CONSTRUIRE ENSEMBLE LE PATRIMOINE NUMÉRIQUE NATIONAL

LE PROJET CORPUS À LA BNF

Emmanuelle Bermès

BnF Bibliothèque nationale de France

2e rencontres Huma-Num | juin 2018

1 Bibliothèque et Recherche

COLLECTER LES RESSOURCES

SIGNALER LES RESSOURCES

CONSTITUER LES COLLECTIONS

ENRICHIR LES COLLECTIONS

CONTEXTUALISER LES RESSOURCES

MAÎTRISER LES RESSOURCES

TOUTE DISCIPLINE

TOUT PUBLIC

« Encourager le partage et la dissémination des savoirs »

Contrat de performance 2017-2021

BnF Bibliothèque nationale de France


INTEROPERABILITÉ TOTALE
MAITRISE PARFAITE DES ACCÈS
TEMPS CONTINU
EXHAUSTIVITÉ DES CORPUS
AUTONOMIE TOTALE DU CHERCHEUR
CAPACITÉS TECHNIQUES INFINIES
GRANULARITÉ PARFAITE
PERFECTION DES ALGORITHMES
UNIVERSALITÉ DE LA (DATA)VISUALISATION
EXHAUSTIVITÉ DE LA DOCUMENTATION
TRACABILITÉ REVERSIBILITÉ
LIBERTÉ JURIDIQUE

2 Parcours dans les collections

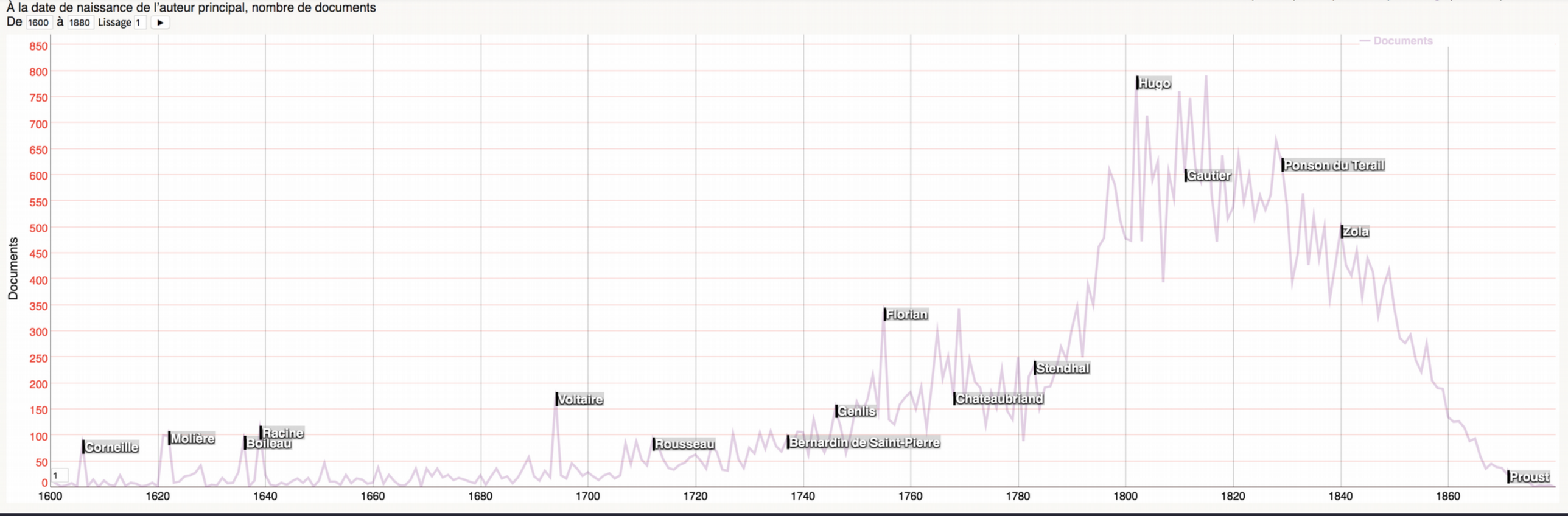
Esprit

Gargantua

soif intarissable de données,
travail sur d’énormes corpus
qui s’enrichissent continument

« Et lui furent ordonnées dix et sept mille neuf cents treize vaches pour l'allaicter ordinairement »
Gustave Doré, 1854
Source gallica.bnf.fr / Bibliothèque nationale de France

BnF Bibliothèque nationale de France



Projet OBVIL
Common places

Fourniture de
135 000
fichiers ALTO
à des fins de fouille
de texte

Site internet :
http://obvil.lip6.fr/tgb/

BnF Bibliothèque nationale de France

2 Parcours dans les collections

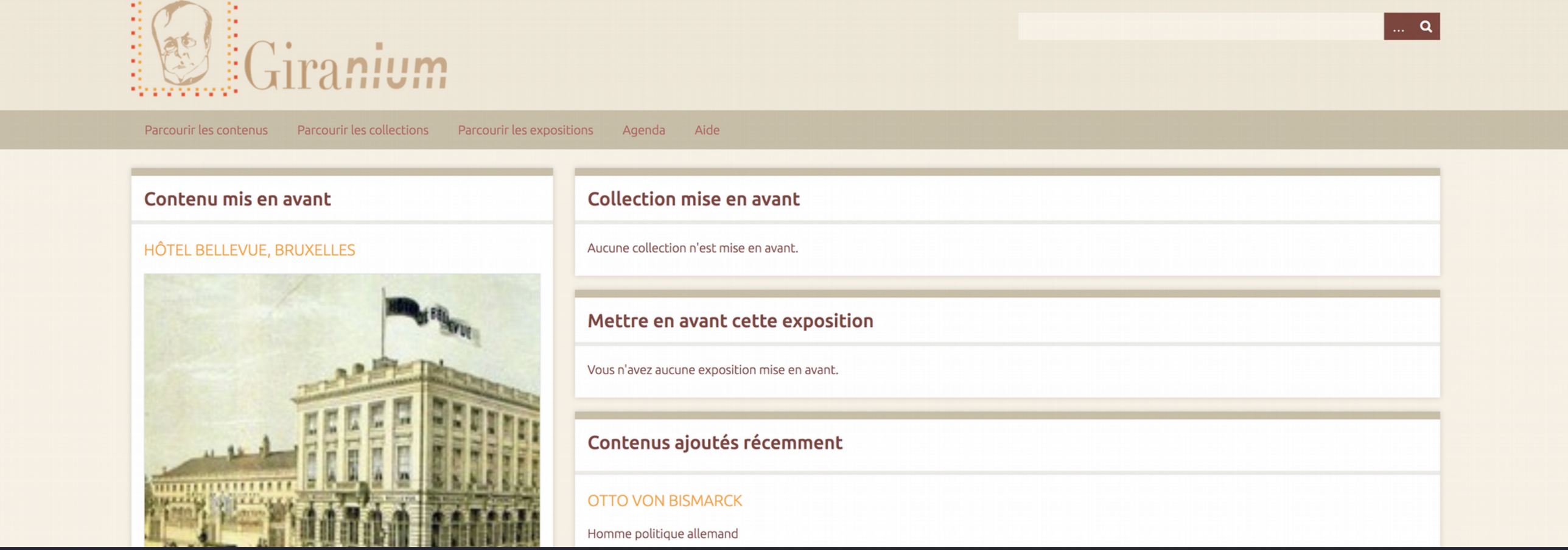
Esprit

Arcimboldo

rassembler puis étudier des sources éparses et/ou hétérogènes

Vertumne
Giuseppe Arcimboldo, 1590
Château de Skokloster

BnF Bibliothèque nationale de France



Projet GRIPIC/ CELSA

Giranium

Analyse de 300 lettres autographes d'Émile de Girardin conservées à l'Institut et mise en relation avec d'autres documents (manuscrits, presse, revues...)

Site internet : http://vintagedata.org/giranium/

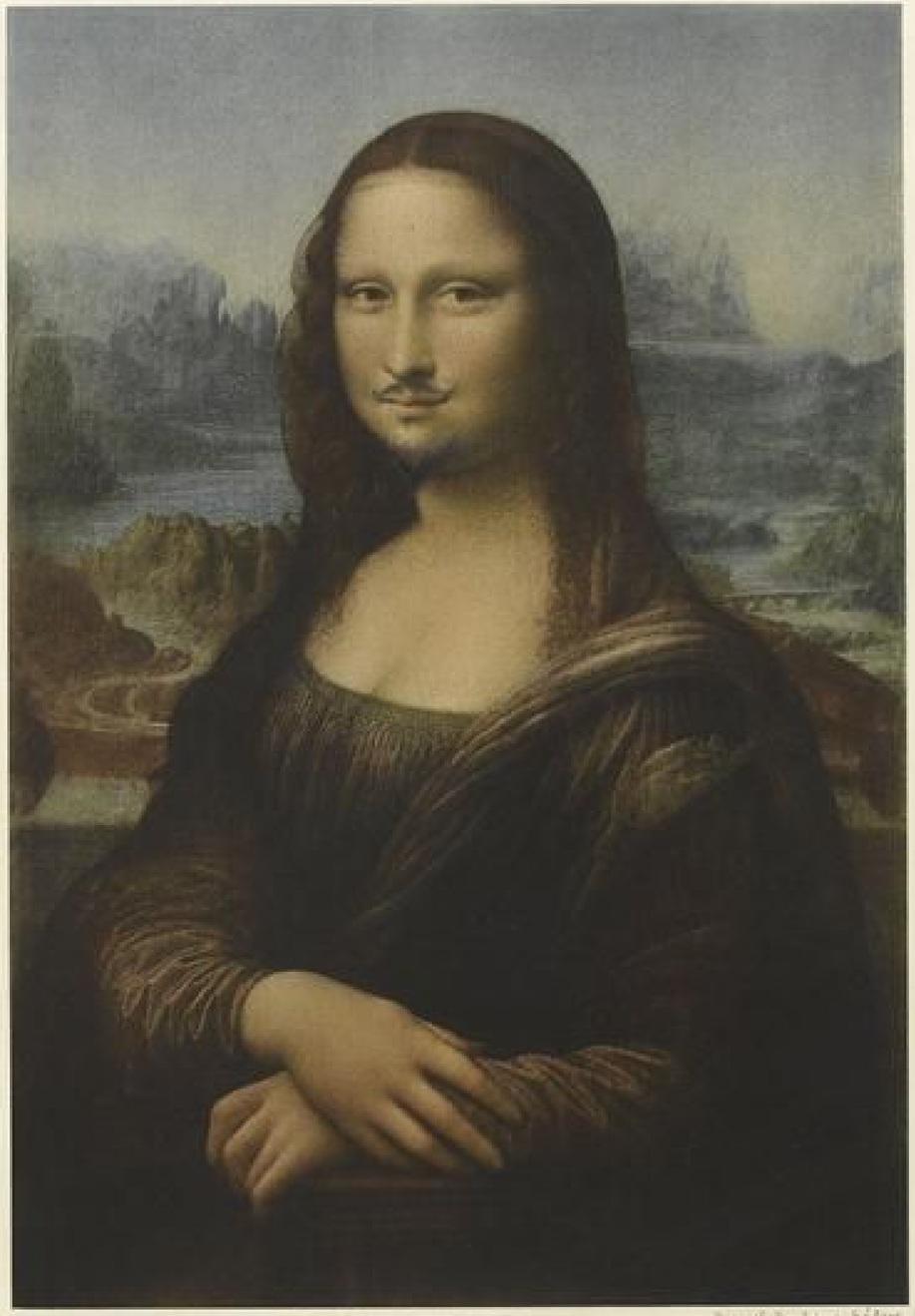
BnF Bibliothèque nationale de France

2 Parcours dans les collections

Esprit

Duchamp

donner à (re)lire des contenus déjà existants, déjà connus



L.H.O.O.Q.
Marcel Duchamp, 1919

BnF Bibliothèque nationale de France

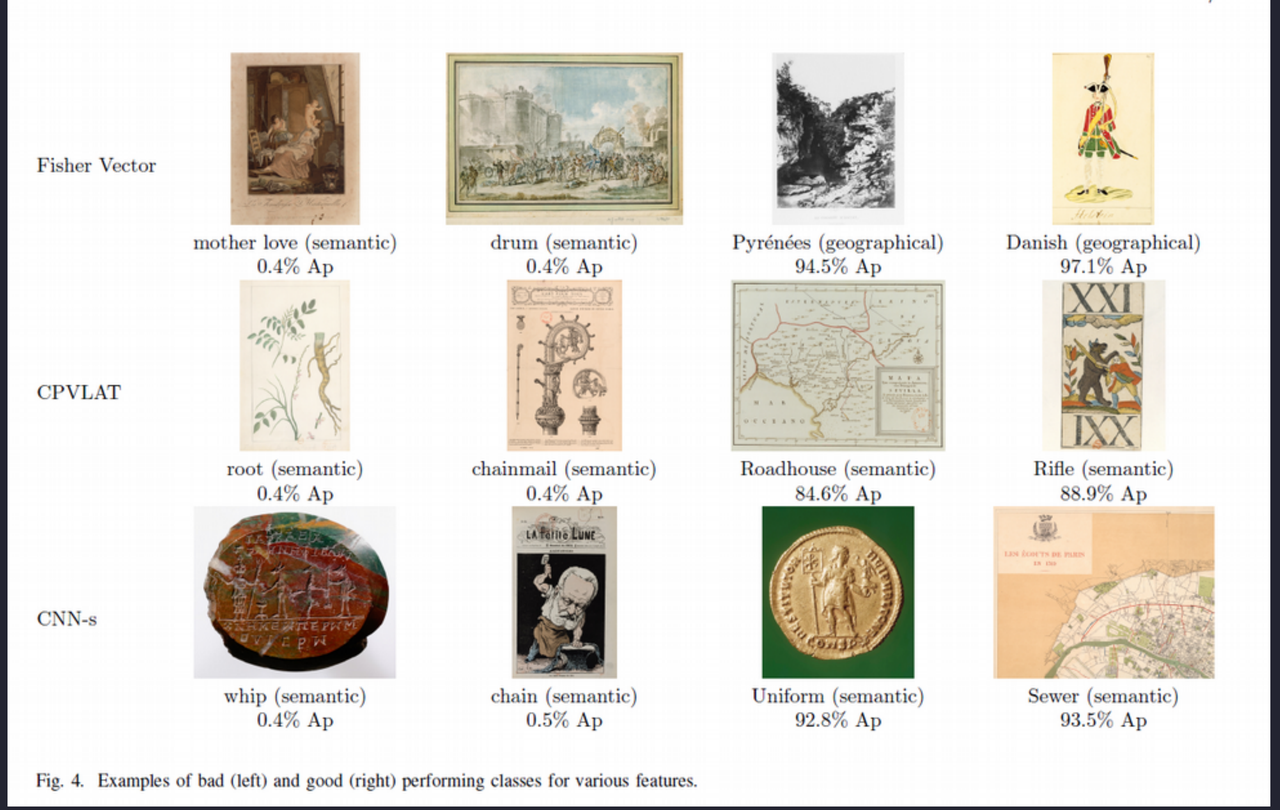


Fig. 4. Examples of bad (left) and good (right) performing classes for various features.

Utilisation d'images issues de la banque d'images de la BnF pour entraîner un algorithme d'indexation sémantique

Laboratoire ETIS

Projet ASAP

BnF Bibliothèque nationale de France

David Picard, Philippe-Henri Gosselin, Marie-Claude Gaspard
« Challenges in Content-Based Image Indexing of Cultural Heritage Collections. » ***IEEE Signal Processing Magazine*****, Institute of Electrical and Electronics Engineers, 2015, 32 (4), pp.95 – 102**
Corpus : http://images.bnf.fr

2 Parcours dans les collections

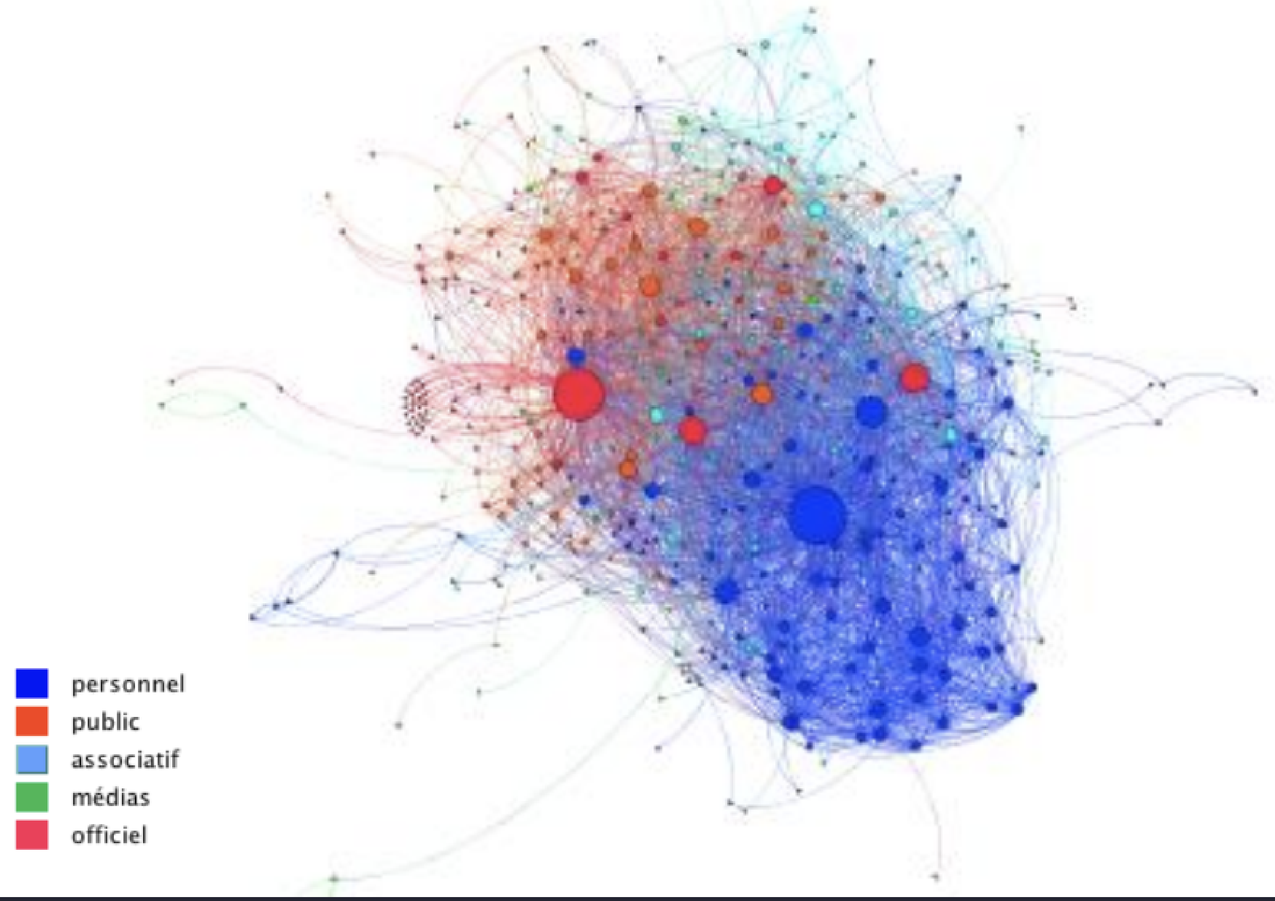
Esprit

J. Cook

explorer, décrire et analyser un terrain inconnu

A general chart of the Island of Newfoundland with the rocks & soundings
Thomas Jefferys, 1771
McMaster University Digital Archive :
https://digitalarchive.mcmaster.ca/islandora/object/macrepo%3A21591

BnF Bibliothèque nationale de France



Projet Labex **Les passés dans le présent**

https://hal.archives-ouvertes.fr/hal-01425600

Le devenir en ligne du patrimoine numérisé : l'exemple de la Grande Guerre

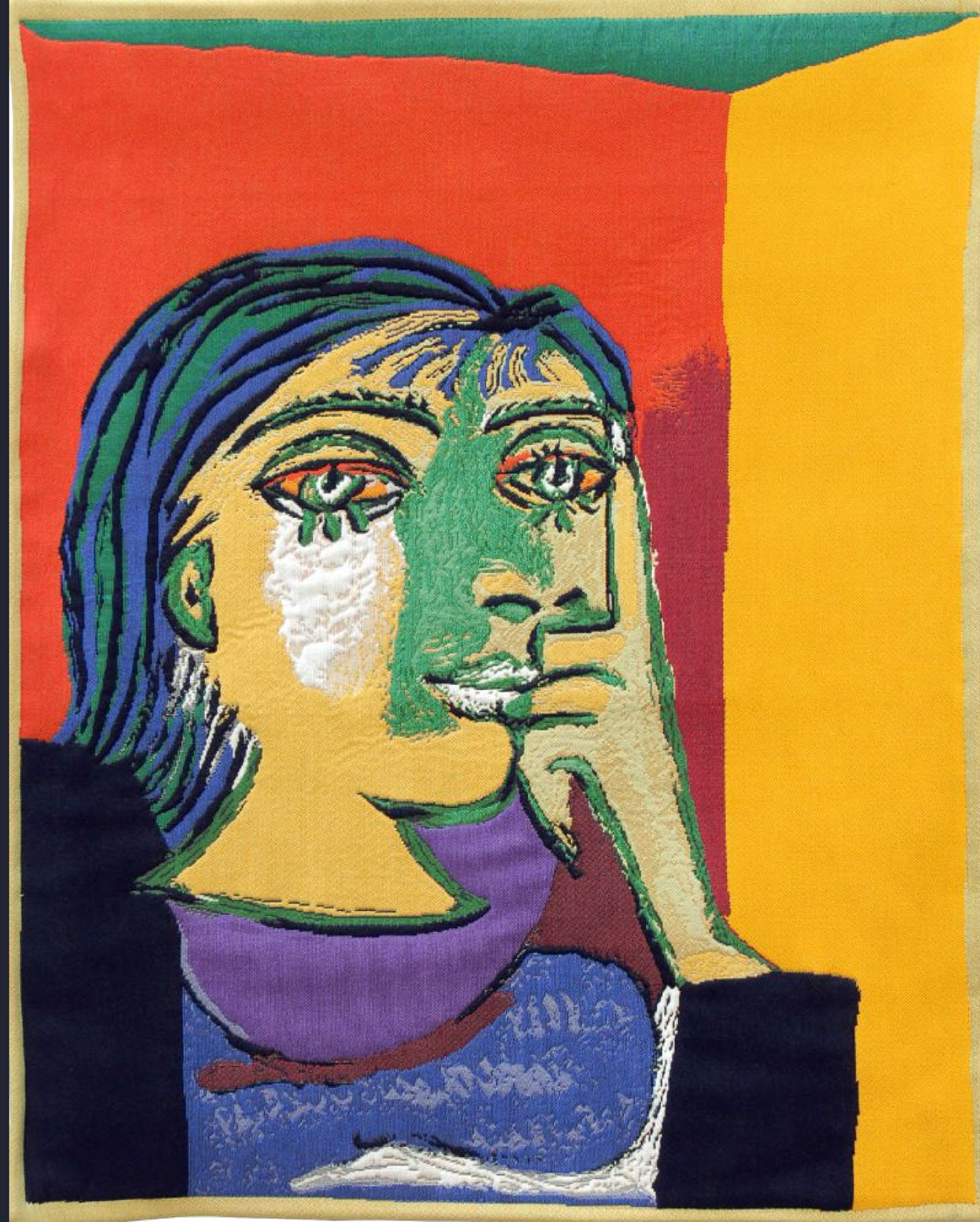
BnF Bibliothèque nationale de France

2 Parcours dans les collections

Esprit

Picasso

façonner une nouvelle forme de patrimoine



Portrait de Dora Maar
Pablo Picasso, 1937

BnF Bibliothèque nationale de France

Projet Gallica Studio

GallicaPix

Développé par Jean-Philippe Moreux (BnF) et Guillaume Chiron (L3i) un module de recherche d’images par le contenu qui fouille dans les collections d'images et d'imprimés de Gallica relatives à la Première Guerre Mondiale (période 1910-1920). Il repose sur différentes techniques d’intelligence artificielle (réseaux de neurones) et sur l’API IIIF pour l’affichage des images.

Site internet :
http://demo14-18.bnf.fr:8984

BnF Bibliothèque nationale de France

« 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

Le programme de recherche CORPUS de la BnF

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 (archives web, numérisation, métadonnées) + 1 année de bilan

4
ans

3
« collections » numérisées ou nativement numériques

2020
horizon

➔Fiche dans la base ANIR :
➔http://c.bnf.fr/fom

Archives de l’Internet

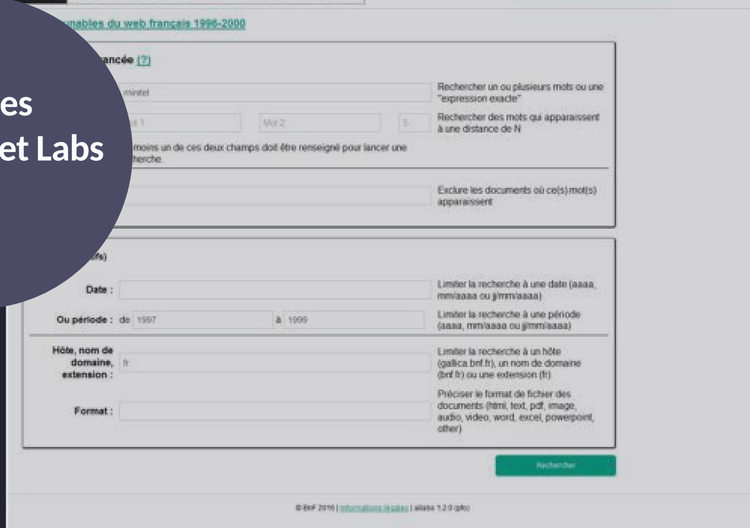
ANNÉE 1

« Il était une fois dans le web : 20 ans d’archives de l’Internet en France »
Journée d’étude co-organisée par la BnF, Vidéos :
http://c.bnf.fr/fse

Mise en ligne du blog Webcorpora
http://webcorpora.hypotheses.org/

Partenariat avec l’équipe Web90 (CNRS/ISCC)

Archives de l’internet Labs



BnF

Documents numérisés

ANNÉE 2

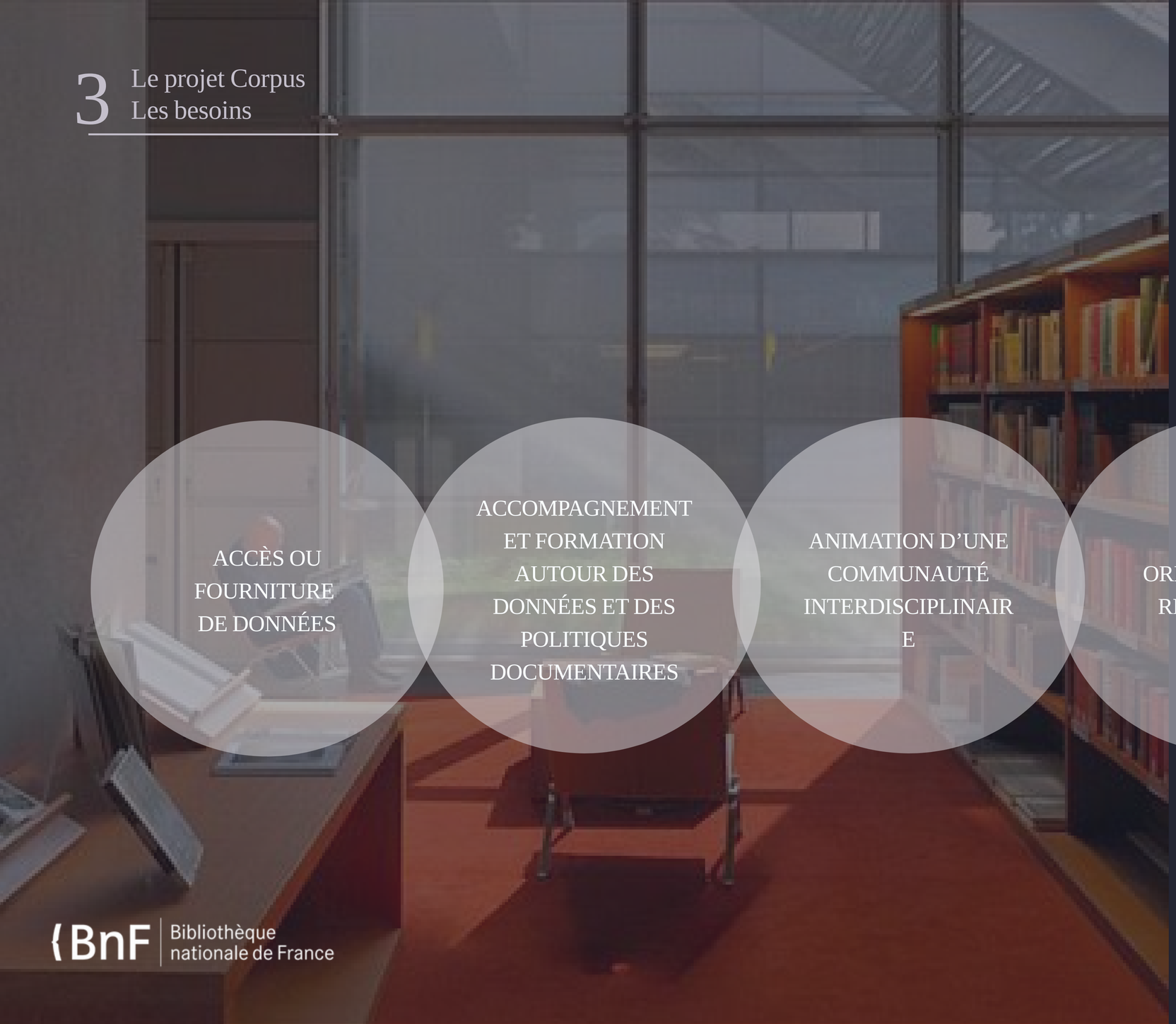
Étude de besoins
https://hal-bnf.archives-ouvertes.fr/hal-01739730v1

Site internet API et jeux de données
http://api.bnf.fr

Projet de recherche GIRANIUM

BnF

3 Le projet Corpus
Les besoins



ACCÈS OU FOURNITURE DE DONNÉES

ACCOMPAGNEMENT ET FORMATION AUTOUR DES DONNÉES ET DES POLITIQUES DOCUMENTAIRES

ANIMATION D’UNE COMMUNAUTÉ INTERDISCIPLINAIRE

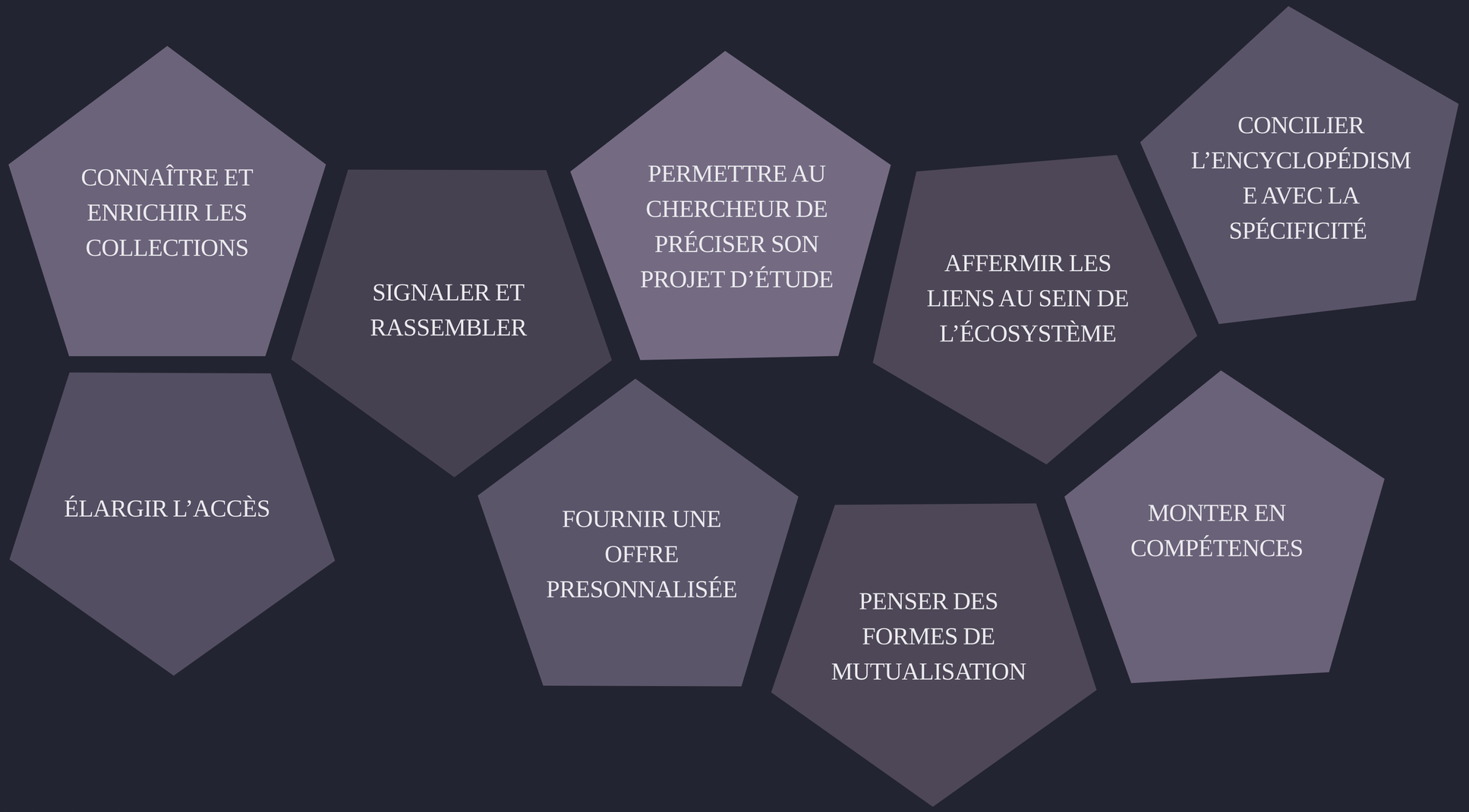
ORIENTATION ET REDIRECTION

VALORISATION ET RÉINTÉGRATION D’OUTILS OU RÉSULTATS

BnF Bibliothèque nationale de France

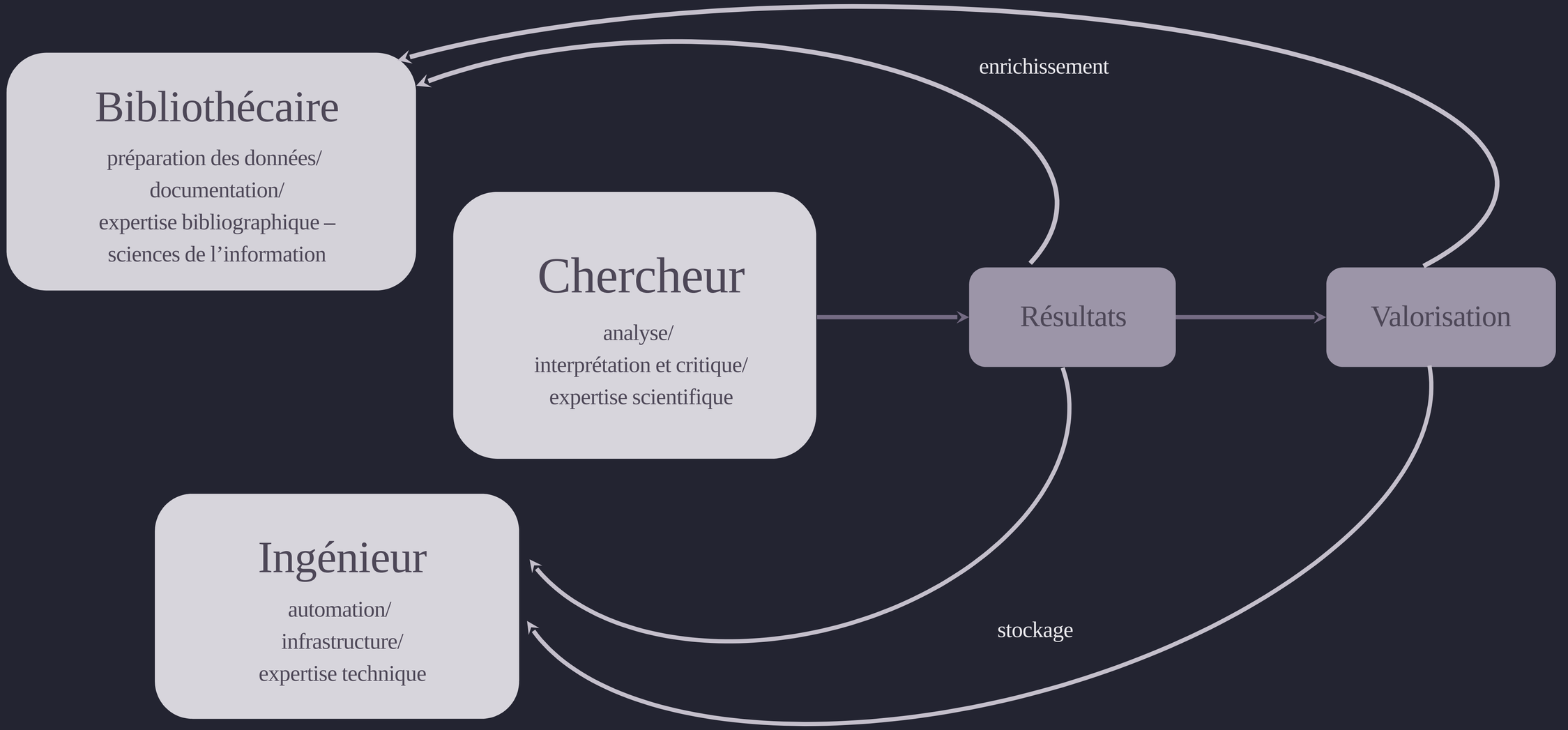

Espace pour événements
pour satisfaire le besoin de
PARTAGE
FORMATION MUTUELLE
DÉBAT
INTERDISCIPLINARITÉ
STANDARDISATION
CONVERGENCE
Bureaux
pour satisfaire le besoin de
PROGRAMMES EN RÉSIDENCE
LOCAUX
VISIBILITÉ
Espace de présentation
pour satisfaire le besoin de
PARTAGE
FORMATION MUTUELLE
INSPIRATION
Loges
pour satisfaire le besoin de
TRAVAIL INDIVIDUEL
LOCAUX
Collections
Numérisées ou nativement numériques
Salles de groupes
pour satisfaire le besoin de
TRAVAIL COLLECTIF
LOCAUX
FORMATION
Espaces du personnel
pour satisfaire le besoin de
ACCOMPAGNEMENT PERMANENT
SERVICE PUBLIC
RENDEZ-VOUS
Zone détente
pour satisfaire le besoin de
CONVIVIALITÉ
ÉCHANGE
Infrastructure numérique
pour satisfaire le besoin de
ACCÈS À DISTANCE
MOBILITÉ
CONTRÔLE FLUX DE TRAVAIL
ORDINATEUR PERSONNEL

3 Le projet Corpus
Les enjeux



BnF Bibliothèque nationale de France

4 Comprendre et exploiter les données ensemble



AVEZ-VOUS DES QUESTIONS ?

BnF Bibliothèque nationale de France

AVEZ-VOUS DES QUESTIONS ?

MERCI DE VOTRE ATTENTION !

BnF | Bibliothèque nationale de France