- ▶ les documents en langues étrangères ou issus des pays étrangers quand ceux-ci ont une importance ou une qualité exceptionnelle dans les collections de la Bibliothèque ou constituent un patrimoine qui peut être considéré comme un bien commun de l'humanité (ainsi, parmi de nombreux exemples, les estampes japonaises, l'estampe flamande, les manuscrits orientaux, la numismatique grecque et romaine, les manuscrits musicaux de grands compositeurs étrangers comme Bach ou Mozart, les phonogrammes, ...)¹⁵. Ces collections contribuent de manière décisive à l'universalisme de la BnF et lui donne une vocation internationale que peu de grandes bibliothèques sont à même d'égaliser.

... imprimé, musical, graphique et audiovisuel...

Dès les premières campagnes de numérisation, la BnF s'est attachée à adopter une orientation résolument multisupports ou multimédia. Manuscrits, imprimés, dessins, gravures et photographies, cartes, partitions, documents sonores..., la numérisation s'applique à tous les types de patrimoine dont elle a la garde pour répondre à une grande diversité de publics et d'usages. La volonté était également d'impliquer l'ensemble des départements de collections dans la numérisation, pour favoriser une mutation homogène des collections et des métiers. La numérisation des objets (objets archéologiques, monnaies,

masques, costumes, maquettes, globes, ...) est en fort développement. L'image animée, largement numérisée pour les salles de lecture, va prendre de l'ampleur dans Gallica.

... majoritairement dans le domaine public...

Si on met de côté les collections audiovisuelles, le patrimoine numérisé est très majoritairement dans le domaine public. La date butoir s'est déplacée et se déplacera au fil du temps : d'abord les années 1914, au début de la numérisation, puis les années 20-30, puis 40, etc. Sauf accords particuliers, la presse est numérisée jusqu'à l'année 1945. La BnF souhaite aller au-delà de cette date. C'est déjà le cas pour certaines revues : par le biais d'accords, de nombreux partenaires ont souhaité que des documents récents soient présents dans Gallica¹⁶. C'est le cas également pour certains documents graphiques (par exemple photographies de spectacles ou maquettes de décors ou de costumes). En 2014, suite à la loi du 1^{er} mars 2012 sur l'exploitation numérique des livres indisponibles du 20^e siècle, votée à l'initiative du ministère de la Culture, la numérisation de ces textes a été lancée¹⁷.

... dans un esprit encyclopédique et d'objectivité

La BnF n'exclut a priori aucun champ du savoir ou de la culture de la numérisation (philosophie, religion, art, littérature, histoire, sciences et techniques, droit, musique, langues et langages, ...) et approche les collections dans un esprit d'objectivité envers les productions éditoriales et plus largement envers l'histoire des goûts et des idées. Elle met à disposition des matériaux pour la recherche et la culture, sans procéder nécessairement à leur contextualisation.

... en prenant en compte le paysage numérique international et national et les évolutions technologiques

L'importance de l'activité de numérisation de la BnF et sa veille technologique font que ses services sont globalement bien informés des évolutions conceptuelles, méthodologiques, technologiques de la numérisation. La BnF est fréquemment consultée par les institutions françaises ou étrangères qui constituent des bibliothèques numériques.

¹⁴On évoque par-là les évolutions du territoire national au fil de l'histoire, notamment l'empire colonial.

¹⁵Il peut se trouver également qu'une segmentation domaine français / domaine étranger soit trop complexe à organiser dans les chaînes de production.

¹⁶Ainsi les revues *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*, *Dix-huitième siècle*, disponibles sur Gallica des origines à 2000. Il en va de même pour des revues de sociétés d'amis d'écrivains, de très nombreuses sociétés savantes, des académies et sociétés nationales (120 conventions de cession de droits à ce jour), ...

¹⁷Voir <https://reliure.bnf.fr/projet-reliure-cadre-légal-et-plus-loin-le-programme-n°10>.

LES MÉTHODES DE NUMÉRISATION

La numérisation de masse
la numérisation de corpus
La numérisation " à la pièce "



La numérisation est organisée par grandes filières de production, qui correspondent aux types de documents (par exemple la presse, l'audiovisuel, les documents spécialisés, ...) et dont certaines peuvent présenter des caractéristiques propres.

Ces filières sont détaillées dans le document quinquennal et les programmes annuels.

Synthétiquement, on peut décrire la granularité de la numérisation en trois approches qui s'appliquent à la plupart des types de documents :

la numérisation de masse

Cette approche est caractérisée par un grand nombre de documents, une ambition d'exhaustivité ou une sélection simplifiée. L'exemple le plus concret en est le traitement des imprimés où on numérise par cote de classification, correspondant néanmoins à une discipline, selon une méthode d'élimination du « non numérisable » pour des raisons juridiques, de nationalité des documents ou encore des raisons techniques ou de conservation¹⁸. La méthode est également utilisée pour certains documents spécialisés, par exemple des cotes entières de cartes et plans. Ce type de sélection a concerné jusqu'à présent près de 80% de la production pour les imprimés.

la numérisation de corpus

Le corpus se définit comme un ensemble de documents (sans limitation de nombre) ayant au moins un point commun. Il peut s'agir des documents d'un seul et même auteur, ou éditeur, d'un type de documents, d'un thème, d'une période, d'une provenance, ou encore de plusieurs de ces éléments combinés. La BnF a de longue date une grande pratique de cette numérisation par corpus¹⁹. Dans les départements d'imprimés, les corpus sont plutôt entendus comme un ensemble documentaire, une sélection intellectuelle, autour d'un thème exploré scientifiquement sur la base de bibliographies savantes. Dans les départements spécialisés, le corpus s'appuie souvent sur des fonds particuliers constitués par regroupement à l'initiative de la Bibliothèque (ainsi les portraits, la topographie, pour les images...) des collectionneurs (les collections Rothschild, Hennin, de Vinck, Chambure ...) ou par des circonstances historiques (par exemple les archives de la prison de la Bastille).

La distinction entre masse et corpus est parfois ténue. Ainsi la numérisation des monnaies grecques est-elle à la croisée de la masse (numérisation intégrale des 120 000 objets menée dans un délai court) et du corpus (une seule et même collection d'un type d'objet homogène).

Il en va de même des projets de BnF partenariats pour le livre ancien antérieur à 1700 ou les disques microsillons, ou encore la numérisation de la presse.

la numérisation « à la pièce »

L'objet documentaire numérisé est dans ce cas une seule pièce, un seul document ou encore une seule unité bibliographique. De façon générale, cette approche résulte de demandes des publics de la BnF (lecteurs, éditeurs, autres bibliothèques, ...). C'est aussi la proposition qui est faite par l'association des Amis de la BnF dans les opérations *Adoptez un livre* ou *Offrez une voix*. On peut y ajouter certaines opérations de gestion interne (numérisation des pièces faisant l'objet de prêts aux expositions extérieures). Si cette numérisation ne relève pas d'une politique documentaire, puisqu'elle résulte de sollicitations, elle n'en est pas moins mise à profit dans les applications numériques de la BnF. Les documents numérisés intégralement sur demande des lecteurs sont versés dans Gallica, lorsqu'ils sont libres de droits.

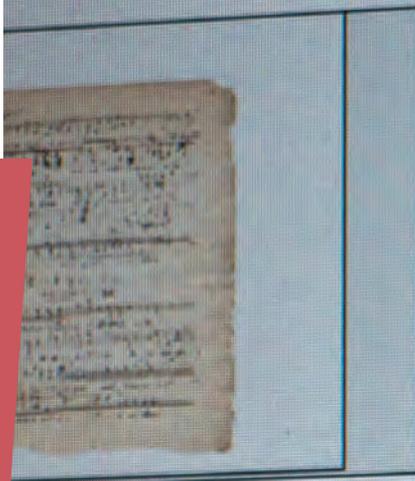
En complément il faut indiquer que la numérisation a pour objectif de donner l'image la plus fidèle possible du document physique. Les documents numérisés le sont toujours intégralement, pages de garde incluses, par exemple pour les imprimés, recto et verso pour les documents planes si ceux-ci présentent au verso des motifs décoratifs ou des mentions. Les documents audiovisuels, quant à eux, sont numérisés au plus proche du format présent sur le support d'origine.

¹⁸À la BnF – quasiment des origines à 1997 – les fonds d'imprimés ont été classés thématiquement (selon la classification Clément). Les lettres du début de l'alphabet ont ainsi été attribuées à la religion (A-D), puis au droit en commençant par le droit canonique (E-F), la géographie et l'histoire (G-P), les sciences et les arts (R-V), enfin la littérature (X-Z), révélant la hiérarchie des savoirs à la fin du XVIII^e s. (voir annexe).

¹⁹Série des *Voyages* (en France, en Italie, en Afrique, réalisée dans les années 2000), le corpus *Anthologie*, qui rassemble presque un millier de trésors de la BnF, le corpus *Utopies*, lié à une exposition des années 2000, les portails avec les pays étrangers comme *La France en Amérique*, ou *France Brésil...*

LES CRITÈRES DE SÉLECTION

L'intérêt patrimonial
L'intérêt documentaire
La conservation



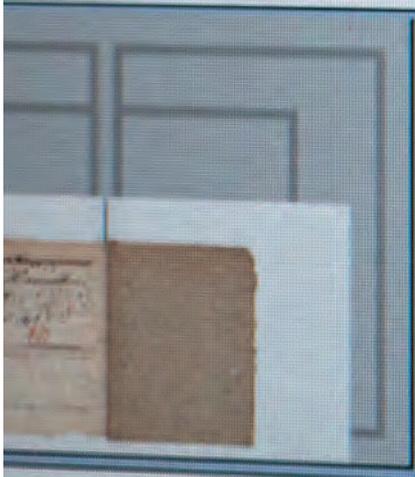
25



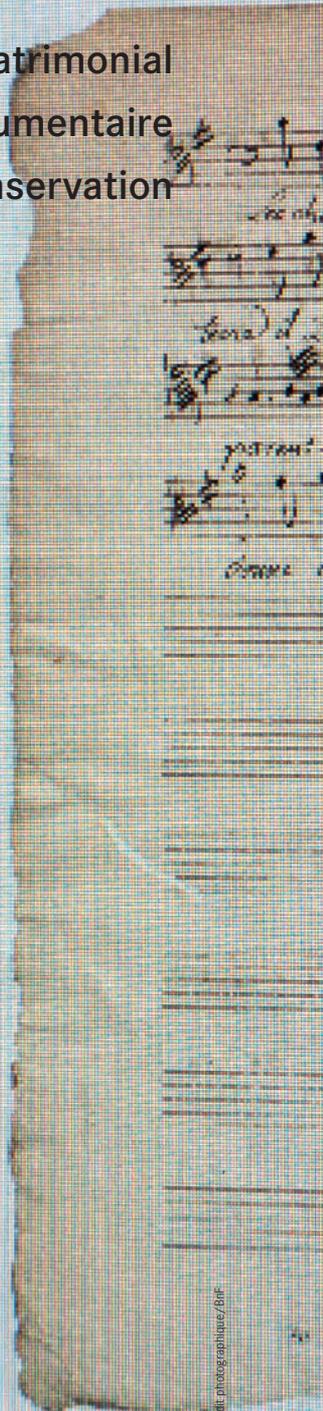
26



27



age en cours



La BnF a fait le choix d'entreprendre la numérisation de ses collections de façon quasi simultanée sur l'ensemble de ses types de collections. L'approche se fait donc majoritairement par grande filière, liée à des supports : les imprimés (ouvrages et publications en série), la presse, les documents spécialisés, l'audiovisuel, ... Pour mieux organiser et suivre la numérisation, ces filières peuvent être décomposées en programmes.

L'ampleur des collections de la BnF peut être appréciée dans le tableau suivant²⁰ :

Imprimés	
Livres et recueils	14 millions ²¹
Périodiques	350 000 titres (dont environ 40 000 titres vivants). Plus de 100 000 de ces titres relèvent de la presse (générale ou spécialisée, de toute périodicité).
Documents spécialisés	
Manuscrits	350 000 volumes (dont 200 000 manuscrits orientaux et 150 000 manuscrits occidentaux)
Estampes et photographies	environ 15 millions
Partitions	2 millions
Cartes plans globes (dont Société de géographie)	900 000
Monnaies et jetons	440 000
Médailles	150 000
Autres objets	60 000 dont 42 000 objets archéologiques, 6 500 costumes, 1 500 marionnettes, plus de 500 masques, ...
Documents relatifs aux spectacles	de 3 à 4 millions
Documents audiovisuels	
Enregistrements sonores	1,5 million
Vidéogrammes	260 000
Multimédias	170 000 dont multisupports : 70 000
Total : environ 40 millions de documents	

La numérisation met l'accent sur les documents du domaine public à dominante française ou francophone, soit probablement plus de 50% de l'ensemble. Le gisement de l'ensemble des collections à numériser représente donc environ 20 millions de documents et ensembles de documents. De ce chiffre, il faut soustraire encore les documents dont l'état physique ne se prête pas aux manipulations liées à la numérisation ou aux matériels disponibles, qu'on estime à environ 10%. Le gisement est donc de l'ordre de 18 millions de documents. Fin 2016, Gallica propose environ 4 millions de documents, numérisés pour la plus grande partie en un peu plus d'une dizaine d'années. Au rythme actuel de numérisation, trois décennies permettraient d'approcher une forme d'exhaustivité. Cette approche est cependant très théorique. Le taux actuel de fonds numérisé est en effet très variable selon les filières ou les départements. S'il est de l'ordre de 20% pour la Réserve des livres rares, de 10% pour le département des Manuscrits, il est sans doute bien inférieur à 5% pour la presse.

À la question souvent posée de l'ambition d'exhaustivité de la numérisation de ses collections, la BnF répond qu'elle n'inscrit pas la numérisation dans une ambition d'exhaustivité quantitative, mais raisonnée, appuyée sur trois grands critères :

- ▶ L'intérêt patrimonial
- ▶ L'intérêt documentaire
- ▶ La conservation

L'intérêt patrimonial

Sous cette notion, peuvent être rassemblés tous les critères qui évoquent la qualité formelle, le caractère majeur de l'œuvre ou du document pour l'histoire des goûts et des idées, son ancienneté, sa notoriété, sa rareté, sa provenance prestigieuse, etc. qui peuvent faire qualifier l'objet de trésor, chef d'œuvre, œuvre de référence, ou majeure, etc. La notion s'applique aux imprimés (grands textes, grandes éditions, grandes provenances, exemplaires annotés par des auteurs, des possesseurs...) comme aux documents spécialisés (grands manuscrits médiévaux ou modernes, portulans – cartes manuscrites du Moyen Âge, maîtres de la musique, de l'affiche, de l'estampe, de la scénographie, de la photographie, collections archéologiques et d'objets d'art...). Une attention particulière est accordée à tous les *unica*, c'est-à-dire les pièces que la BnF est seule à conserver, à l'échelle nationale ou internationale.

L'intérêt documentaire

Cette notion concerne plutôt la qualité informative des documents et les intérêts de la recherche, académique

²⁰Beaucoup de ces chiffres varient notablement selon les sources. Ils sont donnés par support et non par département (ainsi plusieurs départements disposent de cartes ou de photographies). Ils agrègent souvent des ensembles et documents en feuilles, (albums de photographies et photographies, cartes en feuille et atlas, ...). Ils sont donc à prendre avec beaucoup de réserve.

²¹Ce chiffre concernerait plutôt le nombre de volumes (dont environ 2 millions sur les sites de Richelieu, de l'Arsenal et de l'Opéra). Selon plusieurs sources, le nombre de titres serait plutôt de l'ordre de 10 millions.

ou non, globalement thématique (par exemple la presse coloniale, l'histoire de la médecine, la philosophie des Lumières, les romans populaires du XIX^e siècle, la représentation des rois de France, les méthodes d'enseignement de la musique, les monnaies...). Il s'agit souvent de grands ensembles, de cotes sollicitées significativement par les publics. L'intérêt documentaire est garanti par des statistiques de consultation, des témoignages d'intérêt, des sollicitations de collaborations ou de partenariats²². La Bibliothèque peut également sélectionner des documents dont elle estime qu'ils méritent d'être portés plus largement à la connaissance des chercheurs, même s'ils sont peu connus ou sollicités.

La conservation

La numérisation est utilisée particulièrement pour reproduire

- ▶ les documents les plus fragiles (ainsi certains types de photographies comme les daguerréotypes, les autochromes, les plaques de verre...) ou ceux dont la conservation est en péril (notamment les collections dites hors d'usage ou encore globalement les collections de presse)
- ▶ les documents dont la communication sous forme originale est malaisée, – les documents de grande taille par exemple – cartes, gravures, affiches, – ou encore l'ensemble des objets en trois dimensions comme les maquettes, les masques, les globes, ...
- ▶ les documents dont le maintien dans des conditions de sûreté satisfaisantes est délicat (ainsi par exemple, les monnaies, souvent de très petite taille et dont le marquage est impossible).

Grâce à la possibilité d'explorer en détail les documents par zoom, le numérique ajoute en outre à la qualité de la consultation pour certains types de documents (manuscrits enluminés, cartes, monnaies, ...).

On doit mentionner aussi la numérisation massive de microformes – pour les imprimés comme pour les documents spécialisés – qui permet de gagner en confort de consultation et de se dispenser de la communication des supports physiques, sauf nécessité de recourir à l'original. Il faut mentionner enfin le cas particulier de l'audiovisuel, pour lequel le numérique est la technologie de sauvegarde, d'homogénéisation des supports, et de communication dans les salles de lecture²³.

Quand la BnF procède à la numérisation de certains types de documents, ainsi les portulans, ou encore les autochromes, les intérêts patrimoniaux, documentaires et de conservation se recourent évidemment.

La numérisation avance simultanément selon ces trois points de vue, qui sont souvent convergents, intérêt patrimonial, intérêt documentaire, intérêt de conservation et de communication. Les proportions peuvent varier selon les départements et les types de documents.

Autour de ces ensembles de critères, jouent divers facteurs qui ont un rôle important :

- Les principes de sélection restent subordonnés aux filières techniques disponibles : certains types de documents, ainsi les reliures trop serrées, certains grands formats restent inaccessibles à la numérisation (par exemple les formats au-dessus du A0).
- Les collections peuvent nécessiter des travaux de préparation ou de restauration avant la numérisation. La charge de ces travaux préalables influe sur le rythme de numérisation.
- L'existence de référencement (catalogage) : la numérisation doit s'accompagner des informations qui permettent d'identifier et d'interpréter le document. Cette nécessité agit comme une forte contrainte pour les collections spécialisées, où le classement remplaçait souvent le catalogage, mais aussi lorsque les collections sont traitées en recueils, légendées par lots, ou quand les notices de description présentent des défauts. Des chantiers de corrections des données ou de mise à niveau du catalogue sont alors à mettre en place. Ils sont coûteux en ressources humaines et peuvent freiner l'effort.
- Les opportunités en matière de partenariat et de mécénat peuvent accélérer un projet, voire jouer un rôle de déclencheur dans son lancement quand elles apportent des moyens supplémentaires à la BnF pour la réalisation des projets ainsi parmi de nombreux exemples
 - ▶ la Fondation d'entreprise Total pour la numérisation de 250 manuscrits arabes, turcs et persans en 2010,
 - ▶ le partenariat Public Privé de BnF-Partenariats avec Believe digital et Memnon qui a permis la numérisation de l'intégralité du dépôt légal de microsillons entre 1949 et 1962.
 - ▶ l'entreprise japonaise Dai Nippon Printing qui a permis en 2015 la numérisation en trois dimensions d'une partie de la collection de globes
 - ▶ le partenariat avec la bibliothèque nationale d'Israël pour la numérisation de la totalité du fonds de manuscrits hébreux en 2016-2017). Il faut signaler aussi l'essor de la numérisation dans le cadre de projets de recherche.
- Les évolutions globales de l'établissement, dont le meilleur exemple est le projet de rénovation du site historique de la Bibliothèque, le site Richelieu (2010-2020), qui rassemblent les collections spécialisées. L'échéancier des phases de travaux et de transferts de collections est un paramètre important de la programmation de la numérisation.

L'énoncé de tous ces critères permet de dessiner l'ambition globale de la BnF quant à la numérisation de ses collections.

- ▶ Ont vocation à être traités dans une démarche visant à l'exhaustivité selon le périmètre défini (sauf exception, priorité aux fonds français ou francophones dans le

²²Pour les départements du site Richelieu, la communication des collections n'est pas encore informatisée. L'observation plus fine des usages ne sera possible qu'après informatisation de cette communication, au cours de la période 2016-2017.

²³La relation entre conservation et numérisation est décrite de façon plus étendue dans la charte de la conservation, mise à jour en 2015 (disponible sur le site web de la Bibliothèque).

domaine public en état d'être numérisé)

- les documents les plus anciens de toutes catégories (manuscrits, incunables, portulans,...),
- la Réserve des livres rares, qui rassemble les livres les plus précieux, depuis la Bible de Gutenberg jusqu'aux grands livres illustrés contemporains.
- les Réserves des départements spécialisés qui conservent les documents les plus précieux
- les objets muséaux (peintures, dessins, masques, maquettes en trois dimensions, bijoux, monnaies, ...)
- les documents audiovisuels les plus fragiles (tous documents magnétiques – bandes magnétiques, vidéocassettes, ...) et optiques (compact discs, DVDs, CDROMs, ...),
- les documents hors d'usage ou dont la survie matérielle est menacée.
- les fonds de microformes résultant des plans de sauvegarde successifs,
- certains fonds ou parties de collections (Archives de la Bastille, grandes collections historiques du département des Estampes, cartes anciennes, grands fonds d'images de presse, ...) où une sélection serait peu légitime ou trop coûteuse. Sur ces fonds, les programmes annuels feront un point particulier.

► Les fonds étrangers les plus exceptionnels (par exemple estampes japonaises, imprimés chinois anciens, manuscrits persans ou hébreux, ...) feront plutôt l'objet d'une sélection anthologique, sauf partenariat qui permettrait d'atteindre l'exhaustivité.

► Une grande partie du patrimoine imprimé plus courant (livres, revues, presse, ...) mériterait d'être numérisée. On peut considérer en effet qu'à terme, les documents ne disposant pas d'avatar numérique facilement disponible retiendront de moins en moins l'attention des publics. L'ampleur des collections oblige cependant à être sélectif, en prenant d'abord en considération l'importance patrimoniale ou documentaire, la rareté, l'existence d'occurrence sur le web, le niveau de consultation par les chercheurs d'aujourd'hui, l'état de conservation. Faute de séries statistiques établies sur de longues périodes²⁴ pour la consultation par les chercheurs, des objectifs *in abstracto* paraissent difficiles à expliciter. Une numérisation plus ou moins exhaustive sera surtout affaire de moyens sur le long terme.

²⁴Les données de communication des documents imprimés ne peuvent faire l'objet de traitements statistiques de grande ampleur avant 2007.

LES PROCESSUS DE DÉCISION ET DE VALIDATION

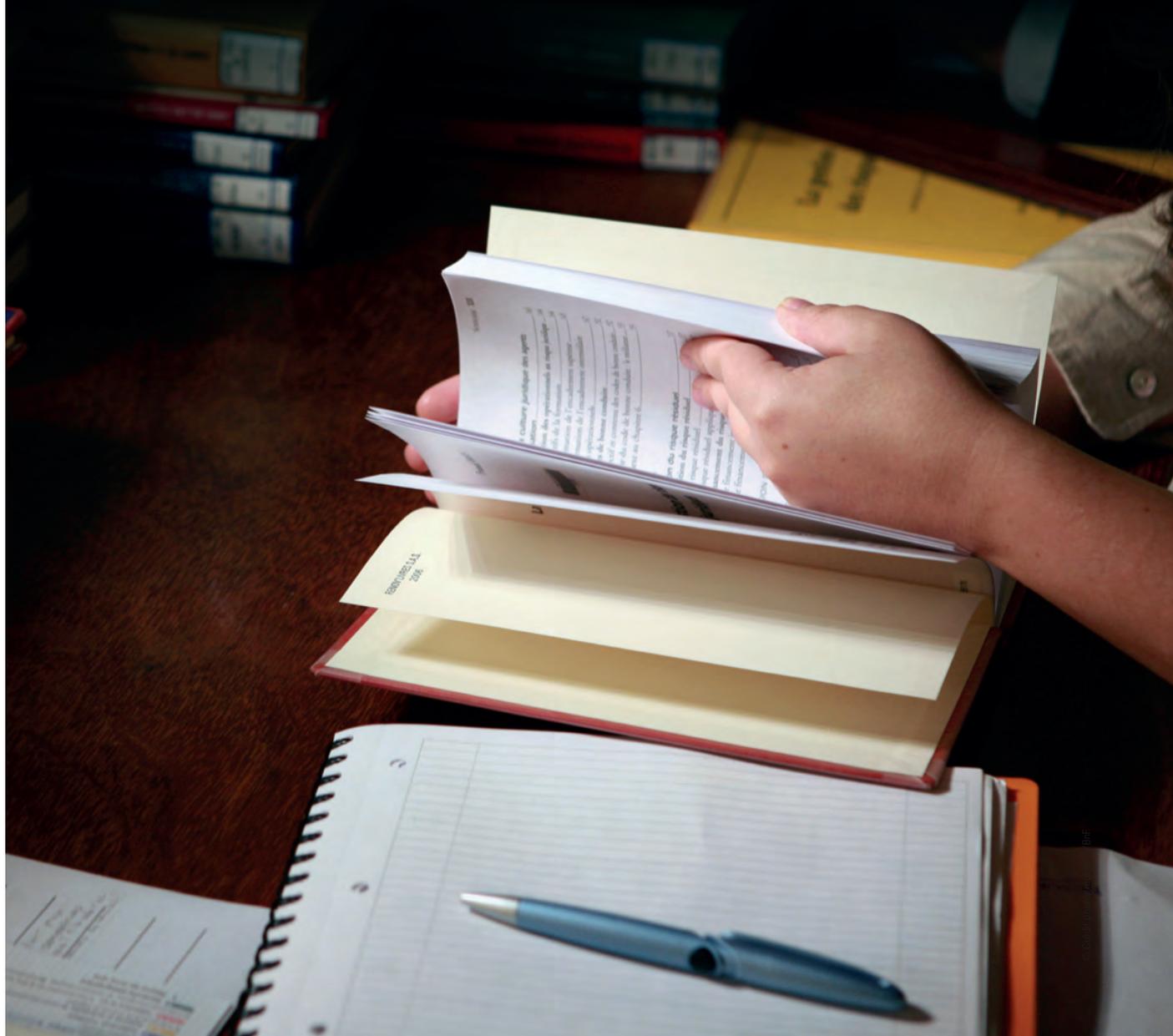


La numérisation de la BnF est planifiée et concertée. Après une étape de recueil annuel des besoins et des projets – qui tient compte notamment des ressources humaines qui peuvent être affectées aux projets – et les arbitrages budgétaires, a lieu une étape de concertation coordonnée entre les départements pour définir des quotas de production, soit le nombre d’images ou de pages numérisées qui sera affecté à chaque collection. Pour les imprimés, l’ampleur des magasins et des collections conservées par les différents départements est un paramètre important de répartition. Pour les collections spécialisées, les moyens humains disponibles sont un paramètre déterminant, du fait des contraintes relatives aux métadonnées.

Après cette concertation, un programme annuel est arrêté puis présenté dans ses grandes lignes aux diverses instances de validation, en particulier le comité de la numérisation interne à la BnF.

De façon régulière, et plus générale, les orientations de la numérisation à la BnF sont évoquées ou présentées aux instances de gouvernance de l’établissement et notamment son Conseil scientifique, au comité de la coopération auquel siège le ministère de la Culture, ...

ÊTRE ACTEUR DE LA POLITIQUE DOCUMENTAIRE DE NUMÉRISATION



La BnF souhaite accroître l'interaction avec les établissements culturels et de recherche comme avec ses publics, dans le domaine de la numérisation comme pour l'ensemble de ses missions. De façon générale, elle souhaite passer à un mode « collaboratif » avec ses partenaires et ses publics – développement des activités de réseaux sociaux, participation des usagers aux corrections d'OCR, ...

En matière de politique de numérisation comme de politique documentaire, prendre appui sur les usages réels ▶ notamment par les statistiques –, sur la demande, sur l'expression des besoins des usagers concourt à la légitimation des choix de la Bibliothèque. Ceci doit s'appliquer aux grandes orientations mais également de façon plus quotidienne, passer par l'expression de besoins particuliers, le signalement de manques, de défauts... Beaucoup des corrections effectuées sont directement issues des remarques des internautes et cette pratique sera encore développée.²⁵

La numérisation peut également être soutenue par des entreprises, des fondations, comme par des particuliers.

Les points d'entrée pour engager des partenariats de numérisation de collections avec la BnF ou contribuer à Gallica sont :

-Bibliothèques musées archives associations etc. en France

Département de la coopération
(gallica@bnf.fr)

-Bibliothèques musées archives associations etc. à l'étranger

Délégation aux relations internationales
(relations.internationales@bnf.fr)

-Programmes de recherche (collections, numérique, ...)

Pour conclure des partenariats de recherche impliquant de la numérisation ou de la mise à disposition de documents numériques.

Délégation à la stratégie et à la recherche
(recherche.coordination@bnf.fr)

-Remarques et suggestions d'utilisateurs

Sommaire de Gallica *Contribuer*, puis *Participer à la qualité de Gallica*. Des informations sur les programmes ou la numérisation de tel ou tel type de collections (par ex. les manuscrits, l'audiovisuel, les cartes, ...) peuvent également être obtenues auprès des différents départements de collections.

(voir www.bnf.fr)

Les points d'entrée pour soutenir la numérisation :

-Entreprises, fondations, associations, ...

Délégation au Mécénat
(mecnat@bnf.fr)

-Les **particuliers** peuvent également contribuer par le biais du programme *Adopter un livre* sur le site de l'*Association des Amis de la Bibliothèque nationale de France*

(voir www.amisbnf.org)

²⁵Les questions et remarques des usagers sont traitées dans SINDBAD (Service d'Information Des Bibliothécaires A Distance) et réparties entre les différents départements concernés, ce qui leur permet de leur donner une suite plus rapidement.

ANNEXES

Bref historique de la politique de numérisation

État actuel de Gallica par type de document

Répartition 2016 de Gallica par disciplines

Répartition de Gallica par siècle

Quelques informations sur les publics et la consultation de Gallica

Brève synthèse sur la coopération

Principaux éléments techniques

Quelques éléments financiers

Bibliographie

A. Bref historique de la politique de numérisation

Très synthétiquement, on peut distinguer quatre grandes périodes.

1989-1992 : un dessein ambitieux

L'historique du projet de la BnF est désormais bien connu. Les initiateurs (J. Attali notamment) avaient bien l'idée d'un bâtiment d'une taille relativement modeste, et d'une très grande bibliothèque numérique (on évoquait déjà à l'époque des chiffres au-delà du million de documents) consultables à distance. Les analyses ont montré rapidement que cette orientation n'était pas envisageable pour des raisons techniques – la reconnaissance optique de caractères en était à ses balbutiements, Internet n'était encore utilisé que par des chercheurs – et économiques – les coûts induits n'étaient pas à la portée d'un projet pourtant doté de moyens considérables. La vision n'était pourtant pas utopique. On peut dire qu'aujourd'hui, la numérisation du patrimoine et Internet se sont rapprochés de la vision originelle des concepteurs.

1993-2005 : une bibliothèque de l'« honnête homme »

Une vision réaliste a substitué au projet originel la création d'une sorte de bibliothèque de l'honnête homme du ^{xx} siècle. Il s'agissait de réunir un fonds numérique des œuvres significatives de l'Antiquité à nos jours en établissant, discipline par discipline, une sélection d'œuvres, de critique et de périodiques consacrées par la postérité. Les documents sont numérisés en mode image, seules leurs tables des matières sont saisies en mode texte. Le chiffre de 100 000 ouvrages qui en définissait le volume est encore inscrit dans les mémoires, et il a été atteint dans les années 2000. Une partie du fonds (environ 20 000 titres) était sous droits – essentiellement des éditions critiques de classiques. Il n'a été que peu utilisé du fait de difficultés d'accord avec les éditeurs. Gallica est lancé en 1997. Le fonds numérique s'enrichissait alors d'environ 5 000 à 8 000 documents par an, choisis très rigoureusement et validés par des comités de pilotage, au plus haut niveau de l'établissement. Dès le début des années 90, l'EPBF²⁶ avait inclus l'audiovisuel et l'image fixe dans le champ du projet (pour des volumétries qui ont été ramenées peu à peu de plusieurs millions à 300 000 images, dont la moitié devait provenir des départements spécialisés).

2005-2009 : l'émergence de la numérisation de masse

Fin 2004, Google annonce son vaste projet de bibliothèque numérique mondiale (12 millions de livres) qui bouleverse le paysage de la numérisation du patrimoine culturel. Sous l'impulsion de J.N. Jeanneney, le président de la BnF à l'époque, la BnF s'efforce de passer elle aussi à une numérisation de masse, ce qu'elle parvient à réaliser en 2007. L'océrisation devient systématique, puis rétrospective sur le déjà numérisé. Le projet est étendu à la presse antérieure à

1945. Un peu plus tard, en 2009, sous l'impulsion cette fois de B. Racine, président nommé en 2007, cette approche quantitative de masse est adoptée également pour les documents spécialisés, grâce à des choix budgétaires volontaristes. Alors que la numérisation « traditionnelle » consistait en une sélection, la numérisation en nombre procède plutôt par « désélection ». Une fois choisis un fonds, ou une cote, on en retire les documents sous droits, les documents étrangers et on numérise tous les documents dont on estime qu'ils peuvent supporter les opérations de traitement (ce qu'on dénomme la « numérisabilité » des collections).

Depuis 2009 : Une politique de numérisation plus collective, plus diversifiée dans sa politique documentaire comme dans ses financements.

Dans cette période, la politique de numérisation connaît de nouvelles mutations. La numérisation de masse continue mais on s'efforce de mieux équilibrer masse et corpus. Elle accentue la priorité sur les documents que la Bibliothèque est seule à détenir – une tendance forte dans les grandes bibliothèques patrimoniales depuis la « rupture » Google. Elle s'appuie de plus en plus sur une logique de partenariat avec des établissements français, étrangers ou dans le cadre de projets bilatéraux, nationaux ou internationaux, en particulier européens. Elle ouvre notamment ses chaînes de numérisation aux collections d'autres bibliothèques et centres documentaires (sur la base de 30% pour les documents imprimés). Elle diversifie ses sources de financement – subventions du CNL, crédits européens, crédits recherche de l'ANR, mécénat, ... Elle s'accompagne d'une médiation numérique plus intense (newsletter, facebook, blog, etc.). Enfin technologiquement, elle franchit encore un palier qualitatif (passage à la couleur pour l'imprimé, et à une résolution de 400 DPI) et s'adapte aux univers de lecture depuis l'ordinateur personnel jusqu'au smartphone en passant par les tablettes numériques (epub). De même, on accentue la prise en compte des publics handicapés (Daisy, EPUB3). Sans préjuger de l'avenir, on peut dire que la politique de numérisation est parvenue à la maturité.

B. État actuel de Gallica par type de document

Nombre de documents consultables sur le site Gallica :

- ▶ 346 091 Livres, dont 239 989 en mode texte
- ▶ 1 778 389 Fascicules de presse et revues
- ▶ 267 011 Documents iconographiques pour plus d'un million d'images
- ▶ 61 310 Manuscrits
- ▶ 40 441 Partitions
- ▶ 53 916 Documents cartographiques pour plus de 100 000 images
- ▶ 40 000 Documents sonores
- ▶ 178 784 Objets
- ▶ 13 Vidéos

²⁶EPBF : Établissement Public de la Bibliothèque de France

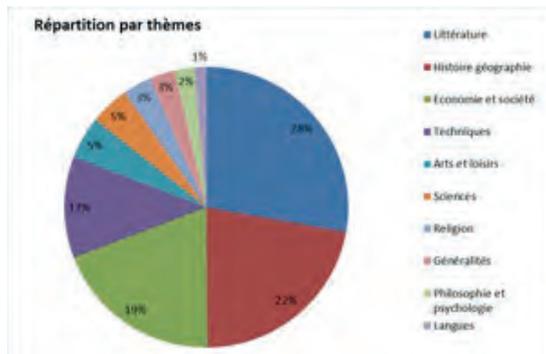
On comptait de plus à fin décembre 2016 309 431 documents référencés dans Gallica par moissonnage (provenant de 79 partenaires).

C. Répartition 2016 de Gallica par disciplines

Le moteur de recherche de Gallica permet d'effectuer une recherche « par thème » sur un ensemble composé d'environ 280 000 notices de livres (97% de livres, 3% de presse et revues). On constate dans les résultats une prédominance de la littérature, suivie de l'histoire – géographie et du thème « économie et société ».

Résultats pour chacun des dix thèmes définis :

Thème	Nombre de résultats
Littérature	79 701
Histoire et géographie	63 130
Économie et société	55 252
Techniques	33 447
Arts et Loisirs	14 021
Sciences	12 809
Religion	9 447
Généralités	8 274
Philosophie et psychologie	6 489
Langues	3 869
Total	286 439



Il est également possible via le mode « recherche avancée » de circonscrire le périmètre de la recherche en choisissant une collection numérique parmi les 39 suivantes :

- Abécédaires
- Arts de la marionnette
- Cartes marines sur parchemin
- Collection Charles Cros
- Collection De Vinck (Histoire de France, 1770-1871)
- Collection Gaignières (histoire de France)
- Collections basques
- Décors du Salon Louis XV
- Documents consacrés à la Première Guerre mondiale
- Fonds régional : Alsace
- Fonds régional : Aquitaine
- Fonds régional : Auvergne
- Fonds régional : Bourgogne

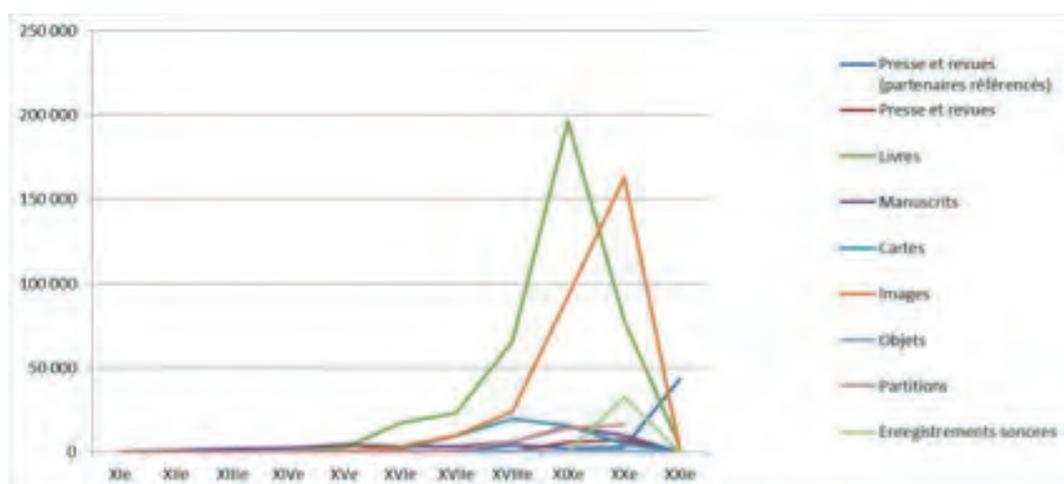
- Fonds régional : Bretagne
- Fonds régional : Centre-Val de Loire
- Fonds régional : Champagne-Ardenne
- Fonds régional : Corse
- Fonds régional : Franche-Comté
- Fonds régional : Guyane
- Fonds régional : Haute-Normandie
- Fonds régional : Ile de France
- Fonds régional : Languedoc-Roussillon
- Fonds régional : Limousin
- Fonds régional : Lorraine
- Fonds régional : Martinique
- Fonds régional : Mayotte
- Fonds régional : Midi-Pyrénées
- Fonds régional : Nord-Pas de Calais
- Fonds régional : Paris
- Fonds régional : Pays de la Loire
- Fonds régional : Picardie
- Fonds régional : Poitou-Charentes
- Fonds régional : Provence-Alpes-Côte d'Azur
- Fonds régional ; Réunion
- Fonds régional : Rhône-Alpes
- France-Amérique
- France-Bésil
- France-Chine
- France-Japon
- Globes
- Jetons français
- Masques
- Monnaies carolingiennes
- Monnaies françaises modernes
- Monnaies grecques
- Monnaies mérovingiennes
- Monnaies obsidionales
- Monnaies romaines impériales
- Monnaies impériales républicaines
- Monnaies royales françaises
- Musée des Monnaies médailles et antiques
- Papiers peints du XVIII^e siècle
- Photographie de spectacle de l'après-guerre à nos jours
- Photographies de l'Agence Meurisse
- Photographies de l'Agence Mondial Photo-presse
- Photographies de l'Agence Rol
- Photographies du XIX^e siècle
- Photographies sur plaques de verre de la Société de Géographie
- Tissus du XVIII^e siècle, Collection Richelieu
- Transibérien

Enfin Gallica propose dans le cadre de la médiation un vaste ensemble de corpus.

D. Répartition de Gallica par siècle

Répartition indicative des documents de Gallica par siècle à la date de fin 2015 (remarque : le calcul porte sur 919 000 documents. Les documents antérieurs à l'an mil ne sont pas comptabilisés dans le tableau ci-dessous : monnaies grecques et romaines, manuscrits carolingiens par exemple).

	XI ^e	XII ^e	XIII ^e	XIV ^e	XV ^e	XVI ^e	XVII ^e	XVIII ^e	XIX ^e	XX ^e	XXI ^e	Total
Presse et revues (partenaires référencés)							6	4 419	2 568	2 953	43 323	53 269
Presse et revues		1			1		46	348	5 744	7 280	128	13 548
Livres	38	22	6	11	2 598	17 093	23 355	65 521	196 530	77 890	3 297	386 361
Manuscrits	637	1 154	2 024	2 582	4 379	3 269	3 395	5 698	15 095	9 864	30	48 127
Cartes	612	7	5	24	58	1 970	9 679	19 616	15 813	4 660	2	52 446
Images	477	9	3	13	660	3 392	9 578	24 013	92 110	162 797	1 060	294 112
Objets	3	3	12	6	17	46	15	687	37	174	58	1 058
Partitions	2	1	4	4	9	273	921	6 555	14 438	16 370		38 577
Enregistrements sonores									4	32 390	20	32 414
Total	1 769	1 197	2 054	2 640	7 722	26 043	46 995	126 857	342 399	343 78	479 18	919 912



E. Quelques informations sur les publics et la consultation de Gallica

Faisant suite à celle de 2011, une grande enquête quantitative auprès des usagers de Gallica a été menée en octobre 2016 par la société TMO Régions, dans le cadre d'un programme de recherche porté par la BnF, le labex Obvil et Télécom ParisTech. 7 600 internautes ont répondu à un questionnaire bilingue (français / anglais) en ligne, soit presque deux fois plus que lors de la précédente enquête. Cette forte augmentation, proportionnellement plus importante que celle de l'audience du site depuis 4 ans, vérifie le fort attachement des gallicanautes au service offert. Au sein des 95% de satisfaits, la part des « tout à fait satisfaits » a notablement augmenté de 2011 à 2016 (+12 points) pour atteindre 48%.

S'agissant des profils d'usagers, l'enquête rappelle que si les gallicanautes résident majoritairement en France, 28% des répondants se connectent de l'étranger : 16% d'un autre pays d'Europe (avec, en tête, l'Italie), 4% des États-Unis et du Canada, 3% du Maghreb. L'âge moyen est nettement plus élevé que celui des salles de lecture de la BnF et a même augmenté depuis la précédente enquête, passant de 48 à 54 ans ; avec un niveau d'étude et des profils socioprofessionnels toujours supérieurs à la moyenne française. 38% des répondants disent par ailleurs avoir déjà fréquenté les espaces physiques de la BnF. Concernant la

fréquence de visite, celle-ci se maintient à un niveau élevé : un répondant sur deux déclare venir au moins une fois par semaine sur Gallica, et 11% tous les jours ou presque. Seuls 3% consultent Gallica sur smartphone, « souvent » ou « à chaque fois », usage minoritaire que vérifient les données de connexion. Évolution notable par rapport à 2011 : la progression, dans l'usage des documents, de la consultation attentive en ligne (73% disent le faire « souvent » ou « à chaque fois »), désormais légèrement plus importante que le téléchargement – effet probable de l'accroissement du confort de lecture et des outils qui lui sont dédiés (zoom). Sans surprise, les domaines les plus consultés restent l'histoire, suivie par la littérature, puis l'art – même si un répondant sur trois déclare consulter plus de trois domaines –, avec un bon niveau de notoriété et d'usage de la diversité des types de documents accessibles (à l'exception des objets). Les consultations de Gallica sont majoritairement le fait de projets de recherche précis, dans le cadre de ses études, de son métier ou d'une recherche personnelle. Malgré cela, les usages exclusifs de Gallica sont rares (seulement 28%), la plupart des répondants déclarant plusieurs motifs qui croisent usage professionnel / personnel, studieux / ludiques, etc. Derrière cette forte porosité des motifs, qui rend désormais délicate toute typologie des usagers, ce sont finalement les recherches personnelles qui pèsent le plus dans les consultations – mutation importante par rapport à la précédente enquête.

F. Brève synthèse sur la coopération

Depuis la fin des années 2000, la BnF a fait de la numérisation l'axe central de sa politique de coopération en France comme avec l'étranger. La coopération numérique, priorité de la politique de coopération nationale de la BnF, vise à créer, diffuser, valoriser et préserver, par le biais d'une bibliothèque numérique conçue collectivement, les plus vastes ensembles possibles de ressources patrimoniales numérisées, quels que soient la localisation des collections et le statut des contributeurs.

Trois enjeux majeurs

Cette action initiée par la BnF répond à trois enjeux majeurs. Le premier est d'ordre culturel et se lit en regard des missions fondatrices de la BnF : offrir au citoyen un accès aisé au patrimoine, national, régional, local ; garantir cet accès sur la longue durée ; rendre plus visibles sur internet la culture et la langue françaises en atteignant une masse critique ; enfin, mettre de nouveaux matériaux à la disposition de l'enseignement et de la recherche en révélant des sources inédites ou méconnues. Cette coopération vise également à permettre des économies, grâce à une mutualisation des ressources et des infrastructures. Par les questions et défis qu'elle soulève et relève autour de l'interopérabilité, de la description des ressources ou encore par les occasions qu'elle offre de développer de nouvelles formes de médiation, cette coopération répond aussi à l'enjeu professionnel que constitue l'évolution des compétences et des services, au sein d'une communauté dynamique, qui partage et échange. À la marge de ces enjeux majeurs, les formes de coopération numérique déployées par la BnF s'adaptent à un contexte administratif et politique changeant, tel celui qu'a connu la France en raison de la réforme territoriale appliquée à partir du 1^{er} janvier 2016, et qui a suspendu plusieurs partenariats régionaux.

Les bases de la coopération

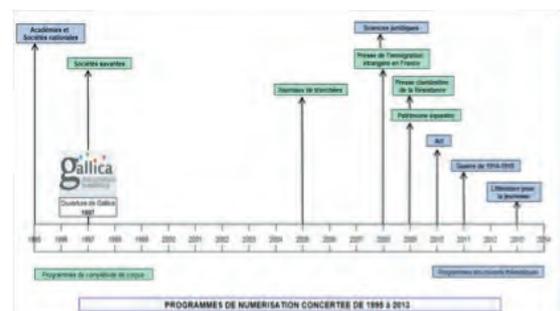
Des programmes structurants

Ces programmes concernent tous types de supports et de partenaires, dans le monde des bibliothèques et de la recherche, en France comme à l'étranger. Ils s'inscrivent dans le cadre de projets bilatéraux ou multilatéraux et s'adossent le plus souvent à certaines exigences telles un signalement systématique au préalable, une complémentarité documentaire maximale pour que les ensembles constitués soient cohérents et fassent sens, ou la recherche de l'exhaustivité des ensembles documentaires constitués.

Sur le plan national, en ce qui concerne les imprimés, la démarche développée avec les partenaires est complémentaire de celle que la BnF mène depuis 2007 en numérisant massivement ses propres collections. Les documents sont choisis en fonction de critères : linguistique : imprimés en français et dans les langues de France ; juridique : titres du domaine public ou dont les droits ont été négociés ; intellectuel : constitution de corpus thématiques ou d'intérêt régional d'envergure, très visibles sur internet.

Sont ainsi mis en œuvre :

- des programmes thématiques : sciences juridiques, lancé en 2008 en co-pilotage avec la Bibliothèque Cujas, le projet compte aujourd'hui 20 partenaires français, histoire de l'art, lancé en 2010 en co-pilotage avec l'institut national d'histoire de l'art, avec 14 partenaires ; littérature patrimoniale pour la jeunesse, lancé en 2013 en co-pilotage avec l'Heure joyeuse, avec 11 partenaires ; Première Guerre mondiale, lancé en 2013 avec 9 partenaires
- des programmes d'intérêt régional : histoire des territoires ; revues des sociétés savantes
- des programmes visant à compléter des corpus plus modestes, mais précieux sur le plan documentaire et scientifique, concernant aussi bien des imprimés (publications des académies et sociétés nationales ; presse clandestine 1939-1945 ; presse de l'immigration étrangère en France, patrimoine équestre, ...) que des documents spécialisés (manuscrits, portulans, ...)



La coopération internationale de son côté vise quatre objectifs principaux : constituer des ensembles numérisés permettant de relier des histoires et des patrimoines communs entre la France et des pays ayant partagé un temps son histoire ; contribuer à des programmes de réunification de patrimoines nationaux dispersés physiquement dans différentes bibliothèques dans le monde ; contribuer à la numérisation et à la diffusion de documents considérés comme des « biens communs de l'humanité » du fait de leur valeur patrimoniale ; contribuer à la sauvegarde de patrimoines fragilisés ou en danger. Elle permet de faire entrer dans Gallica des contenus très diversifiés, dans le cadre de programmes européens (Biodiversity Heritage Library for Europe, Europeana Regia, Europeana Collections 1914-1918), bilatéraux (France-Chine, France-Brésil, France-Japon), multilatéraux (International Dunhuang Project) ou sur projets (Roman de la rose, numérisation de manuscrits hébreux, chinois, arabes, turcs et persans, etc.), souvent ouverts aux autres bibliothèques françaises ainsi associées à des projets scientifiques à large visibilité.

Des enrichissements mutuels

Tous les programmes de numérisation soutenus par la BnF visent à enrichir Gallica, mais, de différentes manières, les ressources patrimoniales numérisées sont également une source d'enrichissement pour les autres bibliothèques numériques.

Les partenaires intégrés bénéficient de l'ensemble des fonctionnalités présentes et à venir de Gallica. Le dis-

positif « Gallica marque blanche » vise en particulier à permettre aux bibliothèques ne disposant pas de bibliothèque numérique de profiter de l'infrastructure développée pour Gallica. La BnF a en effet souhaité mutualiser les moyens qui ont été alloués à celle-ci et faire bénéficier ses partenaires de son savoir-faire. Les deux premiers sites ainsi créés sont la bibliothèque Numistral de la Bibliothèque Nationale Universitaire de Strasbourg en 2013 et le site de la Grande Collecte en partenariat avec le Service Interministériel des Archives de France en 2014. En 2016, trois nouveaux projets Gallica marque blanche ont été engagés : la bibliothèque numérique de Rouen Nouvelles Bibliothèques (Rotomagus), un premier projet international, la bibliothèque numérique patrimoniale du Réseau Francophone Numérique, la bibliothèque numérique du ministère des Affaires étrangères et du développement économique. D'autres bibliothèques manifestent leur intérêt. À l'échelle nationale, la BnF joue aussi le rôle d'agrégateur pour la bibliothèque numérique Europeana. À l'échelle internationale, les documents numérisés dans le cadre des projets de réunification ont vocation à rejoindre des ensembles plus larges constitués par nos partenaires étrangers. Ainsi, les manuscrits hébreux numérisés dans le cadre du partenariat avec la bibliothèque nationale d'Israël seront diffusés sur un site dédié, qui réunira 75 000 manuscrits hébreux dispersés dans une centaine d'institutions dans le monde (British Library, Bibliothèque du Congrès, Deutsche Nationalbibliothek, Bibliothèque Vaticane, ...).

Par ailleurs, Gallica alimente un cercle vertueux en favorisant la dissémination de ses ressources (notamment en valeur les réutilisations ainsi permises : interlocuteurs privilégiés, les Gallicanutes bénéficient par exemple d'une page dédiée dans Gallica, laquelle consacre également une rubrique à ses partenaires.

Le respect des identités

Des pages de présentation des partenaires permettent de valoriser les institutions qui diffusent tout ou partie de leurs collections numériques dans Gallica. Soucieuse de garantir à ses partenaires le respect de leur identité numérique, la BnF présente leurs collections dans Gallica avec des mentions de source individualisées.

Quelques chiffres

Entre 2009 et 2014, la numérisation partenariale a permis la numérisation de 17 380 000 pages. 176 partenaires nationaux ont été soutenus par la BnF, à hauteur de 6 109 000 €. Les différents programmes de numérisation atteignent en 2015 des volumes significatifs : plus de 12 millions de pages numérisées (6 millions pour les quatre programmes thématiques et également 6 millions pour les programmes d'intérêt régional et de complétude de corpus).

Au cours des dernières années, Gallica a fortement accentué sa dimension de bibliothèque numérique collective. La démarche fédérative, engagée dès le lancement de Gallica en 1997, permet de donner accès en 2015 aux ressources numériques de 305 partenaires, pour plus de 400 000 documents. Parmi ceux-ci, on distingue trois types de partenaires :

Les partenaires de Gallica par types de partenaires (toutes filières) 1997-2015

Partenaires des territoires (bibliothèques et archives des collectivités territoriales, structures régionales de coopération, sociétés savantes)	204
Partenaires de l'Enseignement supérieur et de la recherche	45
Autres partenaires (bibliothèques spécialisées essentiellement, avec des statuts variés et rattachées à des tutelles très diverses)	56

Total des partenaires de Gallica : 305

G. Principaux éléments techniques

L'alimentation de Gallica s'effectue grâce aux collections numérisées ou nées numériques de la BnF ainsi qu'aux collections des partenaires, dont l'intégration se fait selon trois modes d'entrée : référencement par interopérabilité OAI-PMH, intégration de documents originaux dans les marchés de numérisation et intégration par dépôt de fichiers numériques.

Référentiels de numérisation

La numérisation se fait en général en 400 DPI. Des résolutions supérieures peuvent être utilisées si le format du document original le nécessite : numérisation à 600 DPI pour les documents de format inférieur au A6, et jusqu'à 2800 DPI pour les documents de très petite taille comme les monnaies. Les images étaient jusqu'à présent produites en TIFF non compressé mais la production bascule progressivement vers le JPEG2000. Une modification du format des métadonnées est également actuellement en cours : passage du format RefNum au format METS.

L'OCR (reconnaissance optique de caractères) des documents textuels (imprimés et presse) est actuellement livré au format ALTO LoC, un format XML maintenu par la Bibliothèque du Congrès. Sur d'anciennes opérations, c'est une version BnF du schéma ALTO qui était utilisée. Les documents téléchargeables, dans Gallica, sous forme de livres électroniques le sont au format EPUB3. Format grand public, il présente l'avantage d'être interopérable et il a vocation à prendre la suite du format DAISY, utilisé jusqu'alors pour rendre accessible les documents numériques aux personnes en situation de handicap.

Afin de garantir l'affichage des documents sur le long terme, chaque document numérique de Gallica dispose d'un identifiant ARK qui permet de générer une url unique et pérenne, indépendamment des évolutions de la plateforme de Gallica.

L'ensemble des référentiels de numérisation de la BnF sont consultables sur son site internet.

Interface de Gallica : fonctionnalités de recherche et de lecture

Le moteur de recherche de Gallica est capable d'interroger des éléments classiques d'un catalogue, mais aussi une table des matières, un texte complet ocrisé... Les caractéristiques du nouveau visualiseur de documents développé en 2015-2016 sont les suivantes : interface unique, différents modes d'affichage (simple page, double page en vis-à-vis ou en recto verso, défilement vertical, vue d'ensemble mosaïque), barre d'outils latérale sur la gauche, nouveau mode d'affichage (texte + image) désormais disponible pour les documents dotés d'un mode texte, lecteur vidéo pour permettre la consultation des images animées, zoom plus fluide qui peut également être décliné en plein écran, navigation dans les tables des matières, navigation au sein des périodiques numérisés et possibilité de naviguer en 3D.

Outils de dissémination et d'appropriation des contenus

Gallica dispose d'outils de base comme les liens pérennes (mise en place du référencement pérenne des adresses url des documents en 2006 : lien ARK / permalien), le lecteur exportable, la possible implantation dans Facebook et Twitter, les partages sur les réseaux sociaux et des interfaces de programmation (API) qui permettent de réutiliser les données de Gallica sur d'autres sites, comme l'API International Image Interoperability Framework (IIIF), un protocole conçu par l'université de Stanford pour permettre la consultation d'images et de portions d'images sur des sites distants.

Afin de s'adapter à un maximum de publics et d'usages, Gallica respecte la norme RGAA (accessibilité) : la nouvelle version de Gallica offre ainsi la possibilité de feuilleter les documents au clavier. Il a d'autre part été

développé une version mobile et des applications pour téléphones portables et tablettes.

Les Gallicanutes peuvent se créer un espace personnel (indexation, annotation, pose de signets) et disposent de fonctionnalités d'export, de téléchargement et d'impression.

H. Quelques éléments financiers

Depuis 10 ans, la BnF opère de vastes chantiers de numérisation, grâce notamment à la mise en place en 2007 du soutien financier du Centre national du Livre.

Le financement par crédits issus du CNL entre 2007 et 2016 se chiffre à 62,3 M€. Sont financés à la fois la numérisation des imprimés, le stockage numérique dans l'entrepôt numérique de la BnF (SPAR), la diffusion via Gallica, le registre des livres indisponibles « Relire » et la numérisation afférente. Cela représente une moyenne annuelle de 6,23 M€, répartis pour 55% sur les marchés de numérisation et 45% sur l'informatique.

À cela la BnF ajoute des crédits sur fonds propres. En moyenne 2,15 M€ pour les marchés de numérisation et 0,68 M€ pour les investissements logiciels et de stockage informatique, soit au total 2,83 M€ par an. Ces chiffres n'incluent pas les ressources humaines.

Globalement, la numérisation au sens large a mobilisé 9 M€ par an depuis 10 ans, 61% affectés aux marchés de numérisation et 39% sur l'informatique; le CNL a assuré 69% des financements, et la BnF 31%, hors mécénats et partenariats qui fluctuent d'une année sur l'autre et hors ressources humaines.

En complément de ces éléments synthétiques, le tableau suivant permet d'appréhender les coûts de la numérisation à l'image ou à la page (référence 2016) :

Prix à l'image			
Type de document	Type de prestation	OCR	Tarif à l'image
Livres imprimés courants	Numérisation livres imprimés tous départements	sans	0,22 €
Livres imprimés courants	Numérisation livres imprimés tous départements	avec	0,37 €
Livres imprimés hors d'usage	Numérisation livres imprimés tous départements	avec	0,46 €
Livres précieux (ie Exceptionnels)	Numérisation livres Réserve et Arsenal	sans	0,59 €
Presse ancienne	Numérisation de presse papier tous départements	avec	0,66 €
Microforme	Numérisation de microfiche ou microfilms, tous départements	sans	0,06 €
Microforme	Numérisation de microfiche ou microfilms, tous départements	avec	0,32 €
Collection spécialisée image simple	Collections des départements spécialisés	sans	0,84 €
Collection spécialisée image complexe	Collections des départements spécialisés	sans	0,84 €

I. Bibliographie

- ▶ *La numérisation en sciences religieuses : un programme en concertation*
Corinne Bouquin,
Bulletin des Bibliothèques de France 1, janvier 2010
- ▶ *La numérisation à la BnF et les investissements d'avenir*
Denis Bruckmann et Nathalie Thouny,
Bulletin des Bibliothèques de France, 4, juillet 2012
- ▶ *Une bibliothèque numérique sur les réseaux sociaux : l'exemple de Gallica*
Collectif equipe@GallicaBnF,
Bulletin des Bibliothèque de France 5, septembre 2012
- ▶ *Gallica (1997-2016) : de la « bibliothèque de l'honnête homme » à celle du gallicanaute*
Aline Girard et Sophie Bertrand,
Bulletin des Bibliothèques de France 9, juillet 2016

François-Mitterrand

Quai François Mauriac, Paris 13^e _ 33 | 0 | 1 53 79 53 79

Richelieu

5, rue Vivienne et 2, rue Louvois, Paris 2^e _ 33 | 0 | 1 53 79 53 79

Bibliothèque de l'Arsenal

1, rue de Sully, Paris 4^e _ 33 | 0 | 1 53 79 39 39

Bibliothèque-musée de l'Opéra

Place de l'Opéra, Paris 9^e _ 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

bnf.fr