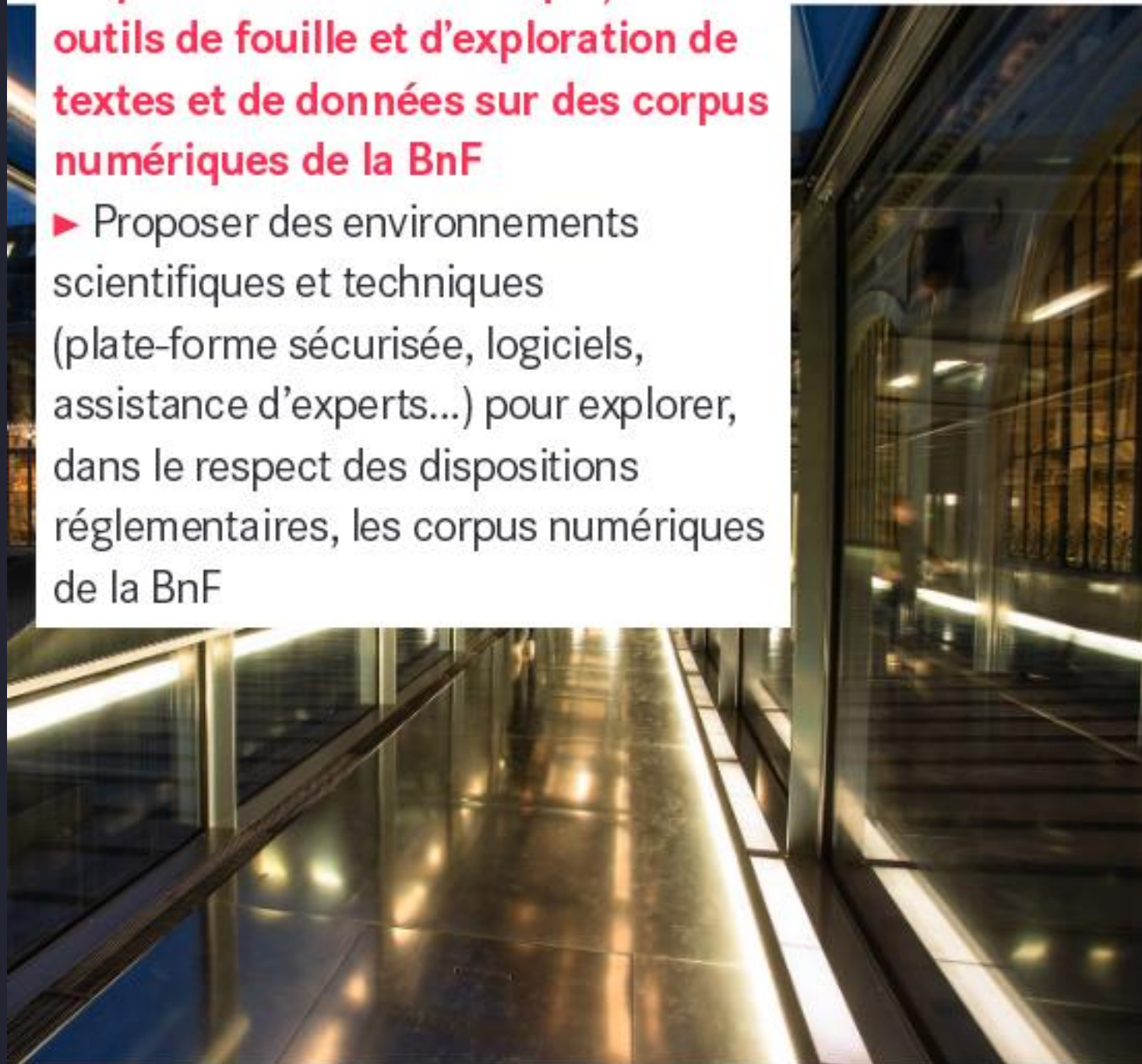

NATIONAL LIBRARIES AS SERVICE PROVIDERS FOR RESEARCH: BNF CASE STUDY

By Emmanuelle Bermès and Jean-Philippe Moreux

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

► Proposer des environnements scientifiques et techniques (plate-forme sécurisée, logiciels, assistance d'experts...) pour explorer, dans le respect des dispositions réglementaires, les corpus numériques de la BnF



CONTRAT D'OBJECTIFS
ET DE PERFORMANCE 2017-2021



{ BnF | Bibliothèque
nationale de France

Delivering BnF digital collections to researchers in order to allow them, on-site, to perform computational analysis of digital corpora »

Performance and strategic report 2017-2021

CORPUS, a BnF research project

Part of the BnF's 4-year internal research programme 2016-2019

OBJECTIVES :

- Designing a future service for providing access to digital corpora for researchers
- Providing researchers with data and tools to analyse them, in agreement with IPR and privacy legislation

3 years of experimentation (web archives, digitization, metadata) + 1 year for conclusions

4 years
3 digitized
and born-digital
collections
2020 by
→ <http://c.bnf.fr/fom>

Founding myths

ENDLESS
TECHNICAL
SOLUTIONS

PERFECT
GRANULARITY

COMPREHENSIVE
CORPORA

COMPREHENSIVE
DOCUMENTATION

FULL
RESEARCHERS'
AUTONOMY

TOTAL
INTEROPERABILITY

TRACKING,
REVERSIBILITY

ALGORITHMS'
PERFECTION

CONTINUOUS
TIME

LEGAL
FREEDOM

UNFAILING
ACCESS
CONTROL
SYSTEMS

EVERYTHING
VISUAL

The Gargantua way

inexhaustible thirst for data,
work on huge corpora
continually enriched

« Et lui furent ordonnées dix et sept mille neuf cents treize vaches pour l'allaiter ordinairement »

Gustave Doré, 1854

gallica.bnf.fr / Bibliothèque nationale de France





OBVIL

“Très Grande Bibliothèque”
project

135,000
monographs
delivered
for text mining

Website:

<http://obvil.lip6.fr/tgb/>

<http://api.bnf.fr/documents-de-gallica-produits-au-format-tei-par-obvil>

13 051 results for the following query:

L'Encyclopédie

TGB

Author "Diderot" x **author = Diderot**

None

1

L'Encyclopédie

Daubenton & Diderot • BECASSE, scolopax • Histoire naturelle | Ornithologie

On peut aussi les prendre aux lacets dans les bois, ou le long des ruisseaux ; ces lacets n'ont rien de particulier. Les bécasses se mangent roties, sans être vidées: quand on en veut faire un ragoût, on ne les laisse cuire à la broche qu'à moitié; on les dépece ; on les met dans une casserole avec du vin, des champignons, du sel & du poivre, & on les laisse bouillir jusqu'à ce que la cuisson soit achevée. Le salmi se fait presque de la même manière ; on ajoute seulement des truffes & des anchois, & on lie la fausse avec le foie & les entrailles de la bécasse. La bécasse considérée comme aliment, passe pour être nourrissante, restaurante

[View passage in context](#)

TGB

Liger, Louis (1658-1717) • OEconomie générale de la campagne, ou Nouvelle maison rustique — Tome 2 • 1700

; après qu'elles font rôties, on y fait une fauce avec orange, sel, poivre blanc, & une ciboulette. Beccajfes en ragoût. Pour faire un bon ragoût de beccasses, on s'en fert lorsqu'elles sont à demy cuites, puis on les coupe en pieces, pour les mettre dans une casserole avec du vin, des champignons, le tout assaisonné de sel & de poivre, après quoy on les laisse bouillir jusqu'à ce que la cuisson foit parfaite, pour être servies après y avoir pressé un jus d'orange. Des Beccajfnes. On apprête les beccassines de la même manière que les beccasses lorsqu'on les veut faire rôtir ; mais quand on souhaite les manger en ragoût, on les fend en deux

[View passage in context](#)

[Show differences](#)

Browse by Metadata Counts

Show options

article topics x

Frequency by L'Encyclopédie Generatedclass

Showing top 100 results

Histoire naturelle	2 166
Histoire de la philosophie	1 381
Grammaire	1 183
Art mécanique	708
Métaphysique	601
Histoire ancienne	550
Histoire ecclésiastique	390
Morale	369
Géographie	350
Commerce	349

2

L'Encyclopédie

Daubenton & Diderot • BEC-FIGUE, ficedula • Histoire naturelle | Ornithologie

est excellent à manger : il se nourrit de figues, de raisin, &c. Voyez Oiseau. (I) * Pour l'apprêter, on le plume ; on lui coupe la tête & les piès ; on le rotit à la broche : à mesure qu'il cuit on le saupoudre de croûte de pain rapée & mêlée de sel, & on le mange au verjus de grain & au poivre blanc .

TGB

Liger, Louis (1658-1717) • OEconomie générale de la campagne, ou Nouvelle maison rustique — Tome 2 • 1700

pour être cfrviesavec une fauce assaisonnée de vinaigre" de sel & de poivre blanc. Des Becfigues, Pour apprêter des becfigues, on se contente de es plumer, & de leur couper la tête ôc les pieds, & à mesure qu'ils cuisent à la broche , on les saupoudre de croûte de pain râpé avec du sel, puis, en les mange au verjus de grain & au poivre blanc .. Du Sanglier. Il n'y a gueres de difference entre les apprêts qu'on fait du sanglier & ceux du porc commun ; j & s'il y en a, les yoicy

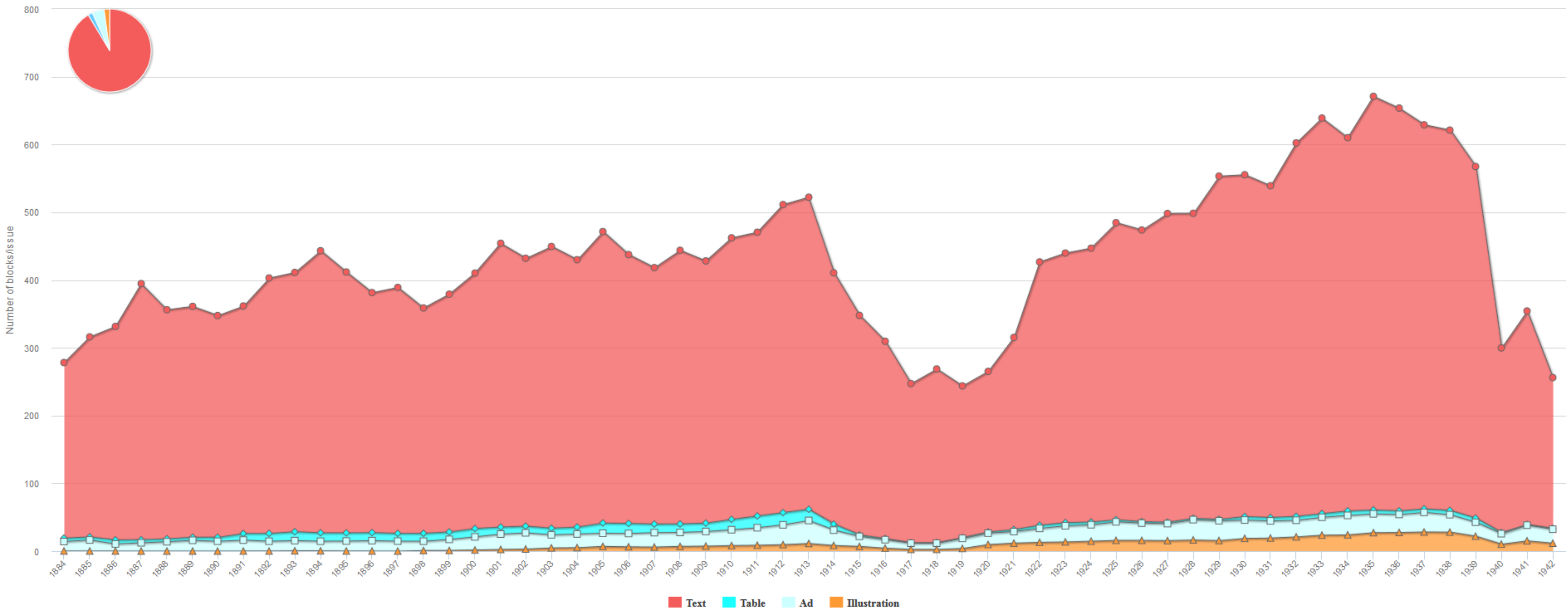
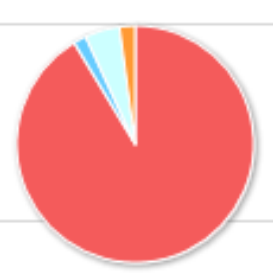
Reuse of L'Encyclopédie within the TGB

Alignments made with PhiloLine:
http://anomander.uchicago.edu/text-align/encyc_vs_tgb

- 310k fragments aligned
- Cross-faceted search on both corpora
- Time series chart

OBVIL, ARTFL (Univ. of Chicago) “Use & Reuse” project

- Investigate reuse models during the XVIIIth c.
- Understand the XVIIIth c. heritage through the reuse during XIXth and XXth c.
- Thanks to sequence alignment algorithms (bioinformatics)
- Output : closed and distant readings analysis, reuse fragments dataset



**Le Matin,
1884-1942**

The impact of wars on the newspaper industry

Europeana Newspapers project

10 M OCRed pages from 12 European Libraries

Data Mining the Metadata

- 900,000 ALTO files used for data mining French dailies
- 6 M metadata dataset for DH

Website:

http://altomator.github.io/EN-data_mining/

<http://www.europeana-newspapers.eu/>

The Arcimboldo way

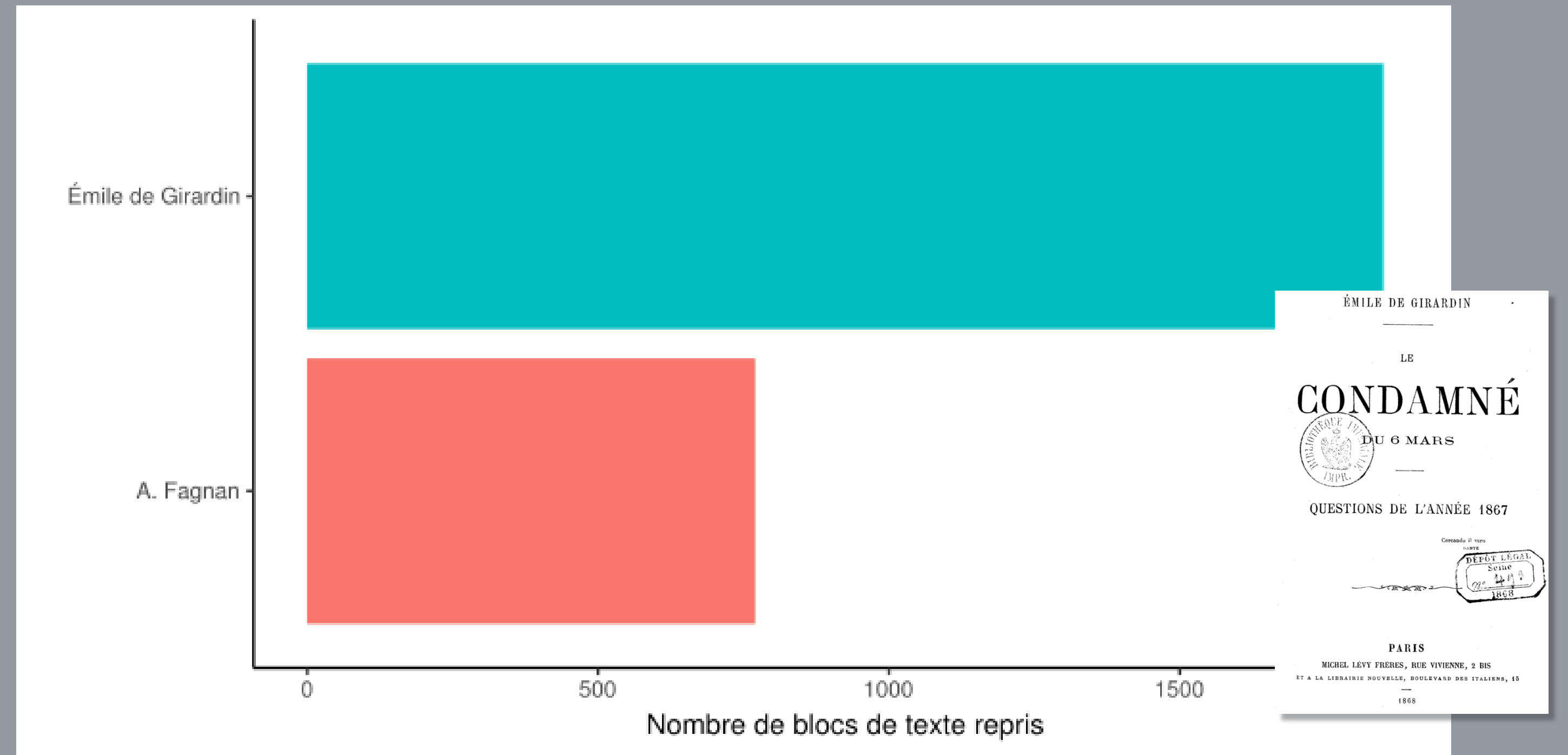
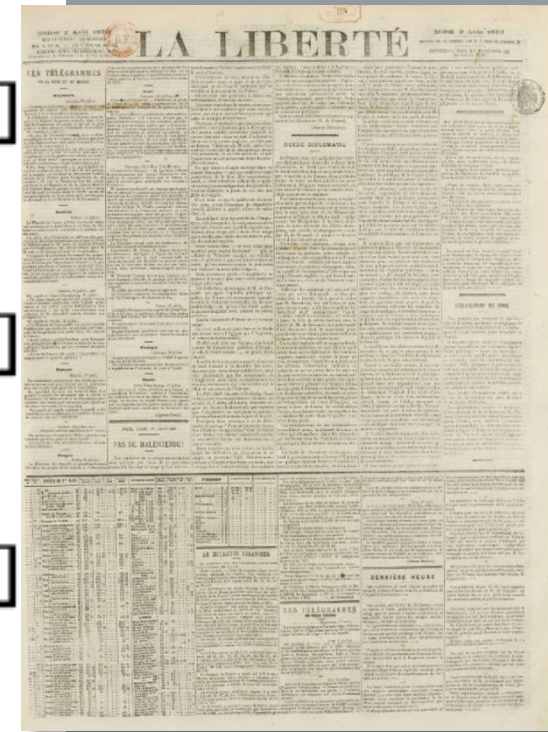
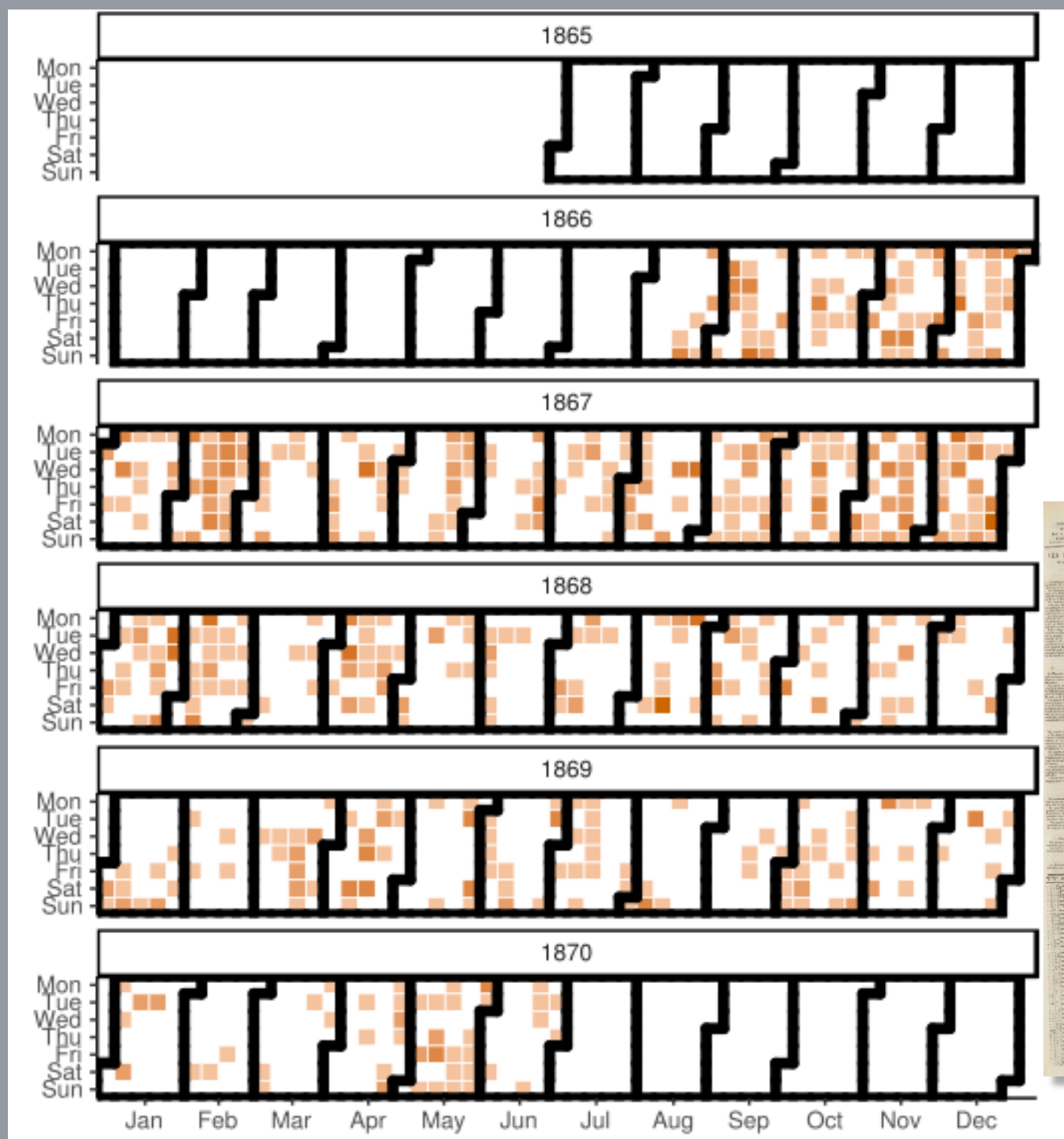
gathering and studying scattered
and/or heterogeneous sources

Vertumne

Giuseppe Arcimboldo, 1590

Château de Skokloster





Who is « A. Fagnan », a regular journalist of *Liberté* newspaper, unknown in authorities files?

The third of Girardin's memoirs (1867) are reprints of *Liberté*. Fagnan=Girardin!

GRIPIC/ CELSA

The Giranium project

Analysis of 300 autograph letters from Émile de Girardin, and other documents (notes, newspapers, works, diaries, maps...)

Website: <http://vintagedata.org/giranium/>

The J. Cook way

exploring, describing and analysing
an uncharted territory

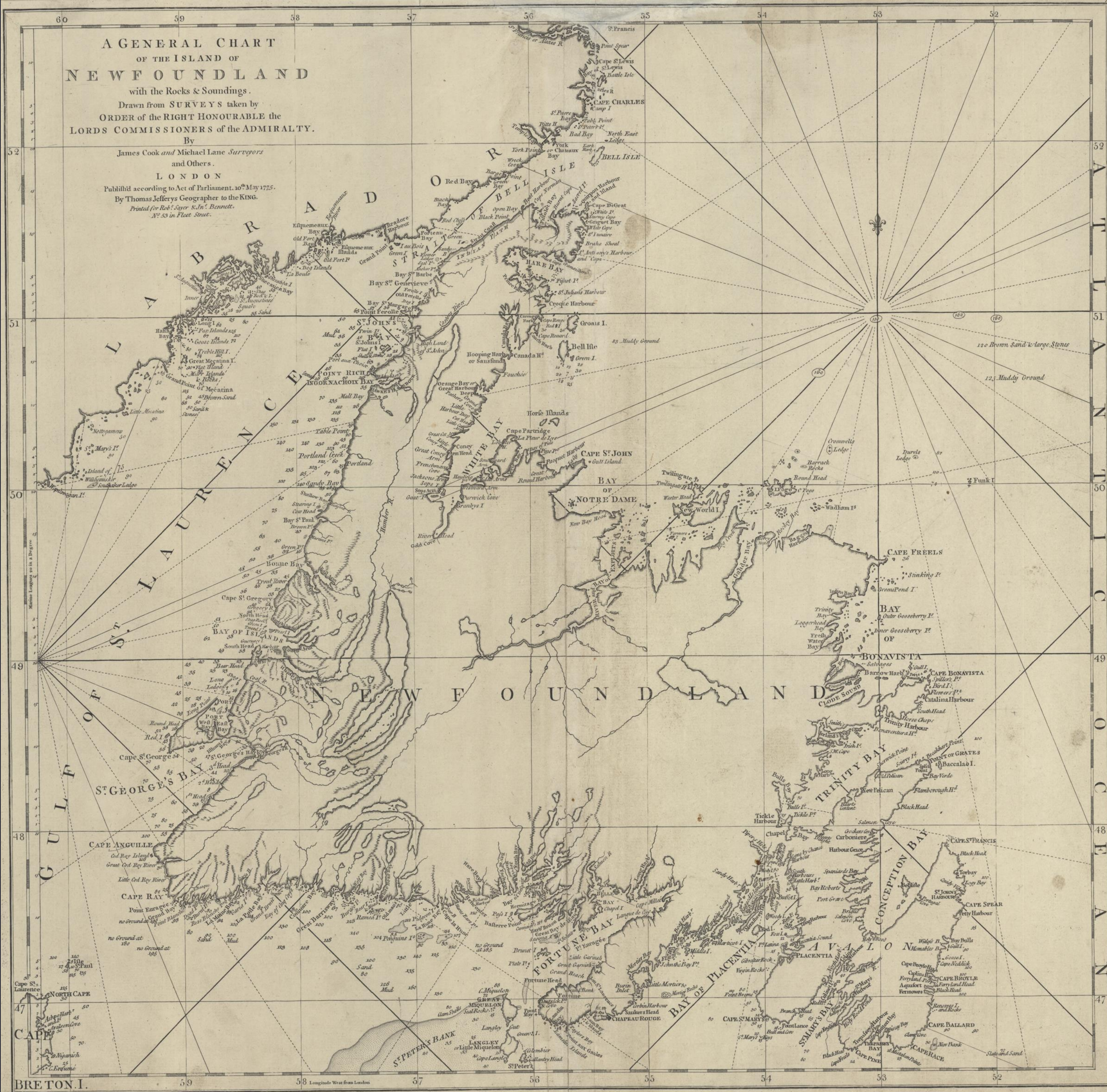
A general chart of the Island of Newfoundland with the rocks & soundings

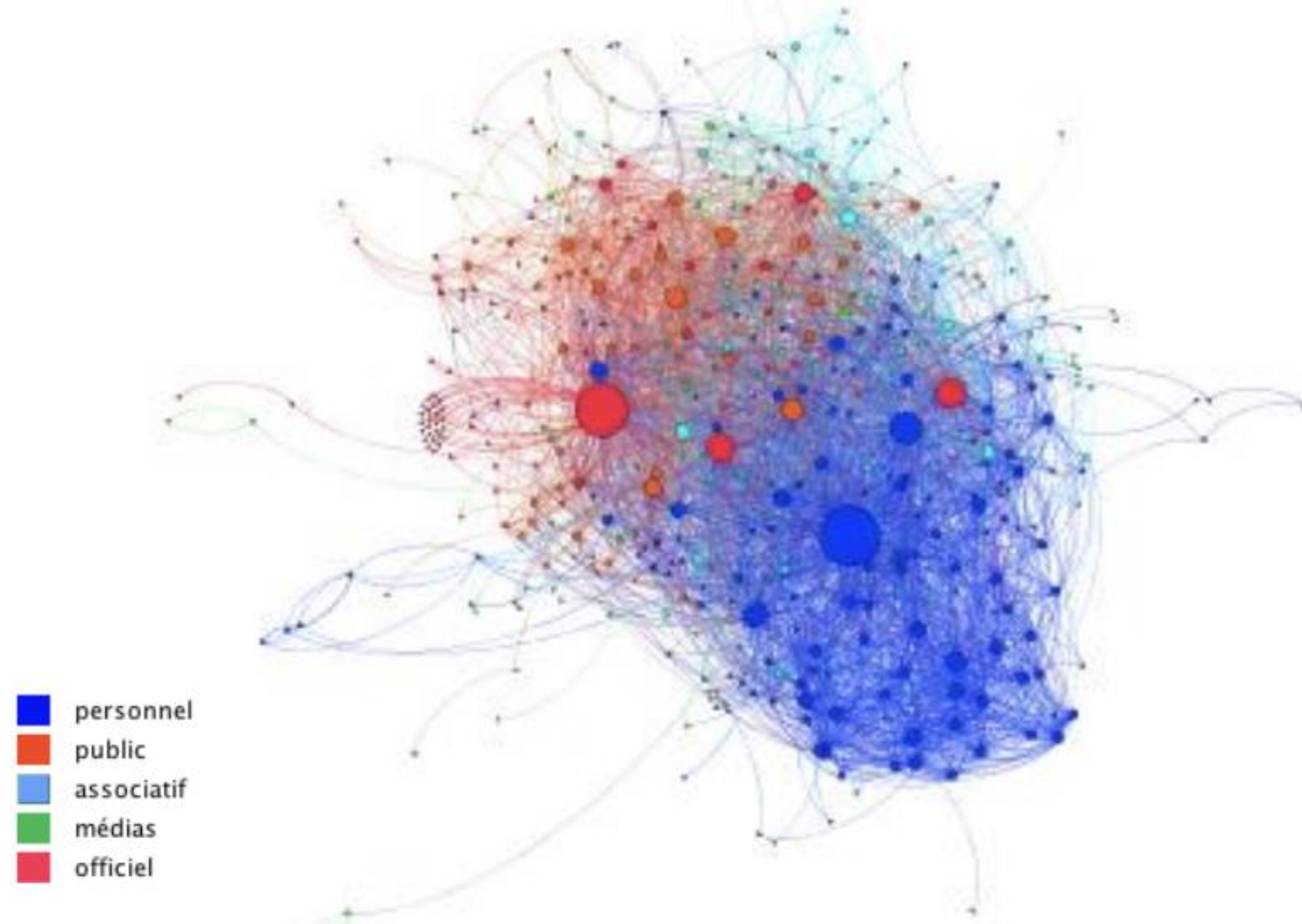
Thomas Jefferys, 1771

McMaster University Digital Archive :

<https://digitalarchive.mcmaster.ca/islandora/object/macrepo%3A21591>

BnF Bibliothèque nationale de France





Project by *Les passés dans le présent*

The future of online digitized heritage: the example of the World War I

Report:

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

The Picasso

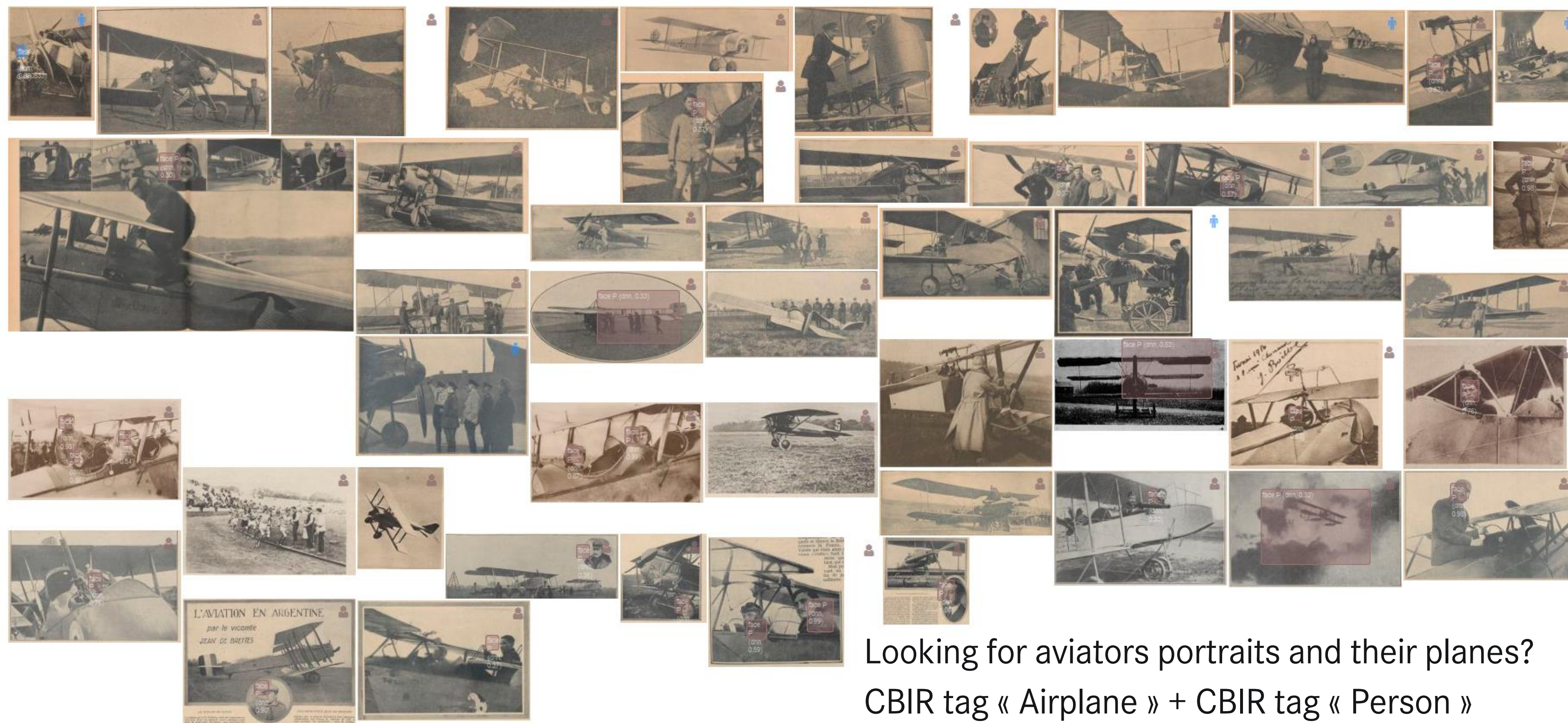
way

shaping new forms of heritage

Dora Maar portrait
Pablo Picasso, 1937
© Picasso estate/ Musée Picasso

{ BnF | Bibliothèque
nationale de France





- Data mining of 500k pages for illustrations
- Mixed collections (print, still image, maps...)
- Image genres classification with a custom CNN model
- Content-based image search (IBM Watson, Google Cloud Vision)
- Cross-modal query (text, metadata, CBIR)
- End-to-end access to data with BnF APIs (inc. IIIF)
- Outputs datasets for DH

Looking for aviators portraits and their planes?
CBIR tag « Airplane » + CBIR tag « Person »

GallicaPix Deep Learning applied to WW1 iconography

Demo:

<http://demo14-18.bnf.fr:8984/rest?run=findIllustrations-form.xq>

Documentation:

https://altomator.github.io/Image_Retrieval/



{ BnF

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

Report | January 2018

“The Corpus project and its potential users.”

A prospective study of users’ needs and expectations.”

Eleonora Moiraghi
Research assistant

<https://hal-bnf.archives-ouvertes.fr/hal-01739730>

What the BnF could do for researchers

...

DATA
ACCESS
AND DELIVERY

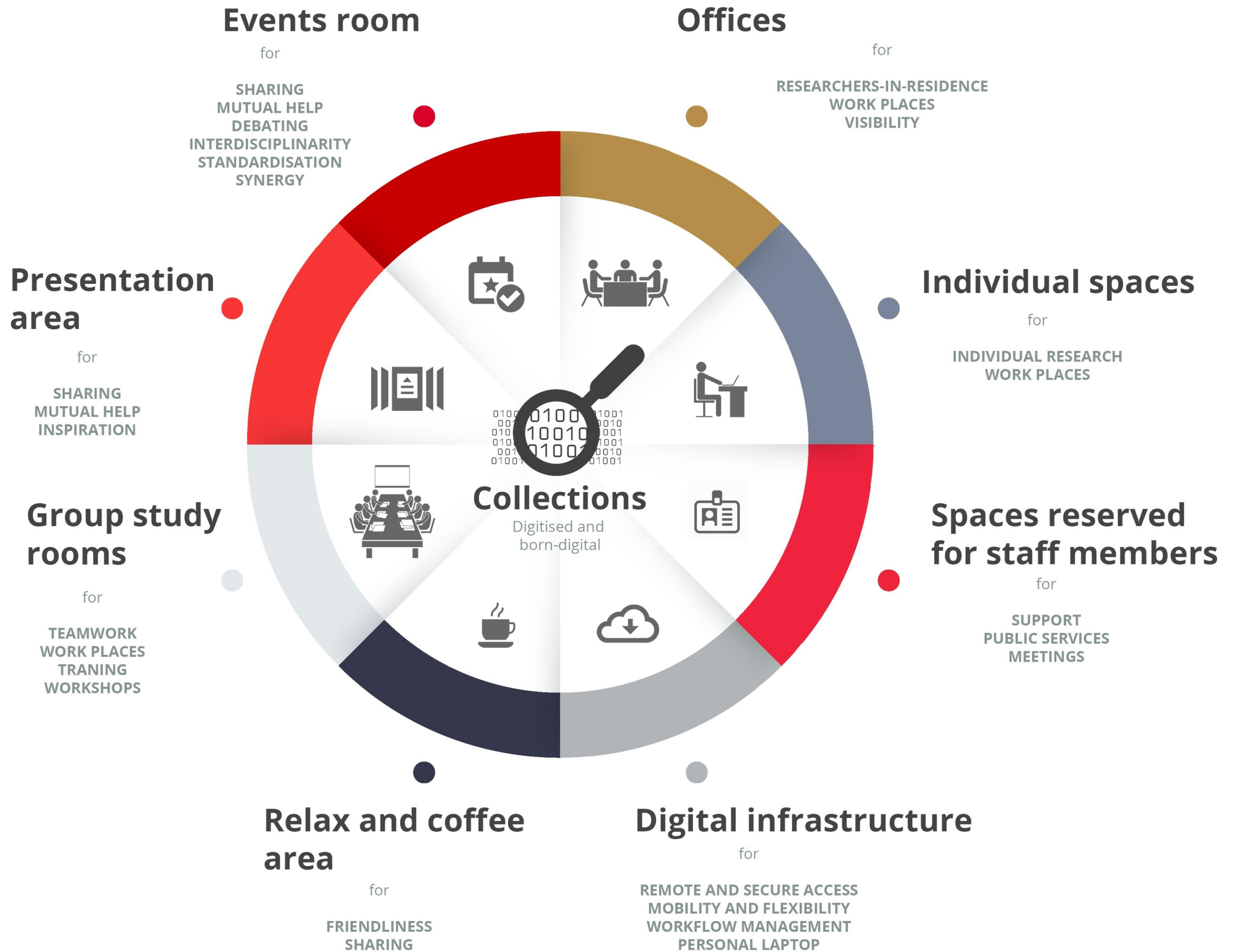
SUPPORT AND
TRAINING
ON DATA MODELS,
FORMATS AND
DOCUMENTARY
POLICY

ANIMATION OF A
CROSS-DISCIPLINARY
COMMUNITY

ADVISE AND
REDIRECTION

DISSEMINATION
AND INTEGRATION
OF TOOLS AND
RESEARCH
FINDINGS

Towards a Digital Humanities Lab



Online/ virtual services

{ BnF API et jeux de données

Découvrez et utilisez les données de la BnF



Le portail BnF API et jeux de données décrit et documente l'ensemble des **API qui permettent d'interroger et de récupérer les métadonnées des catalogues** (notamment BnF Catalogue général, data.bnf.fr, Gallica) et les **collections numérisées de la BnF**. Pour faciliter l'accès aux données et leur utilisation, **des jeux de données (images et textes, métadonnées, statistiques) sont directement téléchargeables** via le portail.

Plusieurs formats, plusieurs technologies permettent de répondre à la diversité des usages des données de la bibliothèque : alimentation de catalogues, création de nouveaux services innovants, fouille de données, datavisualisation, etc.

Développeurs et développeuses, chercheurs et chercheuses, acteurs et actrices du monde de la culture et de la chaîne du livre, *digital humanists*, ou encore amateurs et amatrices de culture, **les données et métadonnées diffusées par la BnF n'attendent plus que vous !**

[A propos de BnF API et jeux de données](#)

Actualités

Journée d'études - Quelles contributions de la BnF à la circulation des données ? rétrospectifs 2017 (20 mars 2018)



000 cam 22 450
001 FRBNF443013680000007
003 <http://catalogue.bnf.fr/ark:/12148/ch443013687>
010 .Sa 979-10-90367-18-0 Sb br. Sd 17,90 EUR
020 .Sa FR Sb 01536092
073 .0 Sa 9791090367180
100 .Sa 201503242015 m y0frey50 ba
101 0 .Sa fe
102 .Sa FR
105 .Sa lllz 00yi

Gallica Studio
autour des co

{ BnF Gallica Stu



{ BnF Archives de l'internet Labs

Accueil | Collections | Services | Aide

Rechercher

» Recherche avancée
» Recherche

» Se connecter

Archives de l'internet Labs permet d'explorer une partie des collections issues du dépôt légal de l'internet, constituées depuis 1996, en utilisant la recherche plein texte. Archives de l'internet Labs est complémentaire de l'application Archives de l'internet qui permet de faire des recherches par adresse URL et par date, de visualiser les pages archivées et de naviguer dans les archives du web français.

Des services complémentaires sont également proposés aux chercheurs dont les travaux sont réalisés sous couvert d'une convention. Trois collections sont actuellement disponibles.



Actualités 2010-2017



Attentats 2015



Incunables 1996-2000

Web archives Labs, an interface designed to meet researchers' needs

Website « API and datasets » created for the hackathon 2017 <http://api.bnf.fr/>

THANK YOU !