Travailler sur des corpus numériques : des collections aux chercheurs

Jean-Philippe Moreux
Département de la Coopération

BnF Bibliothèque nationale de France

© art Buck/Flickr

Plan

Interopérabilité

Coopération

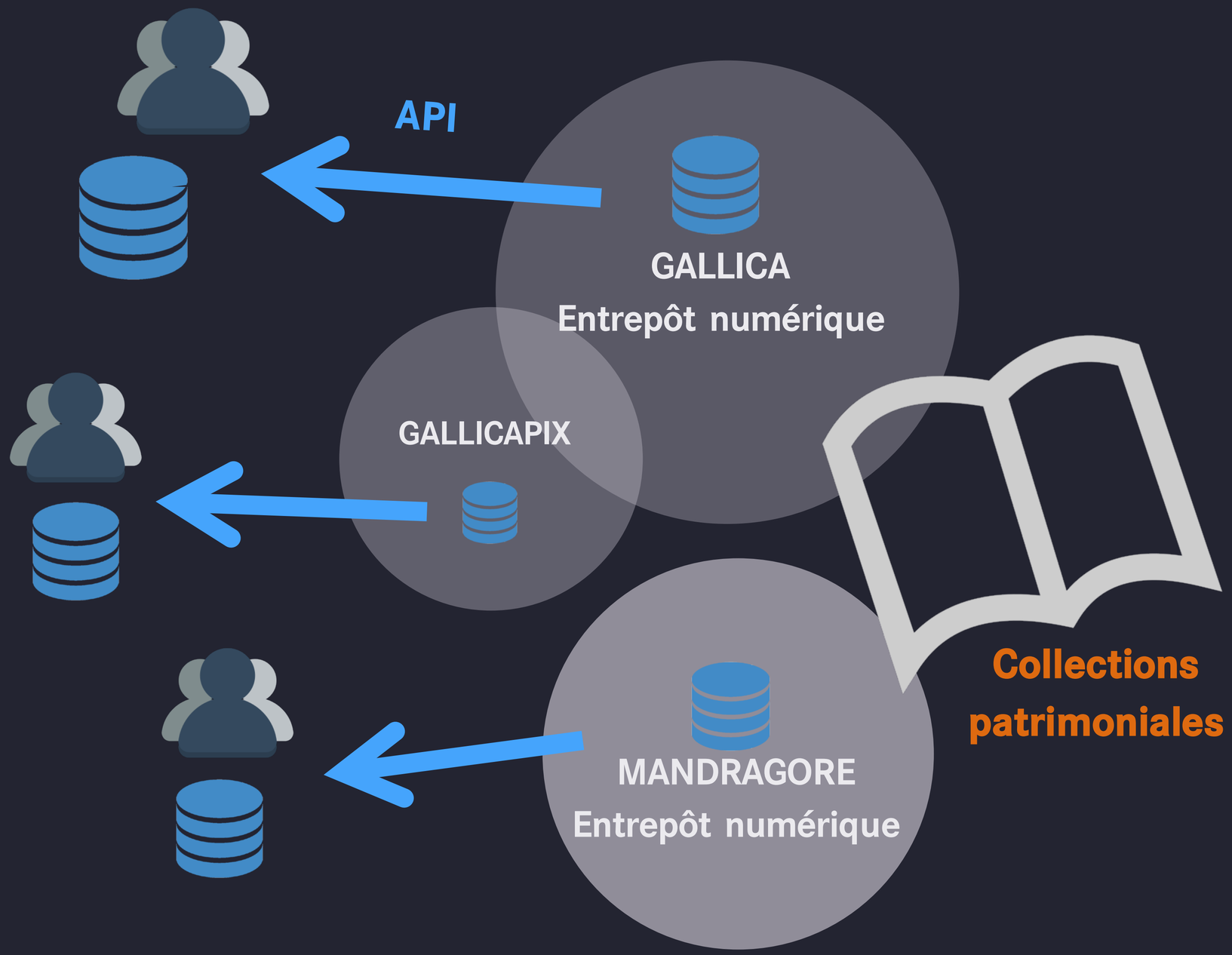
Accès aux collections sous droits

IA

Maintenant

Les chercheurs extraient et utilisent les contenus des collections (API, jeux de données…)

Partager les résultats (transcription, annotation) entre équipes, avec les institutions, est difficile



Demain

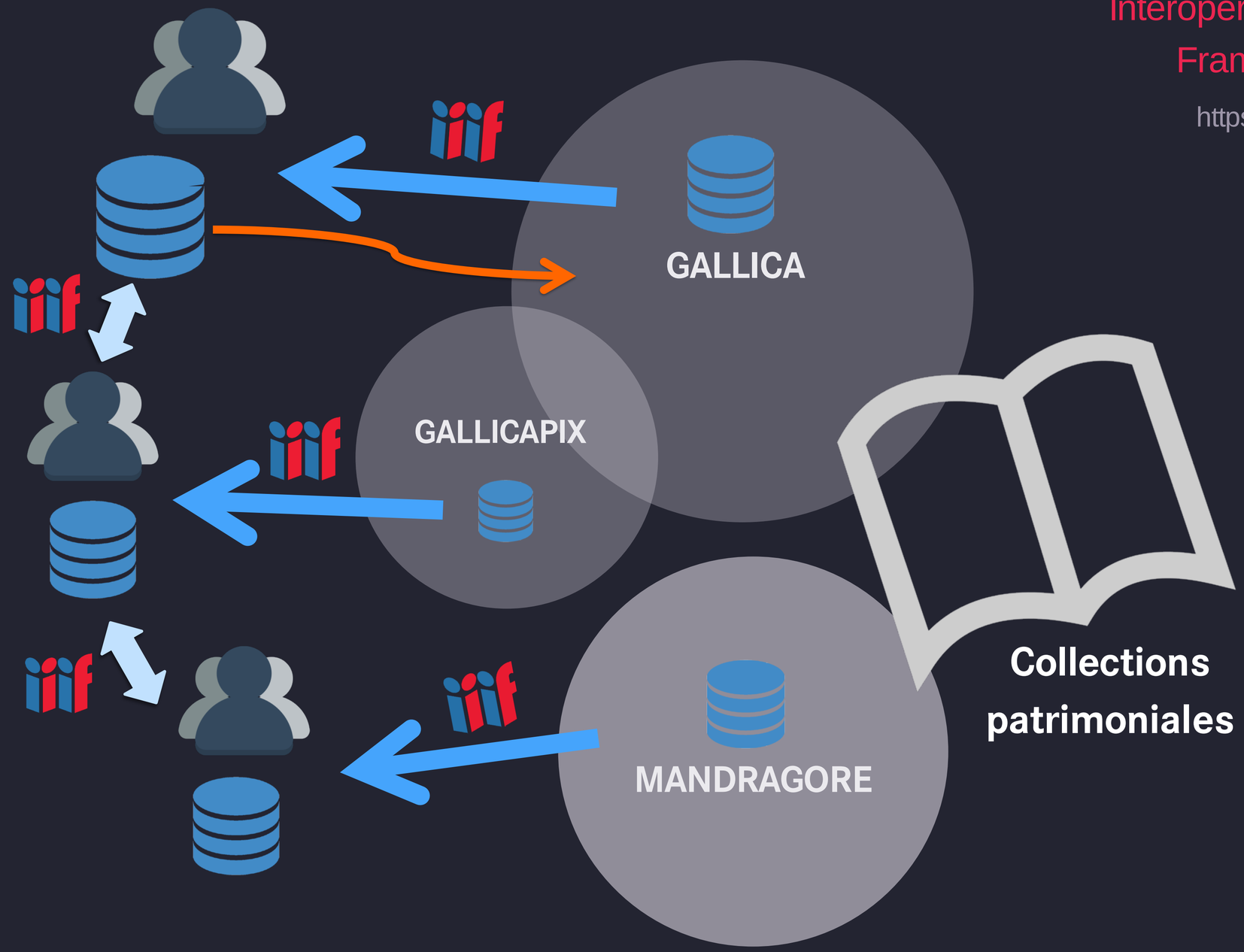
Les chercheurs partagent annotations et transcriptions avec IIIF

Les institutions peuvent exposer plus de métadonnées avec IIIF

Elles peuvent bénéficier du travail des chercheurs

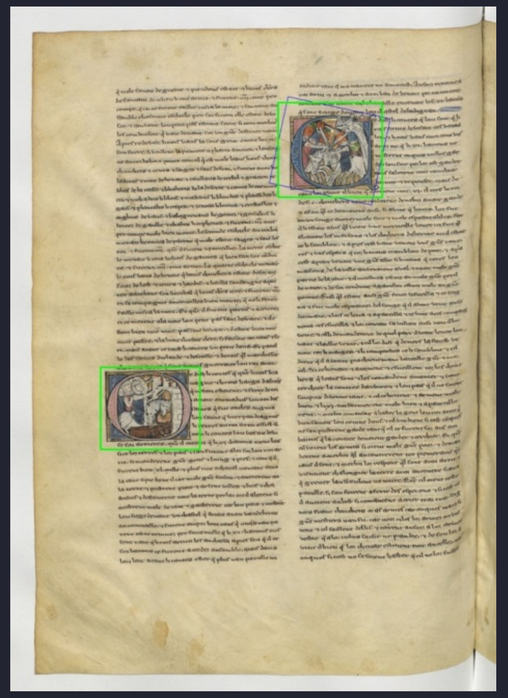
International Image Interoperability Framework

https://iiif.io

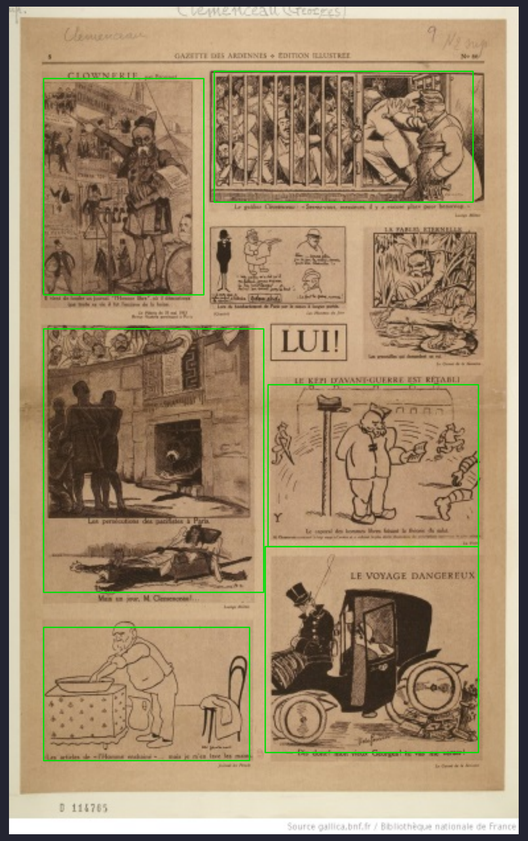


BnF Bibliothèque nationale de France

Quelles métadonnées ?



Manuscrits



Presse

Où sont les illustrations ? (segmentation)

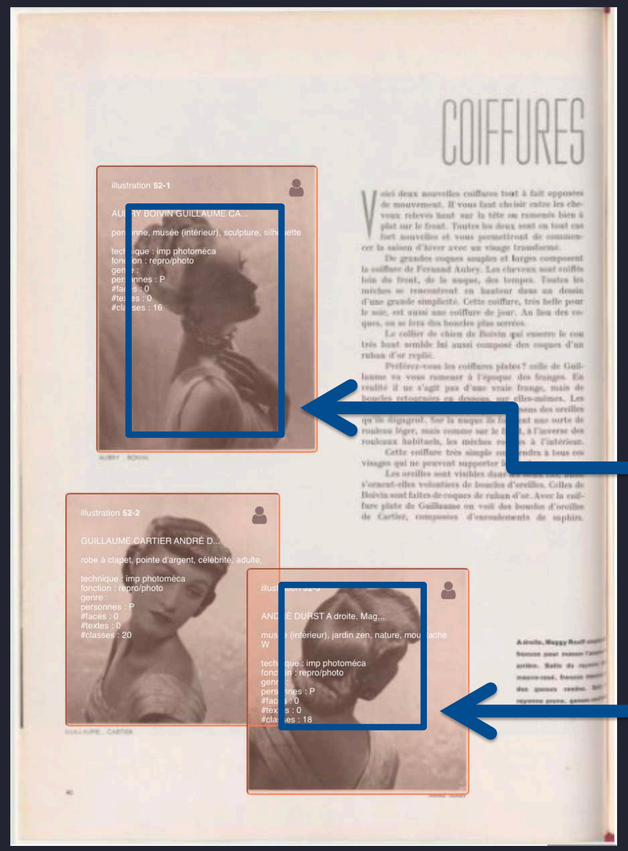
Magazines

x,y,w,h

Canvas, Liste d'annotations (Web Annotations W3C)

BnF Bibliothèque nationale de France

Quelles métadonnées ?



Vogue, 1936

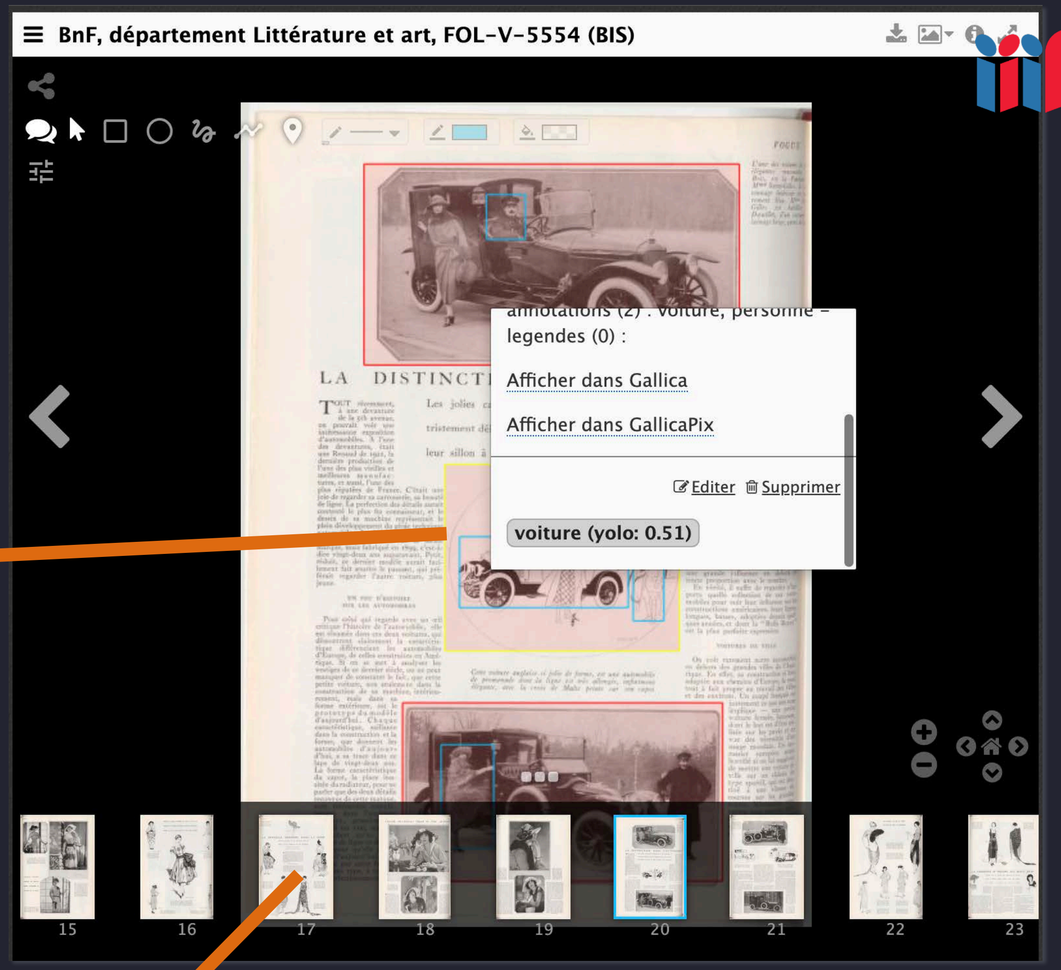
Quel est le contenu des illustrations ? (annotations, transcriptions) + travail de recherche : commentaires, analyse…

Canvas, Liste d’annotations

Cas d'usage : partage des annotations iconographiques

NewsEye : application web et serveur d'annotations IIIF

Document Gallica ouvert dans Mirador et enrichi avec des annotations NewsEye



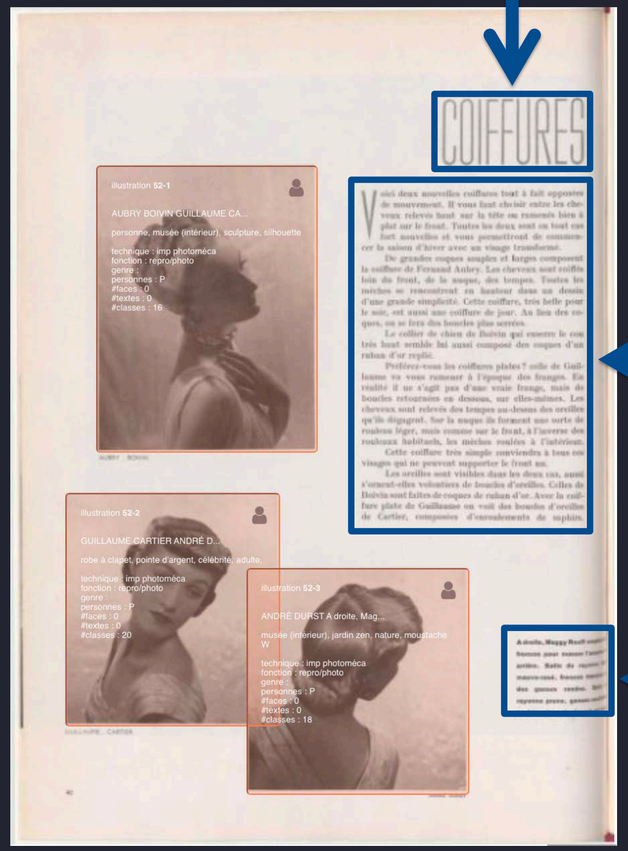
Serveur IIIF Gallica

Démo Vogue : https://api.bnf.fr/fr/node/191

Quelles métadonnées ?

Titre article (annotations)

Quel est le contenu textuel ? (OCR, type d'éléments, entités nommées, événements, …)



Vogue, 1936

Texte (annotations ou lien vers l'OCR)

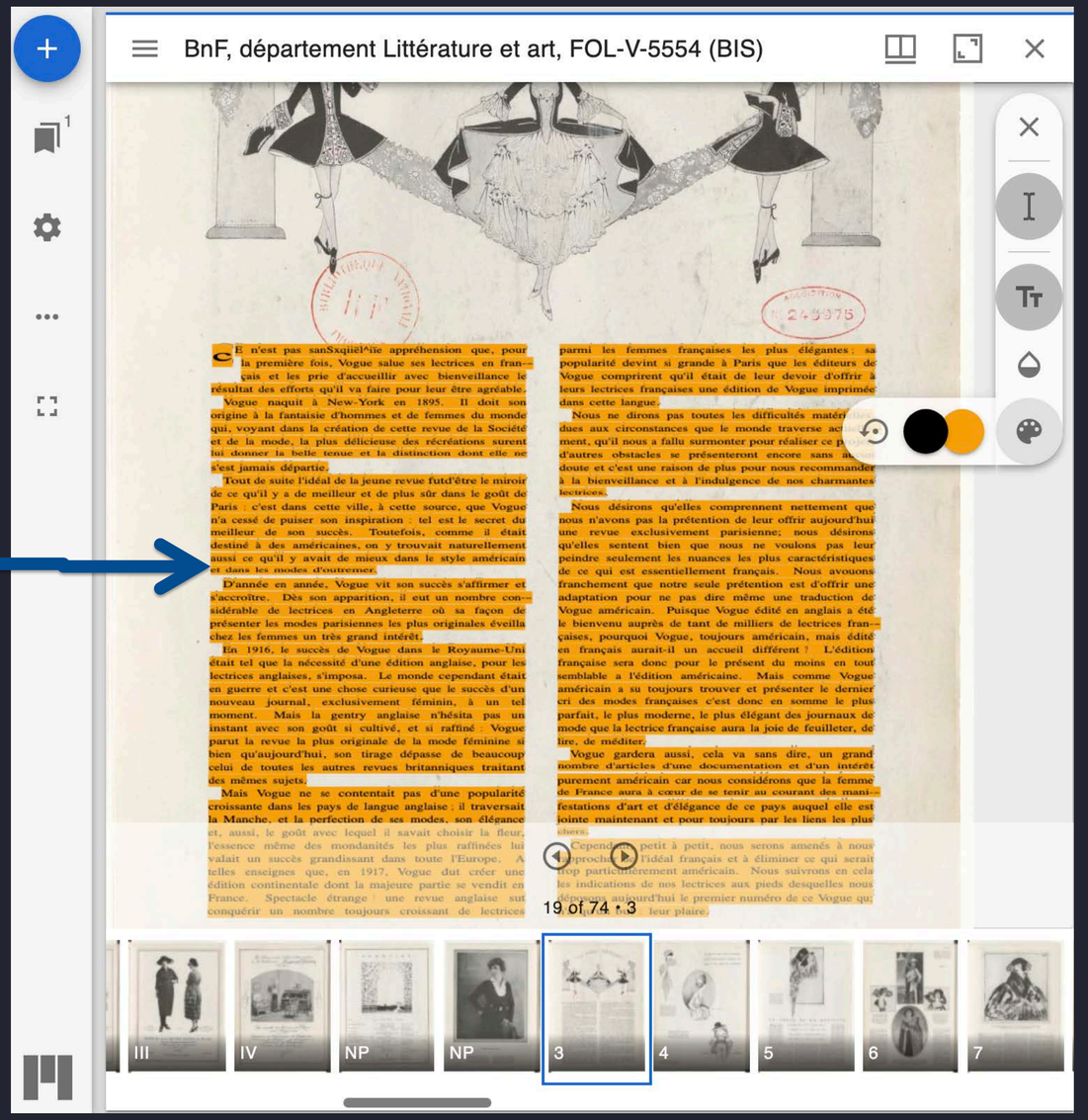
Texte

Canvas, Liste d'annotations, See Also

Cas d’usage : partage des textes et de leur enrichissement

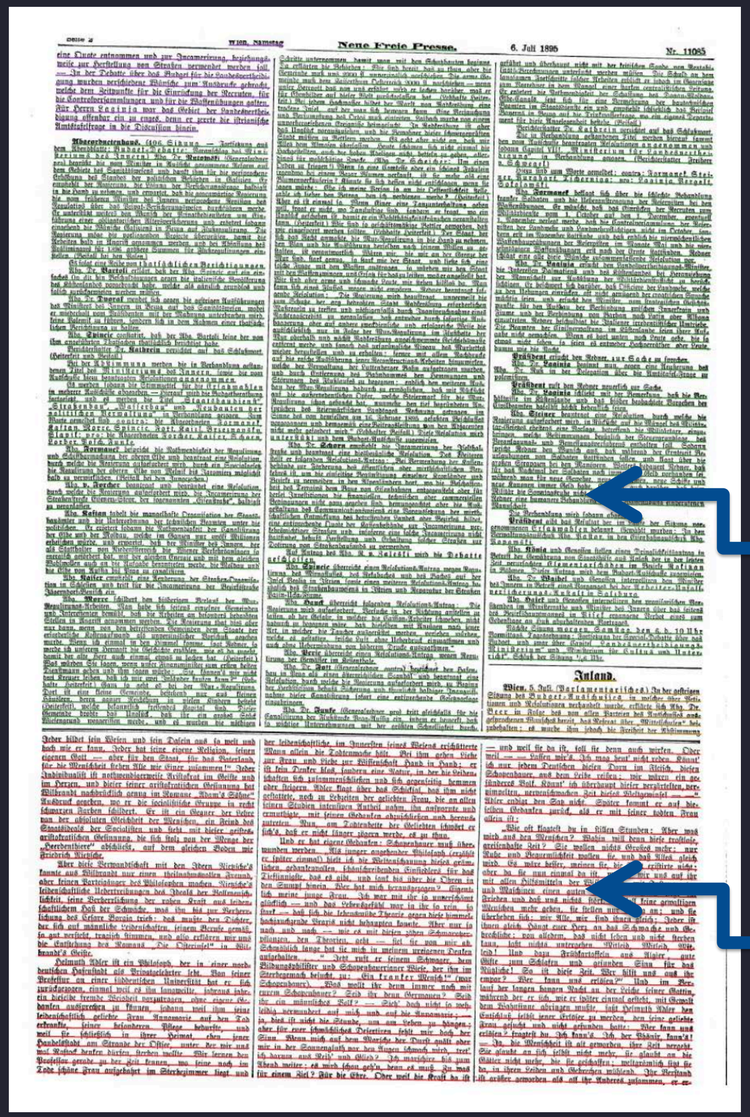
...
"seeAlso": {

"@id": "**https://platform.newseye.eu/IIIF/bpt6k9604118j/X/X00000019.xml**",
"profile": "http://www.loc.gov/standards/alto/ns-v4#",
"format": "text/xml"
}, ...



Démo Vogue

Quelles métadonnées ?

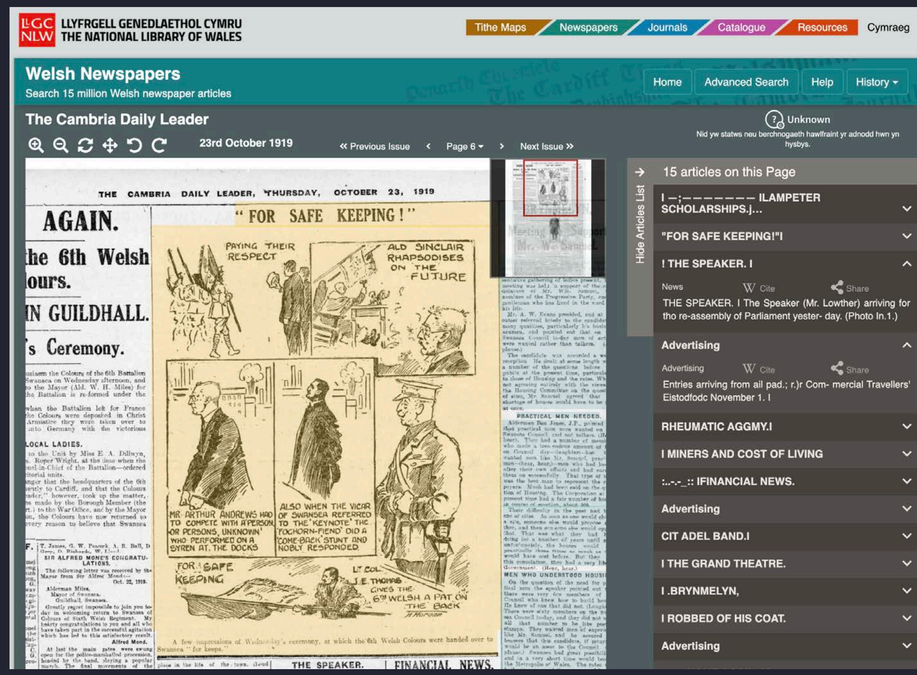

Neue Freie Presse. 6. Juli 1895 Nr. 11065

Article

Feuilleton

Quelle est la structure du document ?

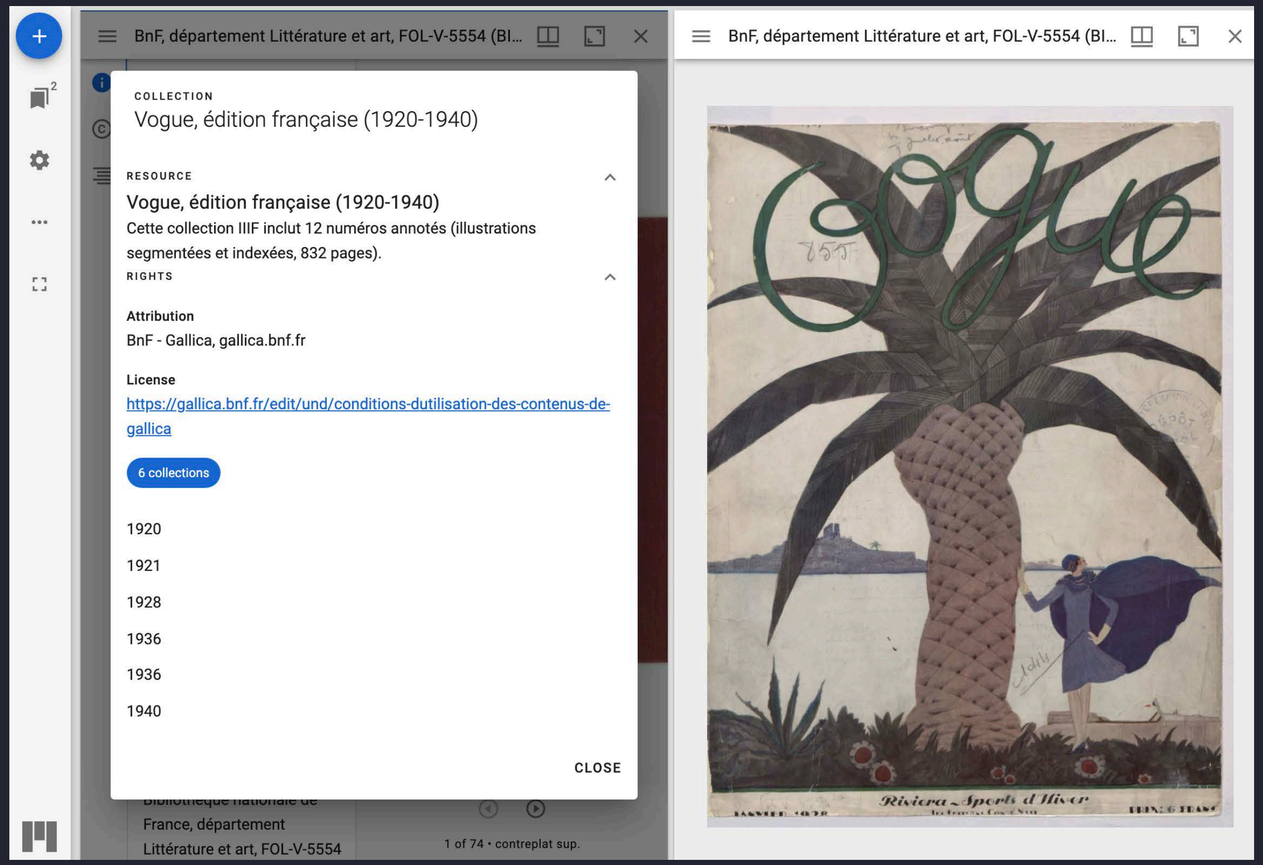
Bibliothèque nationale du pays de Galles



Et aussi

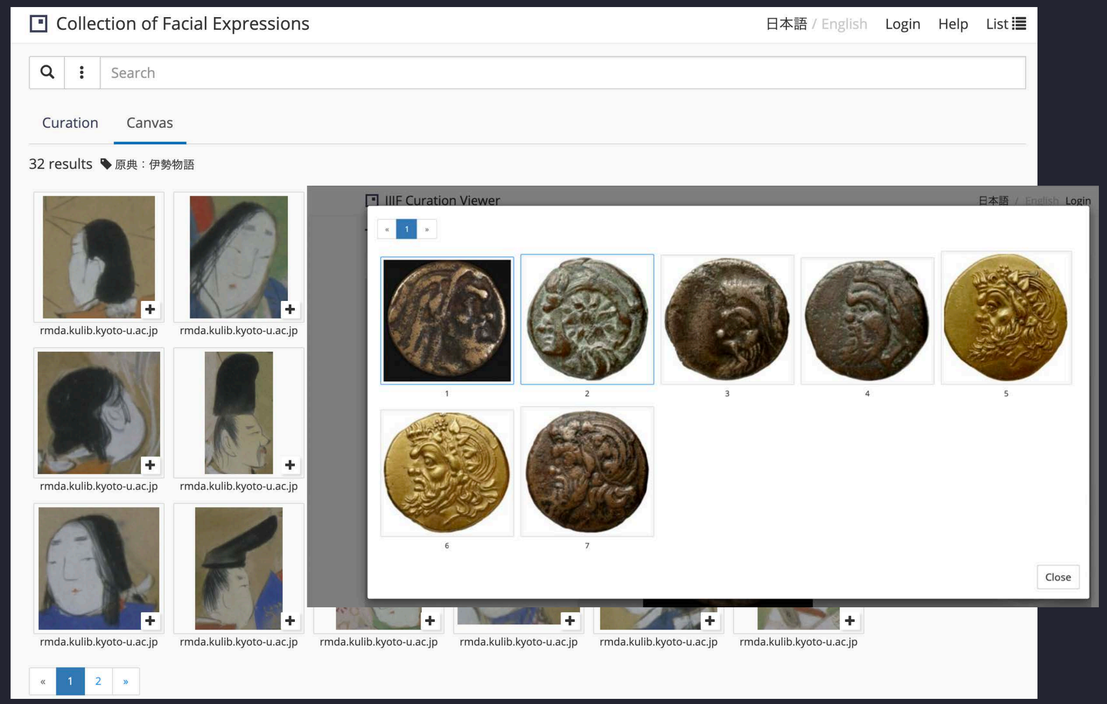
Collections
Curation
Editorialisation

Collection de documents (IIIF)



Démo Collection Vogue

Curation (extension IIIF)

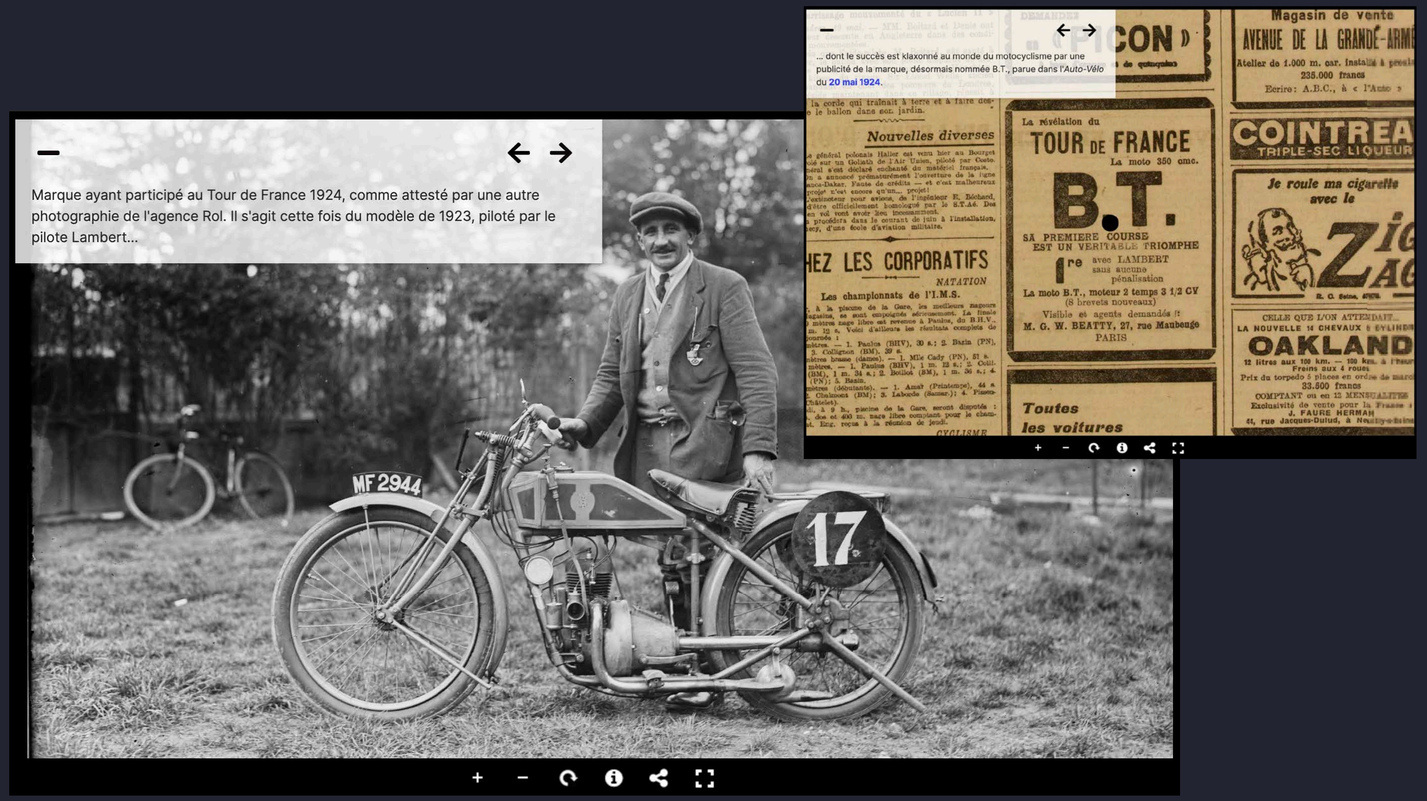


http://codh.rois.ac.jp/icp/index.html.en

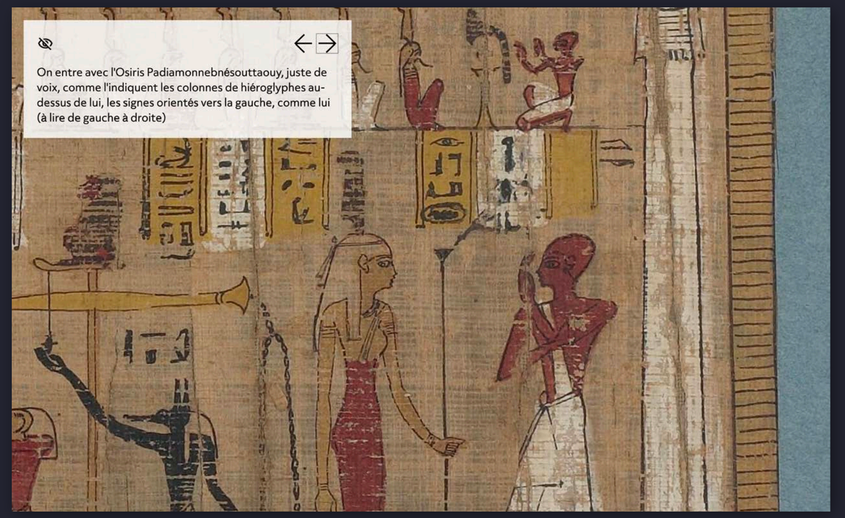
Et aussi

Storytelling

Collections
Curation
Editorialisation



Démo Exhibit



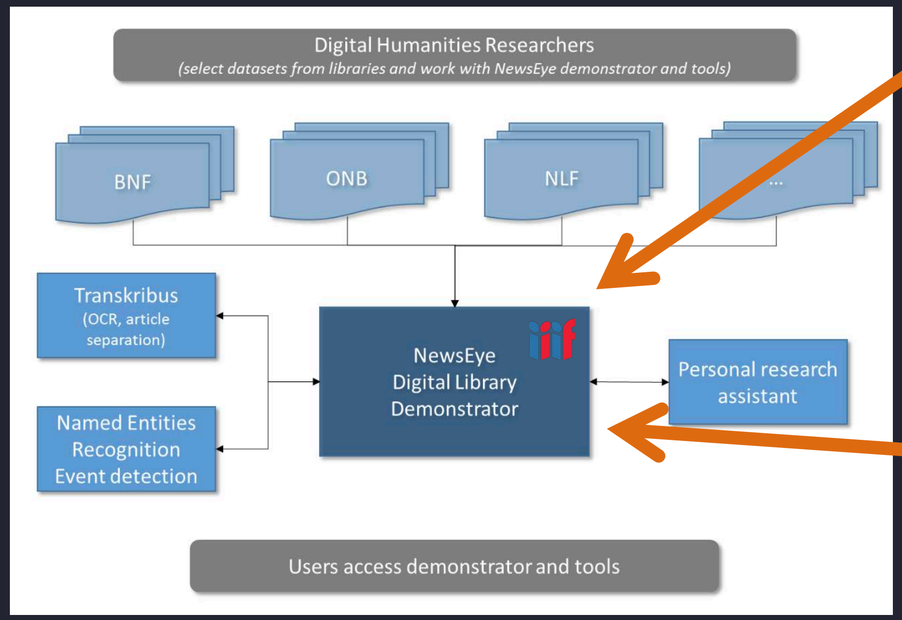
Démo Storiies

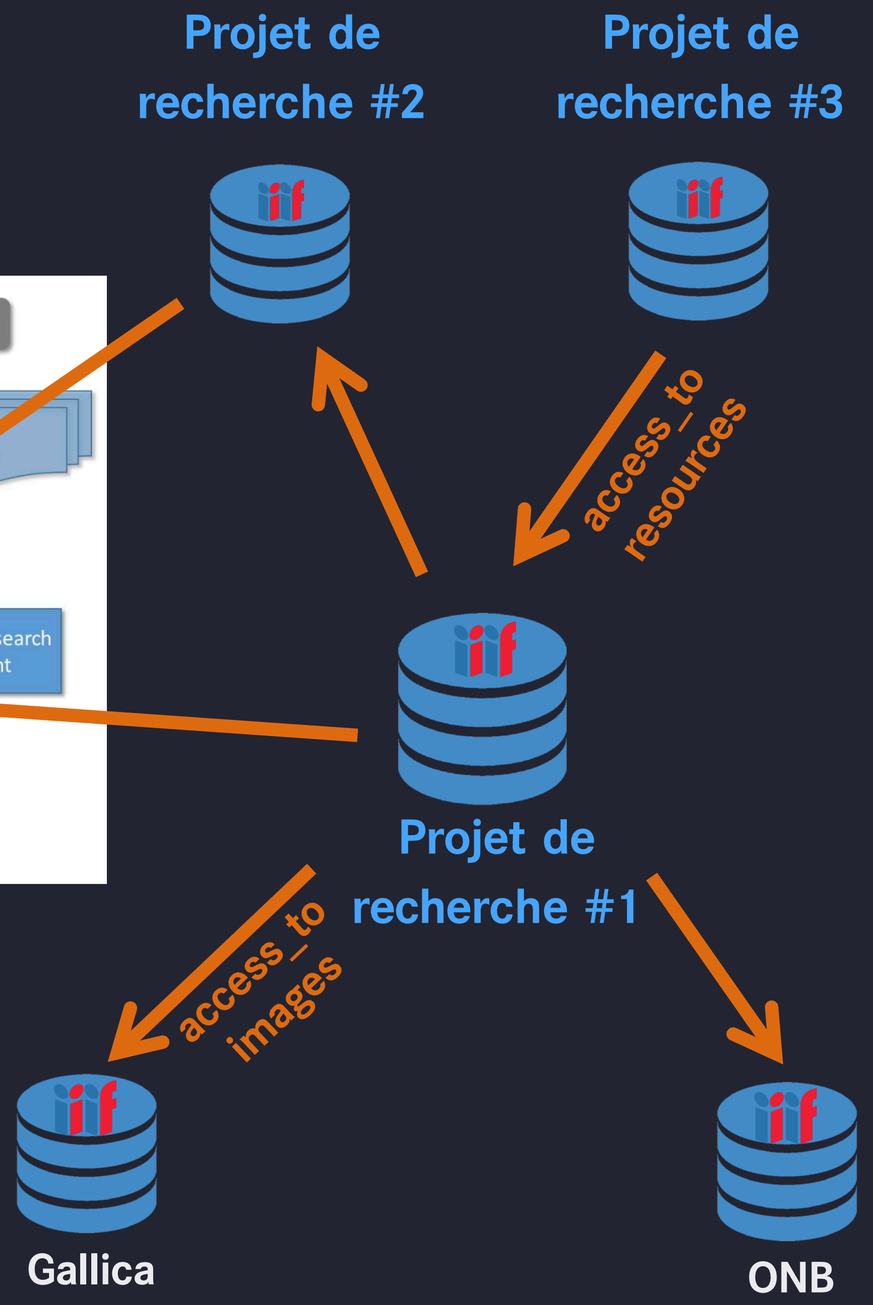
Interopérabilité

IIIF mais aussi :

- les données d'autorité
- le web de données
- TEI

…





Interopérabilité

À la BnF :

- **implémentation des API IIIF v3 ; exposition des ressources textuelles**
- **instance Mirador**
- **rénovation du rapport de recherche Gallica (sous-collections, export CSV, visualisation de données)**
- **portail compatible TEI pour les dictionnaires et encyclopédies historiques**

Coopération

ACCÈS AUX DONNÉES, FOURNITURE DE DONNÉES

CONSEIL ET FORMATION SUR LES MODÈLES DE DONNÉES, FORMATS ET POLITIQUES DOCUMENTAIRES

ANIMATION D'UNE COMMUNAUTÉ TRANSDISCIPLINAIRE

CONSEIL ET ORIENTATION

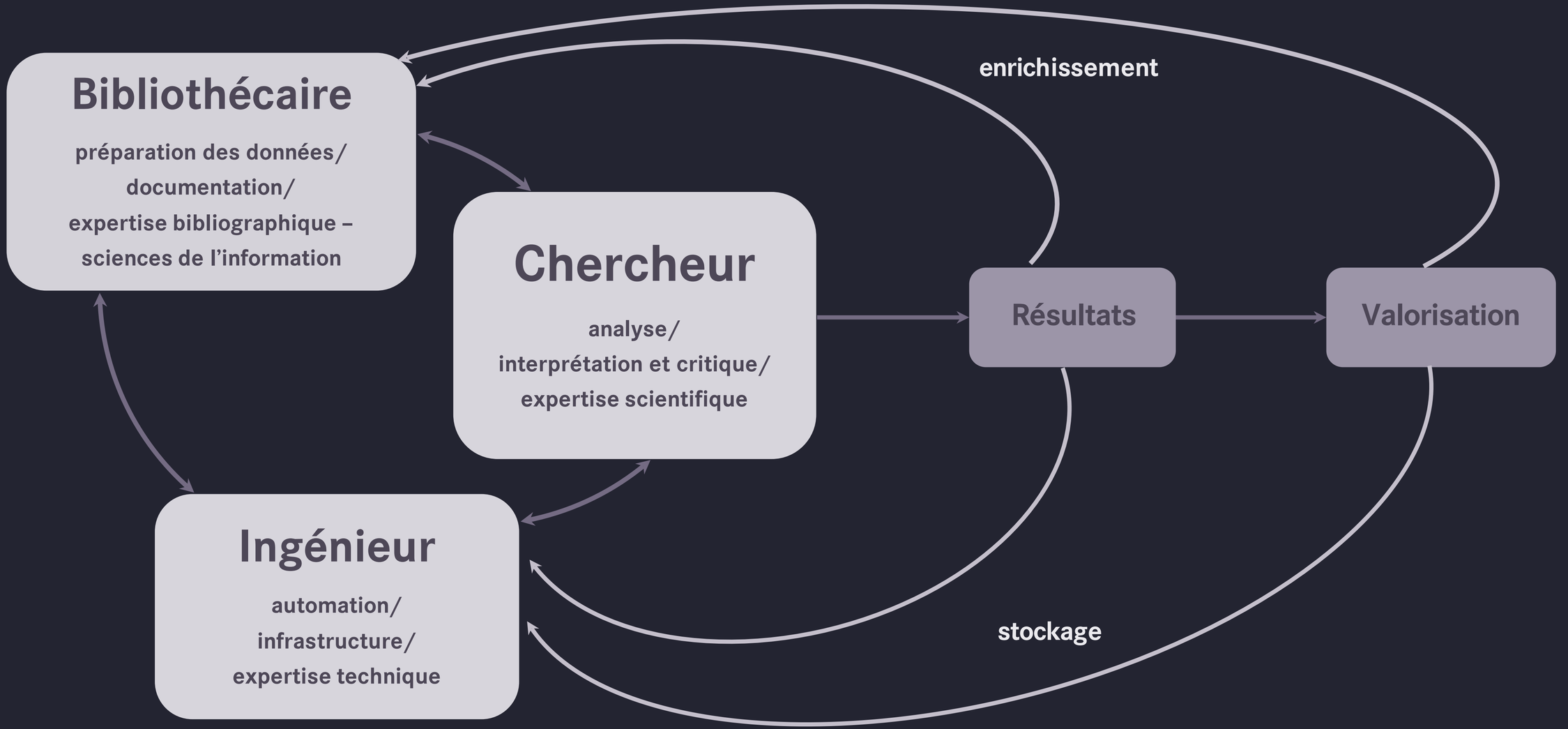
DISSÉMINATION ET INTÉGRATION D'OUTILS ET DE RÉSULTATS DE RECHERCHE

Bibliothèques et chercheurs

BnF Bibliothèque nationale de France

Coopération

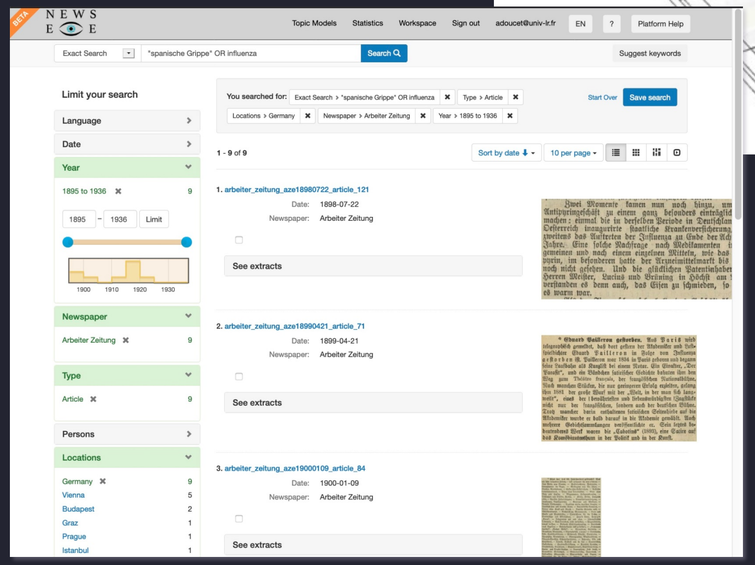
Les métiers



Coopération

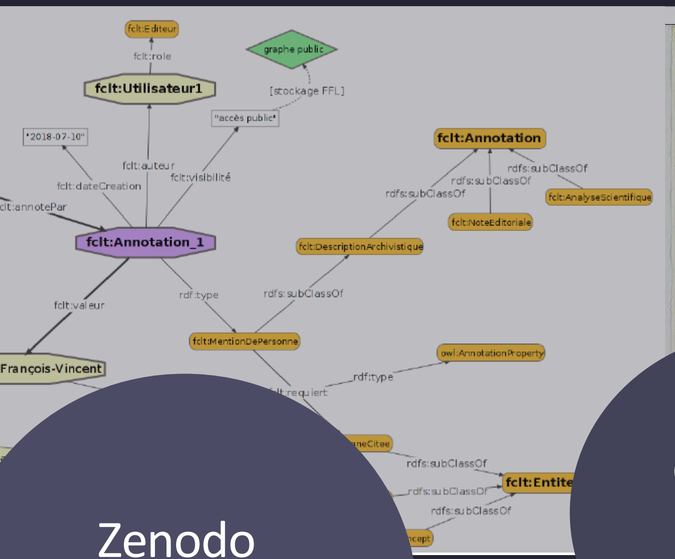
À la BnF : ouverture du BnF DataLab (septembre 2021) en coopération avec la TGIR Huma-Num

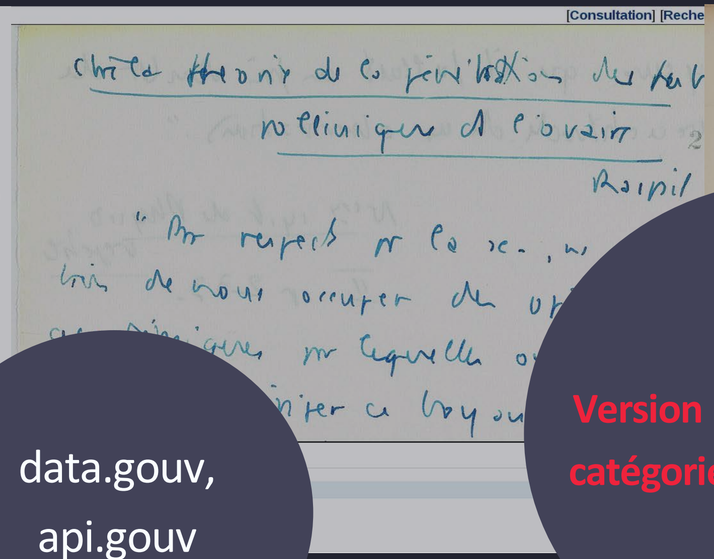
- Accompagnement
- Chercheurs et équipes en résidence
- Catalogue de services
- Boîte à outils
- Infrastructure



Coopération

Diffuser les jeux de données et autres ressources au plus près des utilisateurs





Zenodo

data.gouv,
api.gouv

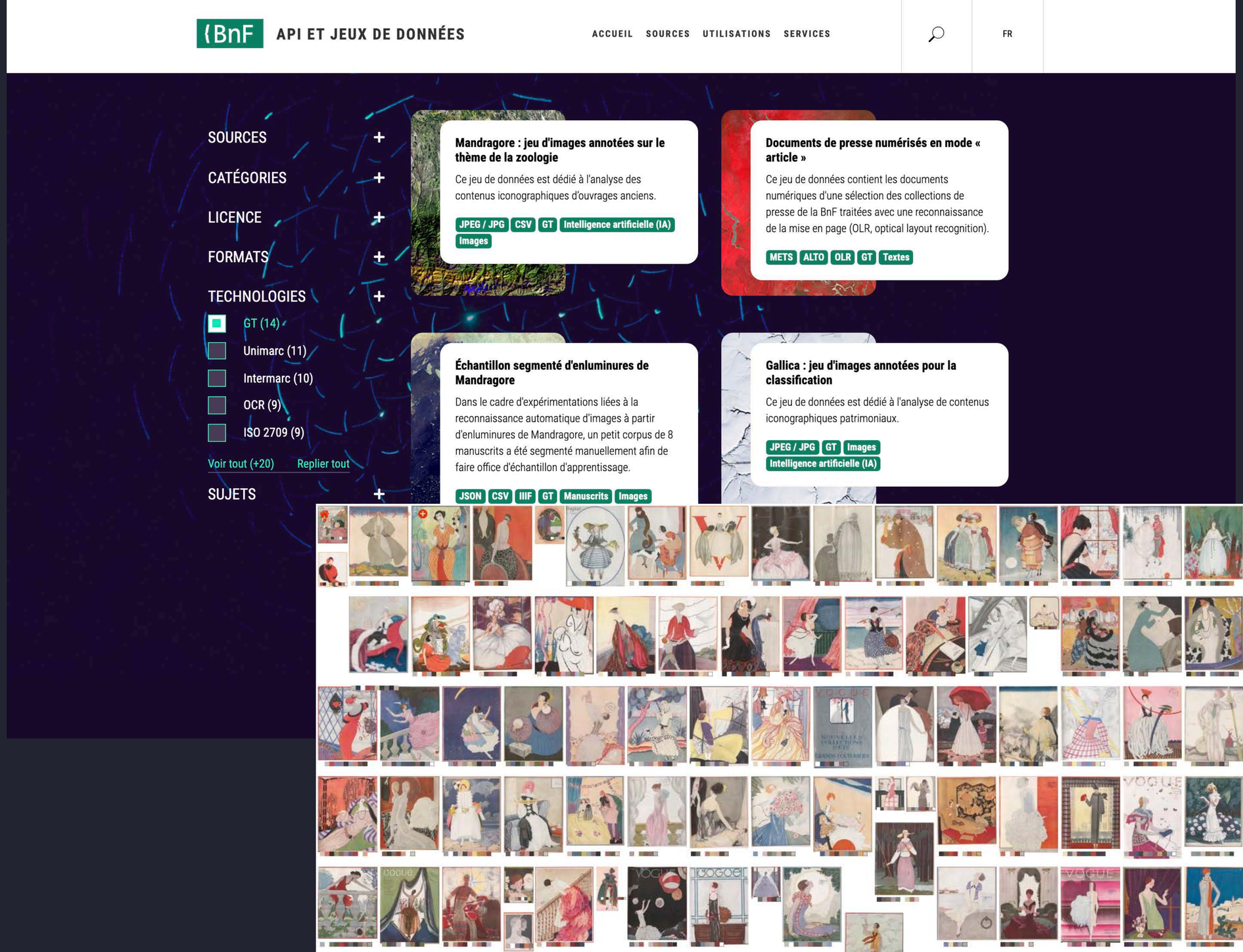
Compétitions
scientifiques

api.bnf
Version 2 du site, nouvelle catégorie « vérité terrain »

CLARIN
Historical Corpora

Coopération

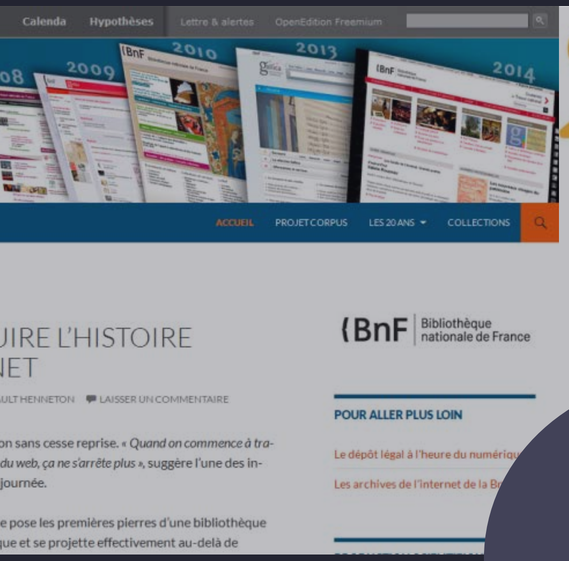
api.bnf.fr :
« vérité terrain »

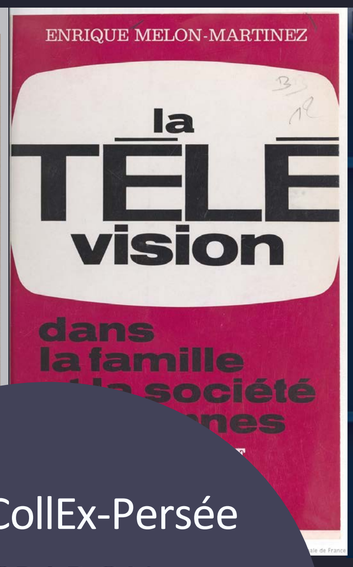


BnF | Bibliothèque nationale de France

Coopération

Trouver des solutions en matière de fouille de données sur les collections sous droits





Capsules
BnF DataLab

Projet CollEx-Persée
RESPADON
Université de Lille, SciencesPo Medialab, BnF, GED Condorcet

BnF Bibliothèque nationale de France

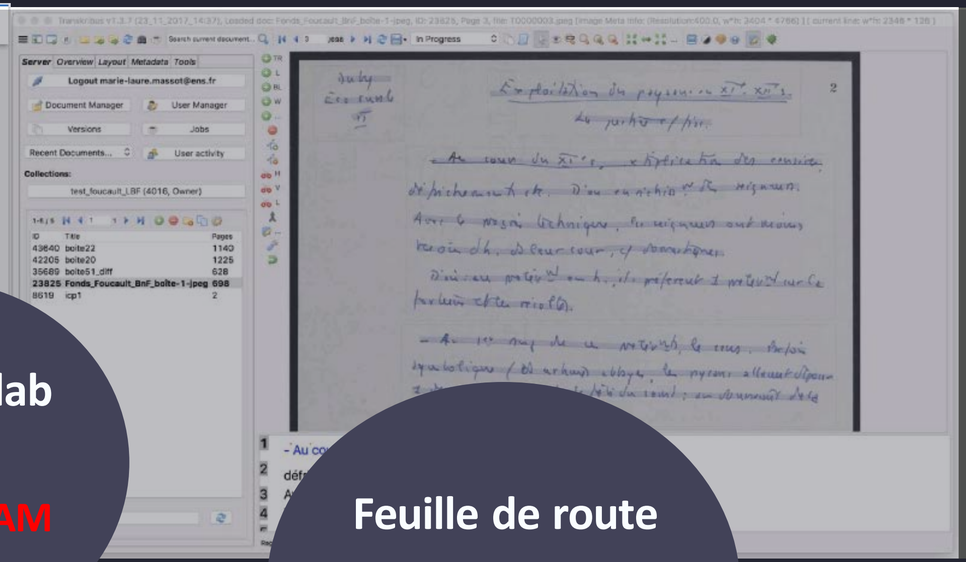
Coopération

Réfléchir à l'usage de l'IA en bibliothèque, expérimenter, réaliser

BnF Datalab

projets

CS-SHS-GLAM



Feuille de route

IA @ BnF

Initiative

IA4LAM

ia4lam.org

{BnF | Bibliothèque nationale de France

MERCI !

jean-philippe.moreux@bnf.fr

BnF | Bibliothèque nationale de France