Le BnF DataLab, un service aux chercheurs en humanités numériques

Marie Carlin, Arnaud Laborderie

Le BnF DataLab, un service aux chercheurs en humanités numériques

The BnF DataLab, a Service for Research in Digital Humanities

Marie Carlin,
Département de l'Orientation et de la recherche bibliographique, Bibliothèque nationale de France
marie.carlin@bnf.fr

Arnaud Laborderie,
Département de la Coopération, Bibliothèque nationale de France
arnaud.laborderie@bnf.fr

Résumé

Depuis une dizaine d'années, l'augmentation massive des collections numériques de la BnF (Gallica, archives de l'Internet, métadonnées, etc.) ouvre de nouvelles pistes de recherche, faisant émerger des problématiques d'exploitation des données patrimoniales par les chercheurs. Pour répondre à leurs usages et besoins, la Bibliothèque nationale de France met en place un nouveau service, le BnF DataLab, dont l'objectif est d'accompagner les équipes de recherche pour constituer et traiter des corpus numériques. Ce lieu, conçu pour le travail individuel et collaboratif, ouvrira ses portes à l'automne 2021.

Mots-clés

Humanités numériques, bibliothèque numérique, numérisation, collections numériques, archives de l'Internet

Abstract

Over the past ten years, the huge increase of digital collections at the National Library of France (BnF) — Gallica, Internet archives, metadata, etc. — opens up new research perspectives, raising issues to utilize heritage datas by researchers. To meet these uses and needs, the BnF is setting up a new service: the BnF DataLab, aimed to support research teams to constitute and to process digital corpuses. This place, designed for individual and collaborative work, will open in fall 2021.

Key-words

Digital humanities, digital library, digitalization, digital collections, French Internet archives

—

Version actualisée le 24/08/21

Notices biographiques

Marie Carlin est diplômée du master Technologies numériques appliquées à l'histoire de l'École nationale des chartes et conservatrice des bibliothèques. Elle exerce les fonctions de coordinatrice du BnF DataLab au sein du département de l'Orientation et de la recherche bibliographique de la Bibliothèque nationale de France.

Arnaud Laborderie, docteur en Sciences de l'Information et de la Communication, est chef de projet au service de la Coopération numérique et de Gallica à la Bibliothèque nationale de France, en charge de l'exploitation des données pour la recherche. Membre du laboratoire Paragraphe de l'Université Paris-VIII, il est professeur associé au département Humanités numériques.

Introduction

Depuis une vingtaine d'années, l'avènement des humanités numériques bouleverse la recherche en sciences humaines et sociales (SHS) et interroge le rôle des bibliothèques. D'un côté, les SHS se trouvent soumises, comme l'a explicité Michel Wieworka (2013), a un « impératif numérique » pour assurer leur suivie, ouvrant une nouvelle ère, celle des pratiques outillées par les nouvelles technologies, ce qui modifie profondément les méthodes de travail. La numérisation des sources et le traitement informatisé des corpus produisent un changement global de la manière de faire de la recherche et d'en communiquer les résultats (Dacos et Mounier 2014).

De l'autre côté, les bibliothèques, qui sont au cœur de l'activité de recherche, se trouvent interrogées sur l'évolution de leurs services avec le développement d'Internet et l'omniprésence du Web dans les pratiques documentaires. Si les humanités numériques, par leur approche transverse et leur propension à décloisonner les disciplines et les services, bousculent la hiérarchie traditionnelle des bibliothèques, elles ouvrent aussi de nouvelles opportunités (Caraco 2012, Barret 2014).

Pour la Bibliothèque nationale de France (BnF), il s'agit d'accompagner les nouvelles pratiques de recherche qui impliquent ses collections numériques et ses données. Celles-ci représentent une masse considérable de documents (plus de six pétaoctets), d'une grande diversité, tant par leur forme que par leur contenu : documents numérisés dans Gallica[1] et Gallica intramuros, métadonnées bibliographiques, jeux vidéo, CD et DVD, ou documents nés numériques comme les archives du Web, les logs de connexion, etc. L'accroissement de ces collections numériques va de pair avec la mise au point de nouveaux outils (de consultation, d'extraction et de fouille de données par exemple) et services, compris comme le recours à des experts des bibliothèques. La multiplication des demandes de corpus massifs, multimodaux, hétérogènes, a favorisé une réflexion plus globale sur les services que la BnF pouvait offrir. Cette réflexion s'est construite d'abord sous la forme d'un projet de recherche interne – le projet Corpus – puis par la définition de nouveaux services dédiés et l'aménagement d'un espace d'accueil : le BnF DataLab, en partenariat avec le CNRS et Huma-Num[2]. Ce lieu, conçu pour le travail individuel et collaboratif, ouvrira ses portes en octobre 2021. Comment la BnF envisage-t-elle les services qu'elle va proposer aux chercheurs dans son DataLab ? Quelles en sont les limites techniques, économiques ou juridiques ? Quels enjeux R&D pour l'exploration des corpus numériques et pour Gallica ?

---

[1] Bibliothèque numérique de la BnF et de ses quelque 400 partenaires, Gallica comprend plus de 8 millions de documents : https://gallica.bnf.fr/. Gallica intramuros propose près de 1 million de documents supplémentaires sous droits, consultables uniquement dans les emprises de la BnF.

[2] Huma-Num est une infrastructure de recherche (TGIR) dédiée aux lettres, sciences humaines et sociales et aux humanités numériques, mise en œuvre par le Ministère de l'Enseignement supérieur et de la Recherche et portée par le Centre national de la recherche scientifique (CNRS), Aix-Marseille université et le Campus Condorcet. Information : https://www.huma-num.fr

La question d'un espace physique pour les humanités numériques a été posée à de nombreuses reprises par la communauté scientifique (Catzaras et Holland 2012, Vensson 2016, ou Fickers et Van der Heijden 2020). Les chercheurs ont discuté l'accueil physique des projets, l'inscription du corps dans ces espaces équipés de technologies, les compétences et les ressources à mobiliser, ainsi que les services à mettre à disposition (notamment en France lors de THATCamp à Paris en 2012, Saint-Malo en 2013 et Lyon en 2014). Interrogeant le rôle des bibliothèques, ils pointaient la nécessaire évolution des espaces physiques et virtuels de diffusion des connaissances, mais aussi du rôle des bibliothécaires, appelés à devenir des « partenaires » pour leur connaissance des collections et leur expertise en matière de données.

Olivier Le Deuff soulignait que les équipes de recherche, confrontées à des problématiques bibliothéconomiques et archivistiques, ont besoin des compétences des professionnels de l'information que sont les bibliothécaires et conservateurs pour produire des données qui soient interopérables et puissent être conservées (Le Deuff 2013). Il s'agit là d'un changement de paradigme, avec des bibliothèques qui ne sont plus seulement tournées vers la conservation et la communication des documents, mais également vers la médiation numérique et la production de savoir, notamment par la création de métadonnées, l'éditorialisation des collections et la constitution de corpus numériques, ainsi que par la formation aux outils de traitement et d'analyse de données. Pour Le Deuff (2014), l'enjeu est d'appréhender la bibliothèque comme « *nouveau milieu de savoir* […] qui permet autant l'accès à l'information que la possibilité de produire et de réaliser ». Miriam Posner (2013), qui a observé une telle évolution aux États-Unis, souligne qu'introduire les humanités numériques dans la bibliothèque est un défi qui nécessite à la fois de nouvelles compétences métiers (ingénieurs, experts, etc.) et de nouvelles infrastructures en termes d'espaces et d'équipements : serveurs, machines virtuelles, logiciels, déploiement d'API[3], etc. Des *DH center*[4] et des « datalab » se développent au sein de grandes bibliothèques universitaires et nationales, parmi lesquelles citons notamment l'Alan Turing Institute et le British Library Labs, le Loc Labs à la Bibliothèque du Congrès (Gallinger et Chudnov 2016), le KB Lab à la Bibliothèque nationale des Pays-Bas ou le Datalab.kb.se à la Bibliothèque nationale de Suède (Snickars 2018).

En France, par leur proximité avec les chercheurs, les bibliothèques universitaires (BU) ont pu prendre la mesure de la mutation à l'œuvre à l'exemple de l'université Bordeaux-Montaigne, dont le service commun de documentation (SCD) se présente comme « fournisseur de

[3] Une API (*Application Programming Interface*) est une interface de programmation permettant à des applications de communiquer entre elles et de s'échanger des services ou des données.

[4] Le *Digital Humanities center* est un espace hybride et mixte, est à la fois distinct et à mi-chemin entre la bibliothèque universitaire et le laboratoire de recherche. Généralement intégré à l'université, le *DH center* accueille les chercheurs et les projets de recherche en humanités numériques tout en formant les étudiants (masterants et doctorants) à ses nouvelles pratiques, méthodes et outils.

services » sur des projets ayant trait aux humanités numériques (Machefert 2014). Avec son DataLab[5] dont l'ouverture est prévue en 2021, la Bibliothèque nationale et universitaire de Strasbourg (Bnu) s'est également engagée dans un nouveau service d'accompagnement de la recherche dans les humanités numériques. À l'université de Lille, c'est une BU d'un nouveau genre que propose LILLIAD[6], « forum des savoirs » avec des espaces dédiés aux apprentissages. En intégrant en leur sein des *learning centers*[7], les BU renouvellent la relation entre bibliothèque et formation par l'articulation entre « l'enseignement (*teaching*), l'acquisition de connaissances (*learning*), la documentation et la formation aux technologies (*training*) » (Jouguelet 2009). En s'adossant aux humanités numériques, il s'agit d'aller au-delà des seuls acquis informationnels pour développer des compétences dans la science des données. C'est ce positionnement au croisement de la bibliothèque, du *learning center* et du *DH center* que le BnF DataLab veut adopter grâce au partenariat avec Huma-Num et le concours d'équipes de recherche en résidence.

Avant de présenter les nouveaux services aux chercheurs que la BnF souhaite proposer, il importe d'ancrer le BnF DataLab dans des initiatives et des projets qui l'ont précédé en matière d'humanités numériques et dans la continuité desquels celui-ci s'inscrit. Nous verrons que cet espace singulier, à la fois physique et virtuel, répond à une attente de la communauté scientifique et renouvelle les relations entre chercheurs et bibliothécaires. S'il apparaît aujourd'hui comme une nouvelle offre de services, le BnF DataLab est l'aboutissement d'un long processus et d'une logique de numérisation des collections engagée il y a une trentaine d'années, dans le contexte d'une institution qui a toujours pensé coopération.

Les humanités numériques à la BnF : expérimentations et coopération en R&D

La coopération figure parmi les valeurs et missions fondamentales de la BnF, avec la numérisation comme axe central de sa politique de coopération en France et à l'étranger. Dès la fin des années 1990, la BnF engageait ses premiers programmes coopératifs de numérisation de fonds (publications des sociétés savantes et des académies nationales) avec des bibliothèques en région (les « pôles associés »).

---

[5] Un projet Collex (2018) en partenariat avec la BnF, l'École des chartes, Persée, l'Unistra, l'Urfist de Strasbourg et le Lorraine Fab Living Lab.

[6] LILLIAD (Learning Center Innovation) est le forum des savoirs à caractère scientifique de l'université de Lille, axé sur l'innovation et inauguré en 2016 au sein du campus de la Cité scientifique. Information : https://lilliad.univ-lille.fr/

[7] Né aux États-Unis dans les années 90, le concept de *learning center* (traduit en français par « forum des savoirs ») désigne une partie ou l'ensemble d'une bibliothèque universitaire où « la dimension pédagogique est essentielle » (Jouguelet, 2009) et qui forme les étudiants en particulier aux compétences informationnelles et technologiques par la maîtrise de bonnes pratiques et d'outils.

En 2007, en réponse au projet Google Books[8] (Jeanneney 2005), une volonté politique et une dynamique européenne conduisaient à un changement d'échelle avec d'importants financements permettant une numérisation de masse des collections, passant progressivement de 5 000 à 100 000 documents numérisés par an. La BnF s'engageait alors dans une stratégie de coopération numérique visant à « favoriser la numérisation du patrimoine écrit des bibliothèques françaises et à créer de manière collaborative de vastes ressources patrimoniales numérisées destinées à enrichir Gallica, mais aussi les autres bibliothèques numériques » (Bertrand et Girard 2016). Cette stratégie repose sur la mise en œuvre de vastes programmes de numérisation concertée avec des partenaires, organisés autour de projets structurants (disciplinaires ou d'intérêt régional) et de complétude documentaire des collections. Pour constituer les corpus, la BnF a mis en place des modalités de coopération scientifique avec des institutions partenaires et des chercheurs sollicités pour formuler leurs besoins et proposer des listes de documents à numériser. Il en est ainsi, par exemple, du programme de numérisation concertée en sciences juridiques, réalisé en concertation avec la bibliothèque Cujas, qui s'est appuyé sur des listes raisonnées d'ouvrages de droit établies par les professeurs Yann Kerbrat, Laurent Pfister et Franck Roumy. En identifiant un besoin quant aux sources du droit, ces chercheurs ont donné le périmètre du champ de numérisation et la structuration du corpus[9].
Le modèle de coopération numérique repose sur une mutualisation des coûts avec les partenaires et sur un impératif, celui de l'interopérabilité, ce qui implique des choix techniques partagés (format Dublin Core, protocole OAI-PMH, etc.). Plus de 120 bibliothèques numériques sont ainsi interopérables avec Gallica.

Aujourd'hui, cette politique de numérisation partenariale se poursuit dans un contexte de baisse généralisée des moyens. Alors qu'est atteinte la masse critique des 8 millions de document accessibles dans Gallica, de nouveaux enjeux apparaissent : proposer au public une médiation des collections numériques, mais aussi permettre aux chercheurs d'exploiter cette masse documentaire considérable. C'est une nouvelle étape qui s'ouvre et un défi auquel doit répondre le BnF DataLab : ceux de l'exploitation massive des collections numérisées et de la création d'outils dédiés.

Travailler sur les collections numériques

L'exploitation des collections numériques à des fins de recherche est expérimentée de longue date par la BnF dans le cadre de partenariats scientifiques. Dès les années 1990, l'équipe de

[8] Lancé en octobre 2004, la bibliothèque virtuelle Google Books comptait déjà plus de sept millions de livres en novembre 2008. (Robert Darnton, « Google and the Future of Books », *The New York Review of Books*, vol. 56, n° 2, février 2009).
[9] Les Essentiels du droit dans Gallica : https://gallica.bnf.fr/html/und/droit-economie/essentiels-du-droit

préfiguration de la BnF travaillait avec des chercheurs émérites (les « grands lecteurs ») pour modéliser leur pratique de lecture savante à travers un projet précurseur de « poste de lecture assistée par ordinateur » (PLAO). À cette époque, l'établissement collaborait avec l'Institut National de la Langue Française (CNRS), participant à l'informatisation du TLF[10] et bénéficiant, pour le lancement de Gallica en 1997, de 250 livres numérisés en mode texte issus de la base Frantext, lesquels venaient en complément des 2 500 livres alors numérisés en mode image. On peut voir dans cette collaboration une des premières initiatives portant sur les questions de rétroconversion de catalogues, d'indexation et d'exploitation de corpus par le numérique.

À partir des années 2010, la numérisation en masse des collections patrimoniales permet l'émergence de projets de recherche particulièrement représentatifs de l'enjeu des humanités numériques pour les bibliothèques et pour la BnF en particulier. C'est en 2011-2012 que s'ouvre véritablement pour les chercheurs une nouvelle ère en matière d'exploitation des collections numérisées. Alors que l'Agence nationale de la recherche (ANR) finance la constitution de corpus de sources sous forme numérique, Gallica franchit la barre du million de document. Le mouvement des humanités numériques commence à s'organiser en France[11] au moment même où le paysage de la recherche change avec la création des Labex et des Equipex[12], dans lesquels la BnF s'implique en apportant son expertise ainsi qu'un accès privilégié à ses collections numérisées (Bermès 2020).
Le labex OBVIL (2012-2019)[13] ouvre la voie à une collaboration plus étroite entre la BnF et chercheurs en humanités numériques, en particulier autour de la fouille de données et du déploiement d'outils d'exploitation sur des corpus massifs de sources primaires et de textes critiques. La constitution d'une collection numérique en TEI de quelque 130 000 monographies issues de Gallica [14] conduit à un changement de paradigme, en passant de la « lecture rapprochée » à la « lecture distante » (Moretti 2013) de documents appréhendés en masse à l'aide d'outils algorithmiques et statistiques. Ce programme de recherche a conduit à une prise de conscience globale à l'échelle de l'établissement. L'impact très lourd en matière d'organisation, de définition de processus de travail, les contraintes techniques liées à la

[10] Trésor de la langue française, projet conduit par l'Institut National de la Langue Française (INALF – CNRS), devenu aujourd'hui laboratoire ATILF Analyse et Traitement Informatique de la Langue Française (laboratoire associé au CNRS et l'Université de Lorraine). En savoir plus : https://www.atilf.fr/ressources/tlfi/
[11] Le premier THATcamp européen, qui s'est déroulé à Paris les 18 et 19 mai 2010, a débouché sur un *Manifeste des humanités numérique*. En ligne : https://tcp.hypotheses.org/318
[12] Les Labex (laboratoires d'excellence) et des Equipex (équipements d'excellence) sont financés par l'ANR (agence nationale de la recherche) dans le cadre des plans d'investissements d'avenir (PIA).
[13] Porté par Sorbonne-Université, le labex OBVIL (Observatoire de la vie littéraire) est un des importants programmes d'humanités numériques en littérature. Il réunit près de 400 chercheurs, répartis dans 24 projets de recherche. Fort d'une équipe de six ingénieurs, OBVIL conçoit des outils d'édition, de fouille de texte, d'alignements et de visualisation de données. En ligne : http://obvil.sorbonne-universite.site/
[14] En savoir plus : http://api.bnf.fr/mise-disposition-de-la-tres-grande-bibliotheque-du-labex-obvil

livraison des données (nécessaire qualité de la numérisation, océrisation[15]) ont participé à la réflexion sur la mise en place d'un service coordonné, capable de répondre dans des délais raisonnables à ce type de demande.

Avec Biblissima (2012-)[16], la BnF s'implique dans un programme à dimension internationale. Ce projet visait à assurer une interopérabilité entre des corpus de manuscrits et d'incunables issus de différentes bibliothèques, afin d'assurer la visualisation des documents dans une même interface, quelque que soient les entrepôts où sont conservées les données. En intégrant le consortium IIIF[17], la BnF s'ouvrait à de l'ingénierie technique développée par une communauté associant ingénieurs et conservateurs autour d'un enjeu de recherche et d'exploitation de collections numériques spécifiques, indissociable de l'émergence d'outils développés en coopération. Cette collaboration a permis d'intégrer un export IIIF des documents dans Gallica et a permis la constitution d'une bibliothèque numérique de manuscrits médiévaux en marque blanche avec la British Library[18].

Autre exemple structurant en matière de R&D, l'implication de la BnF dans le projet Europeana Newspapers (2012-2015)[19], dont l'ambition est d'améliorer l'accès aux collections numérisées de la presse européenne par des fonctionnalités de recherche augmentée et par l'enrichissement sémantique des données. Les processus liés à la numérisation sont optimisés pour reconnaitre les articles de presse à l'unité grâce à l'expérimentation de la technologie OLR[20] (cf. Moreux 2016). Au-delà de la recherche plein-texte (grâce à l'OCR), il s'agit de segmenter les articles pour décrire leurs contenus par classes (typologie d'articles, illustrations, légendes, publicités, petites annonces, tableaux, etc.) et pouvoir y accéder directement. Avec la reconnaissance des entités nommées (REN)[21], on peut identifier et extraire des expressions, noms ou lieux. La catégorisation de ces objets dans des classes permet ainsi d'améliorer les fonctionnalités de consultation et de présentation d'Europeana et de Gallica.

---

[15] L'OCR (*Optical Character Recognition)* est une technologie de reconnaissances de textes imprimés à partir d'images numérisées. Disponible dans Gallica depuis 2005, l'OCR constitue aujourd'hui une chaîne d'entrée interne à la BnF, engagé notamment dans des programmes de rétroconversion.

[16] Doté d'un budget de 7 millions d'euros dans le cadre du programme Équipex (Équipements d'excellence), Biblissima (*Bibliotheca Bibliothecarum Novissima)* est un projet de « bibliothèque des bibliothèques du XXI^{e} siècle ». En savoir plus : https://projet.biblissima.fr/

[17] IIIF (*International Image Interoperability Framework*) désigne à la fois une communauté et un ensemble de spécifications techniques dont l'objectif est de définir un cadre d'interopérabilité pour la diffusion d'images haute résolution sur le Web.

[18] « France et Angleterre, 700-1200 : manuscrits médiévaux de la BnF et de la British Library », en ligne : https://manuscrits-france-angleterre.org/polonsky/fr/content/accueil-fr

[19] Réunissant 18 partenaires, le projet Europeana Newspapers vise au traitement et l'agrégation des journaux libres de droits issus des grands titres de la presse européenne. En savoir plus : http://www.europeana-newspapers.eu/

[20] La technologie OLR *(Optical Layout Recognition*) permet une reconnaissance de la mise en page des documents, utilisée notamment pour la presse ancienne numérisée (cf. Moreux 2016).

[21] Sous-tâche de l'activité d'extraction d'information dans les corpus documentaires, la reconnaissance des entités nommées (REN) consiste à rechercher et identifier un certain nombre d'objets textuels (mots, expressions, noms, lieux, etc.) présents dans les textes.

Ces projets témoignent des premières implications de la BnF dans le champ des humanités numériques et d'un positionnement qui est celui de l'expérimentation et de la coopération R&D avec des chercheurs et des institutions dans le cadre de partenariats scientifiques.

Depuis leur création en 2017, la BnF est également partenaire de projets de recherche Collex[22] portés par des bibliothèques sous tutelle du Ministère de l'Enseignement et de la Recherche. Ces projets poursuivent un double objectif de numérisation concertée de « collections d'excellence » permettant de valoriser de grands gisements documentaires patrimoniaux et scientifiques, et de fourniture de nouveaux services aux chercheurs désormais au cœur des projets de collections. En s'impliquant dans ces programmes de numérisation, la BnF est garante d'un écosystème ouvert permettant l'interopérabilité et l'exploitation optimisée des données par des communautés plurielles.
Enfin, la BnF s'appuie sur des programmes de recherche internes, inscrits dans des plans quadriennaux[23] qui intègrent les notions d'humanités numériques et d'exploitation de données. C'est dans ce cadre qu'est né le projet Corpus, programme de préfiguration des services aux chercheurs en matière de données qui, poursuivant cette stratégie de recherche et d'expérimentation, devait conduire au BnF DataLab.

L'augmentation massive des collections numériques ouvre ainsi de nouveaux champs de recherche et des perspectives inédites d'exploitation de corpus. Cette mise à disposition stimule de nouveaux usages, qui sont autant de défis en termes de mise en place de services dédiés. Une forme de pression du monde académique pousse la BnF à réfléchir globalement son service aux chercheurs. Trois principaux domaines d'innovation sont alors définis pour rendre ces collections accessibles (Pardé et Jacquot 2016) : l'amélioration de l'accès aux données numériques (correction, enrichissement, interopérabilité), la mise à disposition d'outils pour l'exploitation de ces données massives (fouille, analyse, indexation), l'étude des usages numériques (initiés dès 2013 avec le Bibli-Lab[24], en partenariat avec Télécom ParisTech). Dans son contrat d'objectifs et de performance 2017-2021 signé avec le Ministère de la Culture, l'institution se fixe alors pour objectif d'« offrir aux chercheurs, dans les

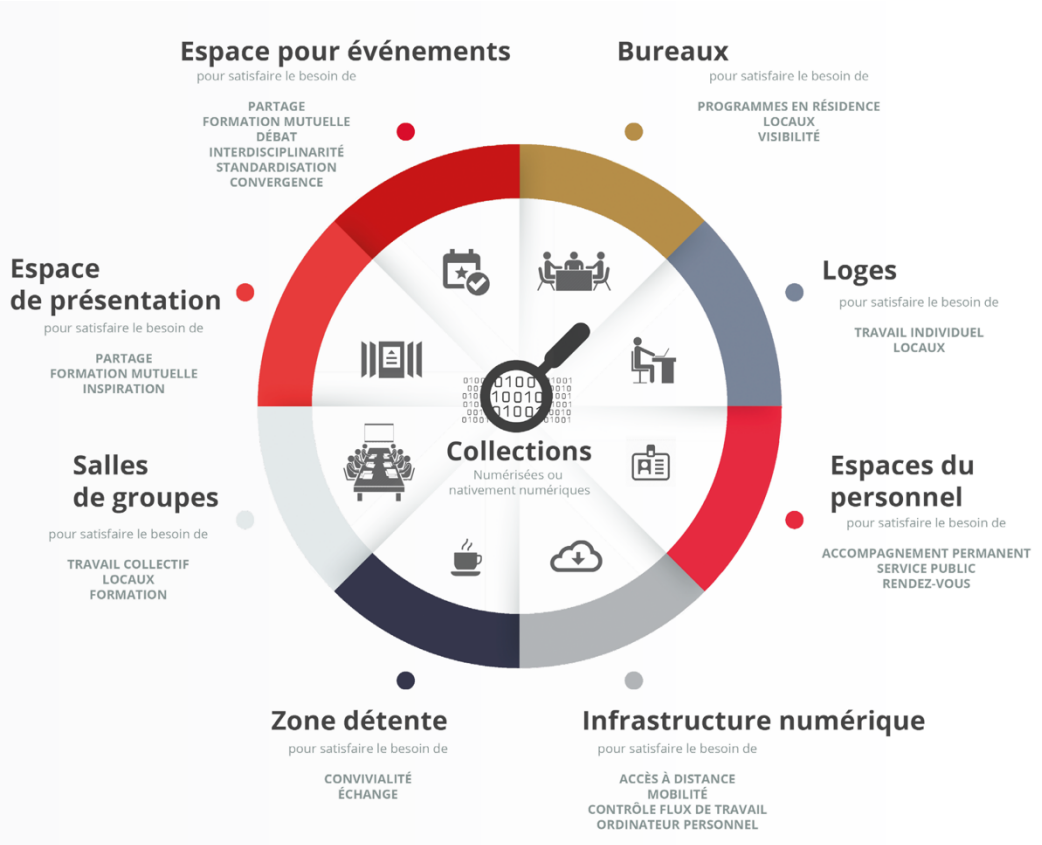
[22] En savoir plus : https://www.collexpersee.eu/les-projets/
[23] En savoir plus : https://www.bnf.fr/fr/plan-quadriennal-de-la-recherche
[24] Quatre programmes de recherche structurants ont été conduits entre 2013 et 2018 dans le cadre du Bibli-Lab, « Laboratoire d'étude des usages du patrimoine numérique des bibliothèques », créé par la BnF et l'école Télécom ParisTech : « Observer et évaluer les usages de Gallica : réflexion épistémologique et stratégique » (2013-2014) ; « Le devenir en ligne du patrimoine numérisé : l'exemple de la Grande Guerre » (2013-2016), « Mettre en ligne le patrimoine : transformation des usages, évolution des savoirs » (2016) ; « Analyser les traces d'usage de Gallica » (2016-2017). Rapports de recherche disponibles ici : https://actions-recherche.bnf.fr/BnF/anirw3.nsf/IX01/A2017000006_bibli-lab-laboratoire-d-etude-des-usages-du-patrimoine-numerique-des-bibliotheques

emprises de la Bibliothèque, des outils de fouille et d'exploration de textes et de données sur des corpus numériques de la BnF »[25].

Du projet Corpus au BnF DataLab

Pour réaliser cet objectif, la BnF a procédé de manière expérimentale, itérative et transversale, à travers un programme de recherche inscrit au plan quadriennal dénommé *Corpus*[26], conduit par Emmanuelle Bermès entre 2016 et 2019 (cf. Bermès et Moiraghi 2020). Découpé en deux phases, le projet Corpus entame une phase exploratoire autour de projets de recherche existants[27] afin d'expérimenter un service de fourniture de données dans trois grands ensembles documentaires : les données de Gallica, les archives du Web, les métadonnées du catalogue général. L'expérimentation s'appuie sur des ateliers collaboratifs menés avec des chercheurs dans une logique de co-construction, afin de mieux cerner les attentes des équipes de recherche.

« Figure 1 : Préfiguration du Laboratoire d'étude et d'analyse de corpus numériques (Moiraghi 2018) »



[25] Contrat d'objectifs et de performance 2017-2021 : https://www.bnf.fr/fr/mediatheque/contrat-dobjectifs-et-de-performance-2017-2021

[26] Programme inscrit au plan quadriennal de la recherche de la BnF (2016-2019), en savoir plus : https://c.bnf.fr/J5q

[27] ANR Web90, ANR Giranium, ANR Foucault Fiches de lecture, etc. Cf. https://bnf.hypotheses.org/

En parallèle à ces explorations, une enquête de besoins réalisée par Eleonora Moiraghi permet d'esquisser une offre de service en se basant sur des entretiens et en définissant des archétypes utilisateurs (*persona*). Cette étude fait le constat que la fouille de données est appelée à se développer dans toutes les disciplines. Elle acte la nécessité de prendre en compte de nouvelles temporalités (traitements de données durant plusieurs heures) et de concevoir des zones dédiées non seulement au travail individuel mais aussi au travail de groupe, à la formation, au partage d'expérience et à la convivialité (Moiraghi 2018).
De ces enquêtes et ateliers émerge le besoin d'un lieu identifié pour accueillir des chercheurs travaillant sur les corpus numériques, afin de permettre non seulement l'échange et la rencontre entre chercheurs et experts BnF, mais aussi la consultation des corpus sous droits et la mise en place d'une infrastructure informatique dédiée (Fig. 1). La dernière année du programme, bilan des expérimentations menées, permet de proposer une première typologie des services et de mettre en place un schéma d'organisation au sein de la BnF.

Le BnF DataLab, nouveau service pour la recherche

Le BnF DataLab, instance applicative des expérimentations conduites lors du programme Corpus, doit permettre l'accueil et l'accompagnement des chercheurs qui souhaitent travailler sur les collections numériques de la BnF : accompagnement dans l'accès aux collections, constitution de corpus, mais aussi développement des usages et des pratiques de la recherche autour des collections numériques avec le concours des experts de la BnF (Bermès 2019). Trois éléments caractérisent le BnF DataLab, qui est à la fois :

- un lieu de travail, d'échange, de résidence pour les chercheurs
- une offre de services, sur place et à distance, pour accompagner les usagers
- un laboratoire favorisant l'expérimentation et la R&D grâce à la mise en place de partenariats scientifiques

Un lieu de travail, d'échange, de résidence pour les chercheurs

La salle X, sise en Rez-de-jardin du site François-Mitterrand de la BnF, a été choisie pour accueillir les espaces du futur DataLab. Gérée par le Département de la recherche et de l'orientation bibliographique (ORB), la salle X est l'héritière de la salle des catalogues de l'ancienne Bibliothèque nationale à Richelieu, lieu pluridisciplinaire où sont rassemblés les outils nécessaires au repérage dans les collections de la Bibliothèque et plus largement à toutes recherches scientifiques (Éloi et *al*. 2019). Sa vocation à l'interdisciplinarité et l'appréhension nécessairement globale et toujours plus complexe des collections, autant que les services autour de la recherche bibliographique mis en place de longue date par les équipes

de l'ORB (chat, rendez-vous en ligne, préparer sa thèse à la BnF, Sindbad) en font le lieu prédisposé à héberger le BnF DataLab.
La configuration de la salle X, avec mezzanine, permet d'aménager deux niveaux distincts mais complémentaires dédiés à l'étude et à l'analyse de corpus numériques. En salle basse, des bureaux et salles de travail collectif de 4 à 8 places ainsi qu'une salle de formation de 12 places permettent des usages collectifs. Dans la mezzanine, des box individuels équipés permettent le travail individuel tandis qu'un espace central ouvert d'une trentaine de mètres carrés permet d'accueillir des manifestations scientifiques telles que des démonstrations, des présentations de travaux de recherche ou encore des ateliers. Une telle configuration rompt avec les traditionnelles salles de lecture silencieuses de la Bibliothèque, obligeant à imaginer des usages mixtes, en bonne cohabitation avec les autres usagers de la salle, qui reste une salle de lecture. Car le BnF DataLab est aussi porteur d'une ambition qui dépasse les services proposés : celle d'un lieu de sociabilité scientifique, « un lieu interdisciplinaire, pluridisciplinaire, voire hybride : un territoire favorisant la rencontre de multiples acteurs de la recherche et de différentes expertises (approfondie des collections, des domaines de recherche, mais aussi des aspects juridiques et informatiques). » (Éloi et *al*. 2019).
Ce lieu, espace physique d'échange et de travail, est doublé d'un espace virtuel pluriel : une infrastructure numérique accessible sur place et, à distance, un site Web compagnon (http://api.bnf.fr), documentant l'usage des API et publiant des jeux de données sous licence Etalab, ainsi que le site Gallica studio (http://gallicastudio.bnf.fr/), dédié à l'expérimentation et valorisant les projets de R&D.

Une offre de services sur place et à distance

C'est donc à partir de cas d'usage expérimentés dans le cadre du programme Corpus que les services du BnF DataLab ont été définis. Reprenant le cycle de vie des projets de recherche, ceux-ci se répartissent autour de trois axes :

- Accueillir les chercheurs : accueil et orientation des usagers, définition d'un parcours
- Identifier les collections : accès aux données, aide à la constitution des corpus
- Travailler les corpus : mise à disposition d'un environnement de travail, formations, suivi de projet avec des experts BnF

Des services relatifs aux collections, aux données et aux outils
Sur place, le DataLab propose un ensemble de services tournés vers l'accueil, le conseil et l'accompagnement des chercheurs dans leurs projets de recherche sur les collections et ressources numériques de la BnF. Des parcours ont été définis à partir de cas d'usage et d'archétypes utilisateurs (*personas*) pour déterminer les services, mais aussi les formations à suggérer à chaque étape du projet de recherche. Ainsi l'offre de services a-t-elle été pensée pour répondre à un parcours usager, dont la première étape consiste à circonscrire et établir

un corpus pertinent : aide à la recherche documentaire et à la constitution de corpus, rencontre avec des experts des collections et des métadonnées, formation aux outils d'extraction des ressources numériques (API), demande de numérisation des pièces manquantes, suivi de projet… Ces services doivent permettre aux chercheurs de savoir ce qui existe et dans quel format, ce qui est disponible et comment y accéder, avec pour objectif de former les équipes de recherche aux collections et aux outils pour permettre une relative autonomie. Il importe en effet d'accompagner les chercheurs dans leur compréhension de la structure des catalogues et des données (formats bibliographiques), de les aider à identifier les outils pertinents et à construire les requêtes en fonction de leurs besoins.
Le deuxième ensemble de services doit permettre d'exploiter le corpus ainsi constitué : se former aux outils de fouille de texte et d'image (par ex. GallicaPix[28]), bénéficier d'une infrastructure dédiée et de l'assistance d'experts (collections, formats, métadonnées, etc.), commander une prestation spécifique, mais aussi échanger avec des acteurs internes et externes. Des formations dans les domaines qui intéressent les collections de la BnF et les outils qu'elle développe seront proposées aux chercheurs, ainsi que des manifestations scientifiques (séminaires, ateliers, démonstrations) permettant aux usagers d'échanger, entre pairs ou avec des experts, de présenter leurs travaux ou de se familiariser avec des outils. Les espaces collectifs du BnF DataLab permettront ainsi de valoriser le travail des chercheurs accueillis et d'être un lieu d'échanges et de formations entre pairs.

Les besoins des chercheurs concernent principalement la fourniture de données des collections patrimoniales : images, textes issus de l'OCR, métadonnées. À distance, ils pourront utiliser des outils de requêtage des données (accessibles en ligne via les API), et bénéficier sur place d'un accompagnement à la prise en main de ces outils, mais aussi de services complémentaires comme l'extraction en masse de données ou un service de numérisation à la demande pour les équipes de recherche souhaitant travailler sur des corpus de documents non disponibles dans Gallica.
Une infrastructure numérique constituant un nouvel environnement de travail (machine virtuelle, extension sur le *cloud*) sera mis à la disposition des chercheurs pour le traitement de corpus (BnF ou externes), notamment à des fins de fouille de données, en respectant le cadre juridique. Certains traitements particuliers sur des types de corpus identifiés pourront être réalisés grâce à une boîte à outils logiciels (HTR[29], segmentation, indexation, extraction, etc.). L'infrastructure informatique permettra également de travailler sur les archives de l'Internet.

[28] Le démonstrateur GallicaPix est un outil de recherche iconographique dans les collections d'imprimés numérisés (livre, revue, presse) qui préfigure la fouille d'images dans Gallica.
[29] L'HTR (*Handwritten Text Recognition*) est une technologie de reconnaissance de l'écriture manuscrite à partir d'images numérisées.

Un service expérimental sur les archives de l'Internet

Le Web est une ressource documentaire particulière dont l'archivage est réalisé par la BnF dans le cadre du dépôt légal de l'Internet[30]. Depuis 2006, la BnF a pour mission la collecte, la conservation et la mise à disposition des archives du Web français. Ces archives, dont les plus anciennes remontent à 1996, s'accroissent chaque année et constituent une ressource précieuse pour la recherche. Les collections Web sont issues de collectes thématiques[31] que tout chercheur accrédité peut consulter dans les salles de recherche des différents sites de la BnF ou dans les établissements partenaires en région [32].

Plusieurs projets de recherche ont préfiguré des outils d'exploitation des archives du Web. Le partenariat avec l'équipe du projet ANR Web90[33] a permis d'élaborer une application « Archives Web Labs » proposant l'indexation en plein texte de deux corpus : les « incunables du Web » (1996-2000) et la collecte « attentats » de 2015 [34]. Ces fonctionnalités ont été déployées sur les postes d'accès de la bibliothèque de recherche, tandis qu'un nouveau corpus, la collecte « Actualités » (2010-2017), venait s'y ajouter dans le cadre du projet de recherche « Neonaute », portant sur la réalisation d'un moteur de recherche et d'études terminologiques sur les néologismes dans la langue française. Un quatrième corpus, consacré à l'épidémie COVID-19 en cours de collecte, est également indexé.

Grâce à ces projets de recherche, de nouveaux services sont ainsi venus progressivement enrichir l'interface de consultation des archives de l'Internet. Une vingtaine de parcours guidés[35], élaborés par des bibliothécaires et des chercheurs, permettent d'explorer, par exemple, les journaux personnels et littéraires en ligne, les mémoires de l'immigration maghrébine ou le Web électoral. L'indexation en plein texte de toutes les archives n'est pas possible mais les chercheurs doivent pouvoir interroger et manipuler les données grâce à des scripts et constituer leurs propres corpus. Pour cela, le BnF DataLab proposera des services et outils de traitement spécifiques aux archives de l'Internet : aide à la constitution de corpus, collectes Web à la demande, extraction des données archivées, accompagnement à la fouille de données Web archivées… Les chercheurs pourront librement traiter les fichiers au moyen des logiciels proposés par le BnF DataLab ou mettre en œuvre des programmes ou des scripts dans le respect du cadre légal.

S'ils sont promis à un grand avenir, ces services autour des archives de l'Internet restent encore expérimentaux et seront limités, dans un premier temps, aux projets en partenariat

---

[30] Les collectes s'appliquent uniquement aux sites du domaine .fr ou dont le producteur est domicilié en France conformément au cadre juridique définissant le dépôt légal de l'Internet. En savoir plus : https://www.bnf.fr/fr/archives-de-linternet

[31] En savoir plus : https://www.data.gouv.fr/fr/datasets/collectes-thematiques-du-web-par-la-bnf/

[32] Les archives de l'Internet sont également consultables dans 19 bibliothèques de dépôt légal imprimeur en région offrant un accès distant. En savoir plus : https://www.bnf.fr/fr/cooperation-regionale-et-action-territoriale-de-la-bnf

[33] ANR Web90 (Patrimoine, Mémoires et Histoire du Web dans les années 1990) : https://web90.hypotheses.org/

[34] Dans le cadre du projet ASAP (Archives sauvegarde attentats Paris) financé par le CNRS : https://asap.hypotheses.org/a-propos

[35] En savoir plus : https://www.bnf.fr/fr/parcours-guides-archives-de-linternet

avec la BnF ou sélectionnés dans le cadre d'appels à projet spécifiques. L'objectif est de poursuivre la co-construction des services et des outils avec les équipes de recherche, notamment lors de dathatons[36] organisés par le BnF DataLab. Ainsi, le projet ResPaDon (2021-2023)[37] mobilise le laboratoire GERiiCO (université de Lille) et le Medialab (Sciences Po Paris) autour des usages des collections Web par les chercheurs, dans la perspective de préfigurer des services accessibles à distance dans un réseau de bibliothèques partenaires.

Un laboratoire favorisant l'expérimentation et la R&D en partenariat

Dès l'origine, le développement du BnF DataLab a été pensé pour travailler en partenariat avec des chercheurs de différents profils et rattachés à différentes institutions scientifiques.

Un partenariat avec Huma-Num et des chercheurs en résidence

Pour mener à bien ces objectifs, la BnF intensifie ses partenariats scientifiques avec des acteurs de la recherche, au premier rang desquels le CNRS et Huma-Num, la très grande infrastructure de recherche (TGIR) dédiée aux humanités numériques. La TGIR Huma-Num favorise la coordination de recommandations scientifiques, la définition de bonnes pratiques technologiques et la co-conception de services numériques. Elle développe un dispositif technologique[38] permettant le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche en Lettres et SHS. Le BnF DataLab partage avec Huma-Num une même préoccupation quant à la pérennisation des données de la recherche et à la mutualisation des outils. Le défi sera d'articuler l'offre de services d'Huma-Num avec l'infrastructure informatique du DataLab afin de permettre un accompagnement complet autour de la valorisation de la recherche en humanités numériques.

Un partenariat a également été noué avec l'École polytechnique fédérale de Lausanne (EPFL) et son laboratoire d'humanités digitales (DHLab). Un autre est à l'étude avec le SCAI (Sorbonne Center for Artificial Intelligence) et l'ObTIC[39] (Observatoire des textes, des idées et des corpus) de Sorbonne Université afin d'accueillir dans le BnF DataLab des chercheurs sur une longue durée. De tels partenariats permettent à la BnF de bénéficier de l'expertise de ces institutions, de leurs moyens humains et de leurs outils de travail dans le champ des humanités numériques. C'est enfin une opportunité pour valoriser les actions qui seront menées dans le BnF DataLab : manifestations, restitutions des travaux des équipes de recherche, publications,

[36] Un datathon est un événement regroupant sur quelques jours des spécialistes du traitement des données pour travailler de façon collaborative sur de la programmation informatique.

[37] Financé par le GIS Collex-Persée, porté par l'Université de Lille et la BnF, en partenariat avec Sciences Po Paris et le Campus Condorcet, le projet ResPaDon (Réseau de Partenaires pour l'analyse et l'exploration de données numériques)) vise à développer et à diversifier les usages par les chercheurs des archives du Web. En savoir plus : https://www.collexpersee.eu/projet/respadon/

[38] Cf. services et outils d'Huma-Num : https://www.huma-num.fr/services-et-outils

[39] L'équipe ObTIC a pris la suite du Labex OBVIL : https://obtic.sorbonne-universite.fr/

etc. Les futurs appels à projet de la BnF seront l'occasion de mettre en valeur le BnF DataLab et de proposer des projets de recherche communs avec Huma-Num.
En raison d'un engagement de la bibliothèque qui dépasse l'offre de services du BnF DataLab, certains projets de recherche nécessitent un cadre de conventionnement entre l'équipe de chercheurs et la BnF : convention d'accueil (sans participation scientifique de la BnF) ou convention de partenariat (permettant une co-construction des résultats du projet).
La BnF peut accueillir en résidence des équipes de recherche et chercheurs individuels dans le cadre de projets de recherche nécessitant un usage intensif du BnF DataLab pendant une période de 3 mois à 1 an (renouvelable). L'accueil en résidence s'inscrit dans un partenariat, un appel à projets ou à chercheurs associés. Le BnF DataLab ambitionne ainsi de devenir le lieu d'accueil identifié par les chercheurs pour échanger, se former, discuter entre pairs et avec des experts, bénéficier de services à la carte et de résidences adaptées, où les partenariats multiples avec d'autres organismes de recherche permettront le partage et la valorisation des données de la recherche.

Coopération et valorisation scientifique

Au-delà d'un espace d'accueil et d'une offre de services, le BnF DataLab se veut un laboratoire favorisant l'expérimentation et la R&D, ainsi que le partage des résultats de recherche sur les collections numériques dans une démarche collaborative et contributive. Les équipes de recherche sont ainsi invitées à participer à l'enrichissement des ressources numériques de la BnF : intégration de données enrichies dans Gallica (OCR corrigé, EPUB, bientôt TEI, etc.) et/ou de corpus thématiques éditorialisés sous forme de sélections dans Gallica ; publication des jeux de données dans le site API et données de la BnF[40]; mise à disposition de scripts. L'aspect collaboratif est une dimension majeure de l'animation du BnF DataLab : les outils, développés dans le cadre d'un programme de recherche, seront reversés dans une boîte à outils et pourront bénéficier à d'autres projets mais aussi aux programmes internes, comme la reconnaissance d'images, le séquençage de la presse ancienne, la détection automatique des styles littéraires, etc. Les outils et les jeux de données ainsi réalisés devront se conformer aux recommandations d'Huma-Num (open access, FAIR data, interopérabilité, etc.). Ils seront également accessibles et documentés sur GitHub.
La coopération scientifique se trouve ainsi au cœur du projet BnF DataLab. Au-delà de la fouille de textes, de nouveaux outils seront développés en R&D avec les équipes de recherche (IA, segmentation des documents, *topic modeling*, etc.) et mutualisés dans le BnF DataLab. À terme, ils viendront enrichir Gallica, servant ainsi tous les usagers, au-delà de la communauté scientifique.

[40] Site API et données de la BnF : http://api.bnf.fr/accueil

Modèle économique et moyens

Si le projet s'inscrit, comme nous l'avons vu, dans une logique et un *continuum* numérisation-recherche-expérimentation, l'ouverture d'un DataLab à la BnF n'en représente pas moins un changement d'échelle qui nécessite d'importants moyens et pose la question cruciale du modèle économique d'un tel laboratoire au sein de l'établissement.

Rappelons que le modèle économique de la numérisation repose, d'une part, sur une politique de subvention qui s'attache à des programmes d'intérêt général dans le cadre de la coopération, et, d'autre part, sur des prestations payantes qui répondent à des intérêts particuliers ou privés. Ces dernières relèvent des services usuels de reproduction des documents patrimoniaux pour les professionnels et les particuliers, ou de fournitures de métadonnées spécifiques à l'interprofessionnel avec des produits bibliographiques sur mesure (dits « à façon ») pour les bibliothèques.

Le BnF DataLab se trouve au carrefour de ces deux voies et pose la problématique centrale des moyens et d'un modèle économique à trouver pour assurer ces nouveaux services. Au-delà de l'effort conséquent auquel consent la BnF afin d'ouvrir ce nouveau lieu, il s'agit de distinguer ce qui va relever du service gratuit ou de la prestation payante, alors que la disponibilité des ressources numériques et leur accessibilité sont des prérequis, conditions indispensables à toute exploitation de corpus.

La BnF souhaite mettre en place un modèle économique qui articule trois niveaux de (1) services, (2) formations et (3) prestations. S'adressant à tous les chercheurs, le premier niveau de services, gratuits, est centré sur l'accueil et l'accompagnement. Le second niveau de formations, gratuites elles aussi, porte sur la connaissance des ressources et formats de la BnF et sur la prise en main des outils disponibles (notamment les API), dans la perspective de rendre les équipes de recherche autonomes dans leur exploitation des collections numériques. Enfin, un troisième niveau de prestations, cette fois payantes[41], concerne les demandes spécifiques et les produits à façon (numérisation, extraction de données, extension de l'infrastructure numérique, etc.) dont pourraient avoir besoin les projets de recherche.

Pour les collections déjà numérisées et disponibles dans Gallica, les chercheurs peuvent utiliser les API et jeux de données de la BnF[42]. Soulignons que les API, développées en 2017 à l'occasion du hacktahon BnF, peuvent constituer une limite technique et qu'elles nécessitent aujourd'hui d'être renforcées et mieux documentées afin de répondre à ces nouveaux usages.

Les questions des moyens se posent non seulement d'un point de vue économique, mais aussi technique et humain, avec une évolution des métiers et des compétences vers de l'expertise en matière de données. Un ingénieur d'étude a été spécifiquement recruté pour faire

[41] Le modèle de la prestation répond à la tarification officielle de la BnF. Pour les demandes de numérisation portant sur de grandes quantités, la décision tarifaire autorise le demandeur à négocier une réduction, en prenant en compte l'économie générale de son projet (ANR, etc.)

[42] Portail BnF API et jeux de données : https://api.bnf.fr/

l'interface entre la TGIR Huma-Num et le BnF DataLab. Sur le plan technique, il s'agit de renforcer les API et de mettre en place d'une infrastructure numérique spécifique qui permette notamment aux chercheurs de lancer des scripts et de travailler sur les collections sous droits. Ce nouvel environnement de travail, qui se présente sous forme de machines virtuelles dédiées, pose des contraintes en termes de ressources (espace serveur, mémoire vive, carte graphique, etc.) et de sécurité pour la BnF. À ces limites techniques s'ajoutent des contraintes juridiques qui encadrent et peuvent restreindre l'exploitation des collections numériques, notamment celles sous droits (Gallica intramuros et archives du Web). Nous n'aborderons pas ici ces questions juridiques qui néanmoins sont impliquantes en termes de fouille de données, d'exposition et de partage des résultats de la recherche.

Conclusion

À l'heure où nous écrivons ces lignes, le BnF DataLab n'est pas encore ouvert et il reste encore de nombreux défis à relever : mettre en œuvre les procédures pour assurer les services et prestations ; former les agents pour accueillir les chercheurs et initier à de nouveaux outils ; enfin animer et faire vivre ce lieu expérimental et singulier.
Le BnF DataLab préfigure une nouvelle génération de services en bibliothèques, articulant espace physique et virtuel, accueillant des pratiques mixtes, individuelles et collectives, tournées à la fois vers l'exploration des ressources numériques et la production de connaissance et d'outils. Cherchant à répondre aux demandes des chercheurs et à les accompagner au mieux pour s'emparer des collections numériques, le BnF DataLab ambitionne aussi de devenir un lieu d'expérimentation et de préfiguration des usages de demain, un lieu d'observation et d'analyse.

Avec plus de 8 millions de documents de tous types dans Gallica, la BnF et ses partenaires deviennent des acteurs centraux dans les humanités numériques. La numérisation a complètement transformé le rapport au document et, pour le chercheur, sa relation aux sources primaires, lesquelles deviennent accessibles facilement et en nombre, mais surtout interrogeables et navigables. Cette relation nouvelle aux sources a évolué dans le temps, avec un premier accès au document en mode image, puis au texte avec les techniques de reconnaissance optique de caractères (OCR et HTR) qui permettent de rendre celui-ci interrogeable (requête, recherche plein texte, etc.). Aujourd'hui, l'enjeu est l'accessibilité de la structure même du document afin de pouvoir naviguer à l'intérieur et cibler le grain d'information (article de presse, notice de dictionnaire, légende d'images, titre des rubriques etc.), par le recours à l'intelligence artificielle (OLR) afin de traiter les documents en nombre.
De nouvelles modalités de lecture et d'exploitation des documents s'ouvrent aux chercheurs avec, d'un côté, l'exploitation immédiate des collections avec des outils de consultation et de

dépouillement ; de l'autre, l'exploitation massive des corpus, qui s'appuie sur des nouveaux outils impliquant une autre temporalité. Cette évolution des usages, qui ne concernent encore qu'un nombre restreint de chercheurs et de bibliothèques, n'en change pas moins le rapport aux collections et à l'espace même de la bibliothèque.

Avec ces nouveaux services, la bibliothèque se positionne moins comme prestataire que comme partenaire renouvelant les relations entre chercheurs et bibliothécaires, pour véritablement travailler de concert dans un échange mutuel de compétences. Car si les humanités numériques transforment profondément le travail des chercheurs, elles modifient tout autant le métier des professionnels des bibliothèques, impliquant l'appropriation de nouvelles compétences pour investir un rôle de médiation numérique et d'acculturation à la science des données.

Pour la Bibliothèque nationale de France, le BnF DataLab nécessite à la fois une transformation des espaces, des services et des métiers. En partant des usages et en instaurant des modalités nouvelles de coopération avec les chercheurs, la BnF réaffirme son rôle de premier plan dans l'écosystème de la recherche, non seulement en fournissant de nouveaux services, mais aussi en cherchant à valoriser et réintégrer les outils et les résultats de la recherche pour coproduire les services de demain.

Références bibliographiques

Barret, Elydia. 2014. « Quel rôle pour les bibliothèques dans les humanités numériques ? » Mémoire de fin d'étude du diplôme de conservateur, École nationale supérieure des sciences de l'information et des bibliothèques (Enssib).

Bermès, Emmanuelle. 2019. « BnF : des métadonnées au service de projets de recherche innovants. » *Arabesques*, 2019 (95) : 8-9. https://publications-prairial.fr/arabesques/index.php?id=1302

Bermès, Emmanuelle, et Eleonora Moiraghi. 2020. « Le patrimoine numérique national à l'heure de l'intelligence artificielle. Le programme de recherche Corpus comme espace d'expérimentation pour les humanités numériques. » *Revue Ouverte d'Intelligence Artificielle* (1) : 89-109. https://roia.centre-mersenne.org/item/ROIA_2020__1_1_89_0/

Bermès, Emmanuelle. 2020. « Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque nationale de France (1997-2019) ». Thèse de doctorat, École nationale des chartes. http://www.theses.fr/2020ENCP0001/document

Bertrand, Sophie, et Aline Girard. 2016. « Gallica (1997- 2016) : de la bibliothèque de l'honnête homme à celle du Gallicanaute ». *Bulletin des bibliothèques de France* (BBF), 2016 (9) : 48-59. https://bbf.enssib.fr/consulter/bbf-2016-09-0048-005

Caraco, Benjamin. 2012. « Les digital humanities et les bibliothèques : un partenariat naturel ». *Bulletin des bibliothèques de France (BBF)*, 2012 (2) : 69-73. https://bbf.enssib.fr/consulter/bbf-2012-02-0069-002

Catzaras, Nicolas, et Johann Holland. 2012. « Quels espaces physiques pour les humanités numériques ? ». Atelier THATCamp Paris 2012. Paris : Éditions de la Maison des sciences de l'homme. https://books.openedition.org/editionsmsh/380

Dacos, Marin, et Pierre Mounier. 2014. *Humanités numériques : État des lieux et positionnement de la recherche française dans le contexte international*. Paris : Institut français. https://hal.archives-ouvertes.fr/hal-01228945

Éloi, Catherine, Eleonora Moiraghi et Virginie Rose. 2019. « Un espace pour les humanités numériques à la BnF ». *Bulletin des Bibliothèques de France*, 2019 (17) : 90-95. http://bbf.enssib.fr/consulter/bbf-2019-17-0090-009

Fickers, Andreas, et Tim Van der Heijden. 2020. « Inside the Trading Zone: Thinkering in a Digital History Lab ». *Digital Humanities Quarterly (DHQ)* 14, 2020 (3). http://digitalhumanities.org/dhq/vol/14/3/000472/000472.html

Gallinger, Michelle, et Daniel Chudnov. 2016. « Library of Congress Digital Scholars Lab Pilot Project Report 2016 ». https://labs.loc.gov/static/labs/work/reports/DChudnov-MGallinger_LCLabReport.pdf.

Jeanneney, Jean-Noël. 2005. *Quand Google défie l'Europe. Plaidoyer pour un sursaut*. Paris : Mille et une nuits.

Jouguelet, Suzanne. 2009. « Les *Learning centres* : un modèle international de bibliothèque intégrée à l'enseignement et à la recherche. Rapport à madame la ministre de l'Enseignement supérieur et de la Recherche. » https://www.enssib.fr/bibliotheque-numerique/notices/48085-learning-centres-les-un-modele-international-de-bibliotheque-integree-a-l-enseignement-et-a-la-recherche

Kear, Robin, et Kate Joranson. 2018. *Digital Humanities, Libraries, and Partnerships. A Critical Examination of Labor, Networks, and Community*. Chandos Publishing.

Le Deuff, Olivier. 2014. « Bibliothèques et lieux de production de savoirs ». *Le Temps des humanités digitales.* Limoges : FYP éditions. https://archivesic.ccsd.cnrs.fr/sic_01487073/

Le Deuff, Olivier. 2013. « Humanités numériques et bibliothèques. » *THATCamp Saint-Malo, Non actes de la non conférence*. Paris : Éditions de la Maison des sciences de l'homme. https://books.openedition.org/editionsmsh/2201

Machefert, Sylvain. 2014. « Quelle place pour les bibliothèques dans les *digital humanities* ? L'exemple de Bordeaux 3. » *Le Temps des humanités digitales*. Limoges : FYP éditions. https://hal.archives-ouvertes.fr/hal-01494347

Millson-Martula, Christopher, et Kevin Gunn. 2017. « The digital humanities: Implications for librarians, libraries, and librarianship » *College & Undergraduate Libraries* (24). https://www.tandfonline.com/doi/full/10.1080/10691316.2017.1387011

Moiraghi, Eleonora. 2018. « Le projet Corpus et ses publics potentiels : une étude prospective sur les besoins et les attentes des futurs usagers. » Paris : Bibliothèque nationale de France. https://hal-bnf.archives-ouvertes.fr/hal-01739730/document

Moretti, Franco. 2013. *Distant Reading*. London/New-York : Verso.

Moreux, Jean-Philippe. 2016. « Approches innovantes pour la presse ancienne numérisée : fouille et visualisation de données. *Carnet de la recherche à la Bibliothèque nationale de France* (blog). https://bnf.hypotheses.org/208

Pardé, Thierry, et Olivier Jacquot. 2016. « Les humanités numériques à la Bibliothèque nationale de France ». *Culture et recherche*. Paris : Ministère de la Culture et de la Communication. https://hal-bnf.archives-ouvertes.fr/hal-01379908

Posner, Miriam. 2013. « No Half Measures: Overcoming common challenges to Doing Digital Humanities in the Library ». *Journal of Library Administration* (53).

Roustan, Mélanie (dir.). 2016. *La recherche dans les institutions patrimoniales : sources matérielles et ressources numériques*. Villeurbanne : Presses de l'Enssib.

Snickars, Pelle. 2018. « Data Labs: Datalab.kb.se – A Report For the National Library of Sweden ». https://www.infodocket.com/2018/10/17/data-labs-datalab-kb-se-a-report-for-the-national-library-of-sweden/

Sula, Chris Alen. 2013. « Digital Humanities and Libraries: A Conceptual Model ». *Journal of Library Administration* (1). http://chrisalensula.org/digital-humanities-and-libraries-a-conceptual-model/

Svensson, Patrik. 2016. *Big Digital Humanities: Imagining a Meeting Place for the Humanities and the Digital*. University of Michigan Press, Ann Arbor. https://doi.org/10.2307/j.ctv65sx0t.1.

THATCamp Saint-Malo, 2013. « Quels espaces physiques pour les humanités numériques ? » *Non actes de la non conférence*. Paris : Éditions de la Maison des sciences de l'homme. https://books.openedition.org/editionsmsh/2211