

Le BnF DataLab

« Offrir aux chercheurs, dans les emprises de la Bibliothèque, des outils de fouille et d'exploration de textes et de données sur des corpus numériques de la BnF. »

Contrat d'objectifs et de performance
2017-2021

Humanités numériques

Analyse
computationnelle

Analyse
de
réseaux

Modélisation

Reconnaissance
automatique
d'écritures
manuscrites

Fouille de
données

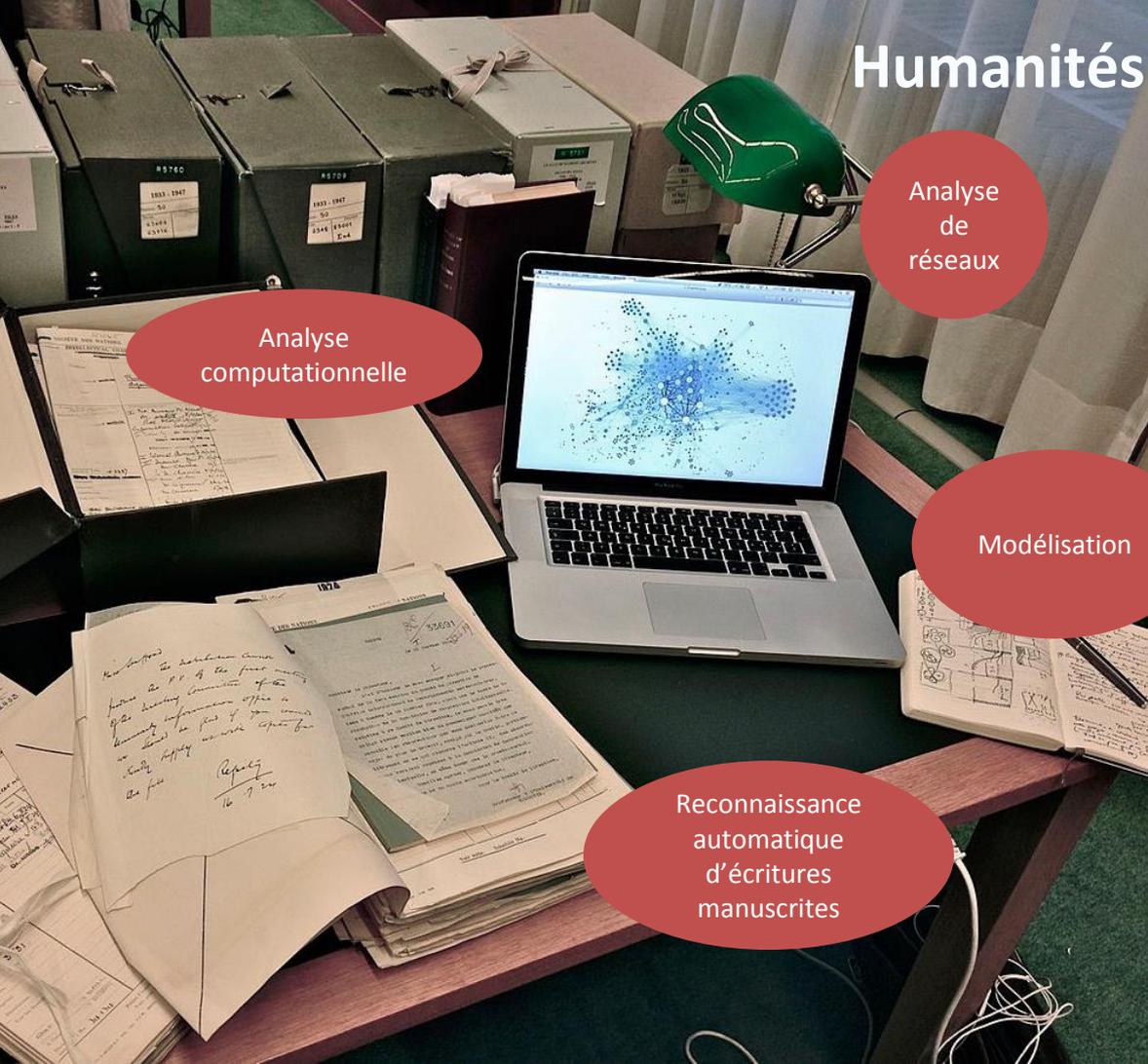
Entraînement
machine

Encodage
de texte

Reconnaissance
d'images

Programmation

Alignement de
données



Le projet Corpus : 2016-2019

Un projet inscrit au plan quadriennal de la recherche de la BnF 2016-2019

Objectifs :

préfigurer un service de fourniture de corpus numériques à destination de la recherche
fournir à des chercheurs des données et des outils pour les analyser, dans le respect du droit d'auteur et de la vie privée

3 années d'expérimentation, autour de 4 collections numériques : archives web, numérisation, métadonnées, images et cartes

+ 1 année de bilan

Foucault Fiches de Lecture



Projet qui a pour but de numériser, mettre en ligne, décrire et enrichir les manuscrits de notes de lecture de Michel Foucault, en utilisant une plate-forme numérique de travail collaboratif.

Numérisation des fiches

Reconnaissance de l'écriture manuscrite

Annotation et alignement avec data.bnf.fr

Enrichissement et visualisation des données

GiraNium (Paris-Sorbonne, CELSA)



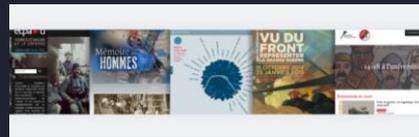
Projet autour des écrits d'Émile de Girardin. Corpus hétérogène composé de la correspondance, des monographies et d'articles de presse.

Numérisation d'un corpus hétérogène

Transcription

Extraction d'entités nommées

Mémoire de la Grande Guerre sur le web



Ce parcours guidé propose une exploration des axes forts de la présence de la Grande Guerre sur internet : travail participatif des internautes et la circulation des informations - entre historiens, chercheurs, enseignants, grand public -, qui ont parfois permis de faire émerger des thématiques inédites.

Gallicarte



Nouvelle fonctionnalité développée au cours du premier hackathon BnF. Affiche les résultats d'une recherche effectuée dans Gallica sur une carte et permet de visualiser l'ensemble d'un fonds photographique géolocalisé sur une carte interactive en présentant les documents et leurs métadonnées.

Le projet Corpus et ses publics potentiels.

UNE ÉTUDE PROSPECTIVE SUR
LES BESOINS ET LES ATTENTES
DES FUTURS USAGERS.

Eleonora Moiraghi, assistante de recherche
sur le programme de recherche Corpus porté
par la Bibliothèque nationale de France

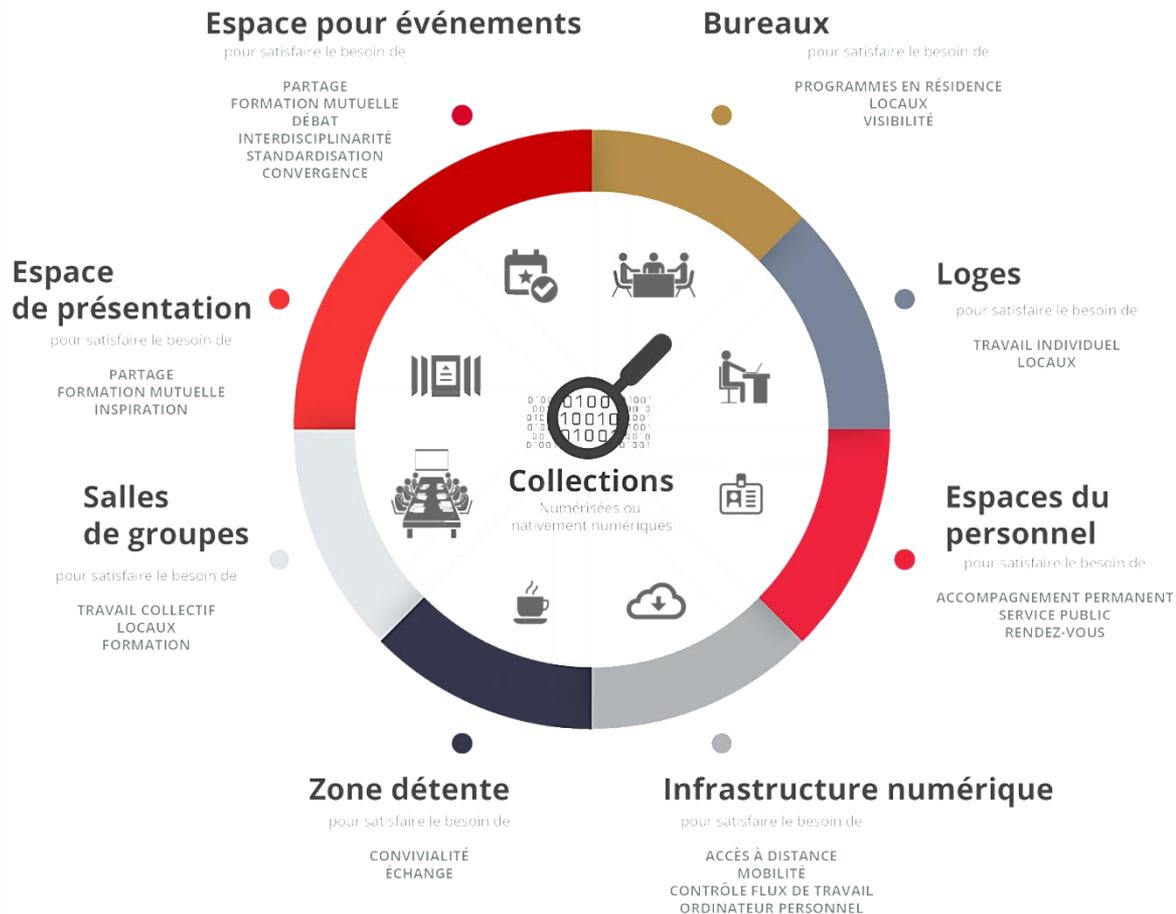
Janvier 2018

Étude prospective sur les besoins et les attentes des futurs usagers

- une enquête qualitative par entretiens auprès de trois populations (seize universitaires, onze agents de la BnF et trois experts du domaine) ;
- deux ateliers avec des équipes de recherche et des experts BnF (« Décrire, transcrire et diffuser un corpus documentaire hétérogène : méthodes, formats, outils », « Géolocalisation et spatialisation de documents patrimoniaux ») ;
- un atelier collaboratif utilisant la méthode des personas (six universitaires et quinze agents de la BnF) pour vérifier les besoins et les attentes manifestées lors des entretiens individuels.

Eleonora Moiraghi, *Le projet Corpus et ses publics potentiels : une étude prospective sur les besoins et les attentes des futurs usagers*. [Rapport de recherche] Bibliothèque nationale de France. Janvier 2018. (hal-01739730)

Un espace dédié aux humanités numériques à la BnF



Des services d'accompagnement



Les services du DataLab

- Accueil et présentation des services
- Rendez-vous experts DataLab
- Découverte du DataLab et présentation des outils

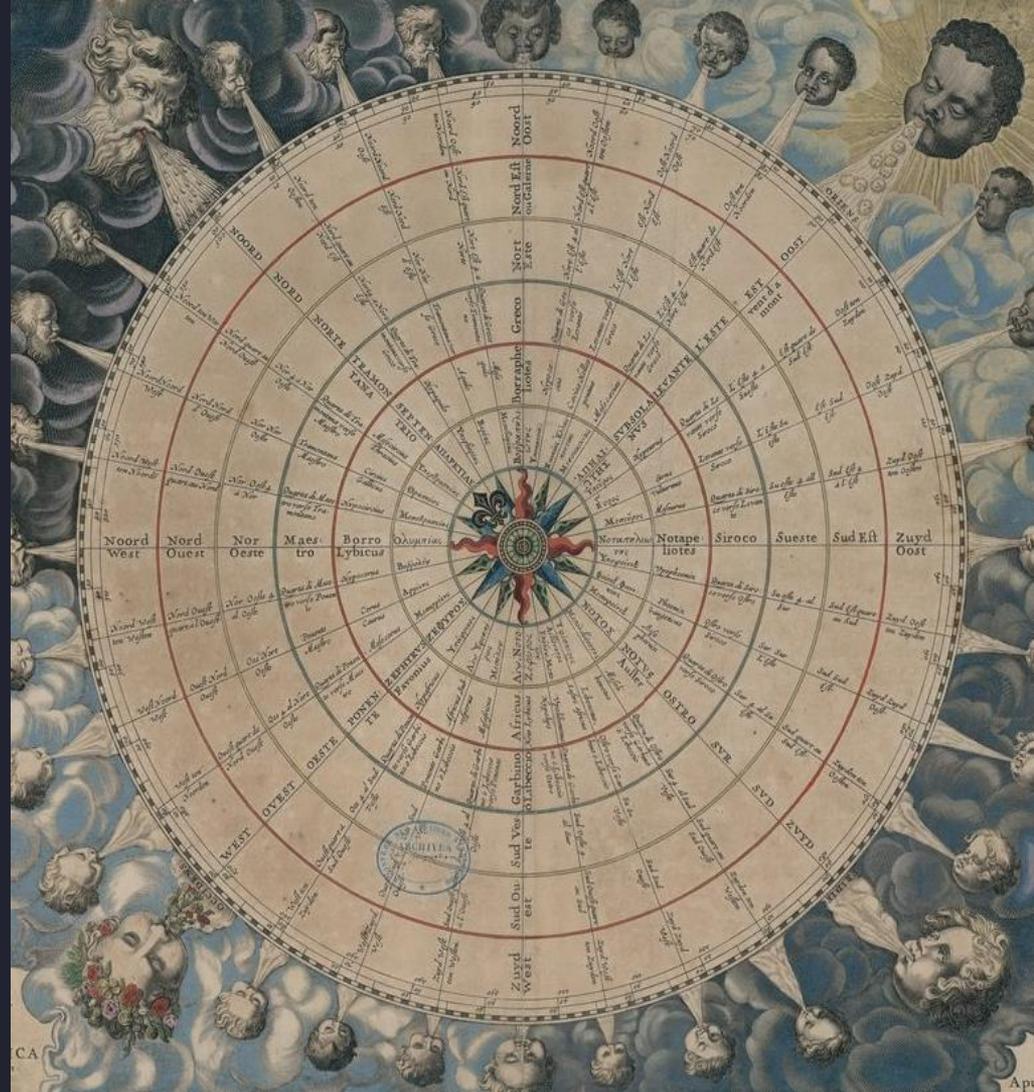
- Aide à la constitution de corpus
- Aide à la formulation de requêtes et à l'utilisation des outils et services en ligne (API)
- Numérisation à la demande (DIP)
- Extraction de collections numériques
- Collecte web à la demande
- Extraction d'une archive d'un site web
- Manifestations scientifiques

- Accompagnement à la fouille de corpus web
- Extraction et traitement de données
- Offre de formation
- Mise à disposition d'une infrastructure numérique
- Mise à disposition d'une boîte à outils logiciels

- Conventionnement et partenariat
- Résidence de chercheurs
- Coopération et valorisation scientifique
- Accompagnement de projets de recherche

Centraliser les demandes pour mieux les orienter

- Accueil et présentation des services
- Rendez-vous experts DataLab
- Découverte du DataLab et présentation des outils



Constituer son corpus

- Aide à la constitution de corpus
- Aide à la formulation de requêtes et à l'utilisation des outils et services en ligne (API)
- Numérisation à la demande (DIP)
- Extraction de collections numériques
- Collecte web à la demande
- Extraction d'une archive d'un site web
- Manifestations scientifiques



Travailler sur son corpus

- Accompagnement à la fouille de corpus web
- Extraction et traitement de données
- Offre de formation
- Mise à disposition d'une infrastructure numérique
- Mise à disposition d'une boîte à outils logiciels



Habit de Barbier Perruquier,

Partenariat et résidences

Les projets de recherche sélectionnés pourront faire l'objet d'un accompagnement de la BnF sous diverses formes :

- Convention d'accueil
- Convention de partenariat
- Accueil en résidence

<https://gallica.bnf.fr/ark:/12148/btv1b90120733>



POUR VOS VOYAGES VILLÉGIATURES

CONSULTEZ LES CARTES-TOURISME DE L'INDICATEUR DU TOURISME

EN VENTE PARTOUT

LE MOINS EMBARRASSANT LE PLUS PRATIQUE 1^{fr} 50

Renseignements 13, rue Drouot, PARIS

OFFICE CENTRAL DU TOURISME FRANÇAIS

A. S. P. M. 13, rue Drouot, PARIS

Identifier d'autres acteurs

Tisser des partenariats

Compléter les services

Élargir les possibilités



Parcours usager pour intégrer le DataLab



Chercheur individuel,
doctorant, post
doctorant



Equipe de recherche,
laboratoire



Partenaires
institutionnels BnF

Parcours usager chercheur « simple »

L'utilisateur prend connaissance du service et veut en savoir plus



Il est orienté vers la page internet et invité à remplir le formulaire



Accueil des chercheurs par le binôme ORB



Constitution du dossier de suivi



Prise de contact avec les départements
Actualisation du dossier (calendrier, livrables...)



Soumission du dossier au COS



Élaboration d'un parcours partenariat ou prestations



Démarrage du projet

Élaboration d'un parcours gratuit 8h et/ou prestations



Parcours usager équipe de recherche

L'équipe prend connaissance du service et veut en savoir plus



Le porteur de projet est orienté vers la page internet et invité à remplir le formulaire



Accueil des chercheurs par le binôme ORB + DSG



Constitution du dossier de suivi



Prise de contact avec les départements
Actualisation du dossier (calendrier, livrables...)



Soumission du dossier au COS



Élaboration d'un parcours partenariat ou prestations



★ ★ ★ Démarrage du projet ★ ★

Bibliographie

Ateliers Corpus

Olivier Jacquot, « Décrire, transcrire et diffuser un corpus documentaire hétérogène : méthodes, formats, outils ». *Carnet de la recherche à la Bibliothèque nationale de France*, 29 novembre 2017. Disponible en ligne : <<https://bnf.hypotheses.org/2214>> (consulté le 7 octobre 2019).

Eleonora Moiraghi, « Penser, classer, modéliser. L'exemple du projet Foucault Fiches de Lecture ». *Carnet de la recherche à la Bibliothèque nationale de France*, 21 décembre 2018. Disponible en ligne : <<https://bnf.hypotheses.org/7445>> (consulté le 7 octobre 2019).

Eleonora Moiraghi, « Explorer des corpus d'images. L'IA au service du patrimoine ». *Carnet de la recherche à la Bibliothèque nationale de France*, 16 avril 2018. Disponible en ligne : <<https://bnf.hypotheses.org/2809>> (consulté le 7 octobre 2019).

Eleonora Moiraghi, « Géolocalisation et spatialisation de documents patrimoniaux : trois heures de partage autour de la cartographie numérique ». *Carnet de la recherche à la Bibliothèque nationale de France*, 20 décembre 2017. Disponible en ligne : <<https://bnf.hypotheses.org/2299>> (consulté le 7 octobre 2019).

Eleonora Moiraghi, « Données liées et données à lier : quels outils pour quels alignements ? ». *Carnet de la recherche à la Bibliothèque nationale de France*, 19 juillet 2018. Disponible en ligne : <<https://bnf.hypotheses.org/4128>> (consulté le 7 octobre 2019).

Présentation du projet Corpus

Catherine Éloi, Eleonora Moiraghi, Virginie Rose, « Un espace pour les humanités numériques à la BnF ». Bulletin des bibliothèques de France (BBF), 2019, n° 17, p. 90-95. Disponible en ligne : <<http://bbf.enssib.fr/consulter/bbf-2019-17-0090-009>> (consulté le 7 octobre 2019).

Eleonora Moiraghi, *Le projet Corpus et ses publics potentiels : une étude prospective sur les besoins et les attentes des futurs usagers*. [Rapport de recherche] Bibliothèque nationale de France, 2018. Disponible en ligne : <<https://hal-bnf.archives-ouvertes.fr/hal-01739730/document>> (consulté le 7 octobre 2019).

Emmanuelle Bermès, « BnF : des métadonnées au service de projets de recherche innovants ». Arabesques, 2019, n° 95, p. 8-9. Disponible en ligne : <<http://www.abes.fr/Publications-Evenements/Arabesques/Arabesques-n-95>> (consulté le 17 octobre 2019).

Emmanuelle Bermès, Eleonora Moiraghi, « 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 d'Intelligence Artificielle (RIA), à paraître. Disponible en ligne : <https://hal-bnf.archives-ouvertes.fr/hal-02122073/document> (consulté le 17 octobre 2019).

Histoire des collections numériques

Emmanuelle Bermès, *Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque nationale de France (1997-2019)*. Histoire. Paris, École nationale des chartes, 2020. Français. ([NNT : 2020ENCP0001](#)). (<tel:02475991>)