

**CO-CONSTRUIRE LES HUMANITÉS  
NUMÉRIQUES ENTRE BIBLIOTHÈQUES ET  
CHERCHEURS : LE PROJET CORPUS**

**Emmanuelle Bermès**

**Journée sur la diffusion des données à la BnF**

**20/03/2018**

# LE PROJET DSR-CORPUS

- 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

→ Fiche dans la base ANIR : <http://c.bnf.fr/fom>



## USAGES DES CORPUS : 3 CAS DE FIGURE

- Le corpus comme source Newspapers
- Le corpus comme bac à sable
- Le corpus comme interface

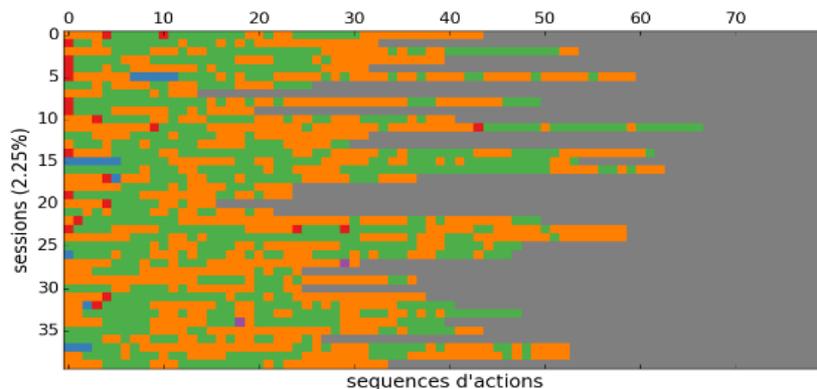
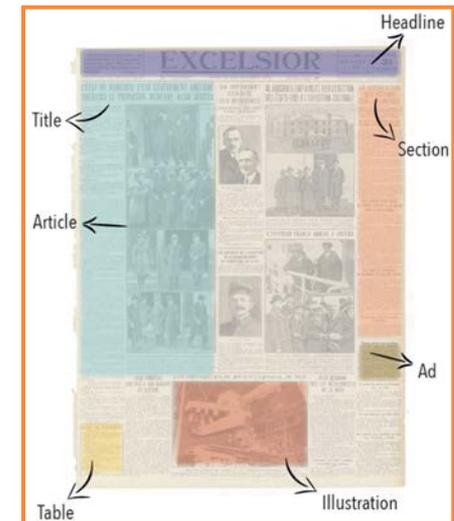


FIGURE 3.5 – cluster 4

Adrien Nouvellet et. al. *Analyse des traces d'usage de Gallica : Une étude à partir des logs de connexions au site Gallica*. [Rapport de recherche] Télécom ParisTech; Bibliothèque nationale de France. 2017. [〈hal-01709264〉](https://hal.archives-ouvertes.fr/hal-01709264)



JP Moreux,  
“Approches innovantes  
pour la presse  
ancienne numérisée :  
fouille et visualisation  
de données” in *Carnet  
de recherche à la  
bibliothèque nationale  
de France*, 3 décembre  
2016

<https://bnf.hypotheses.org/208>

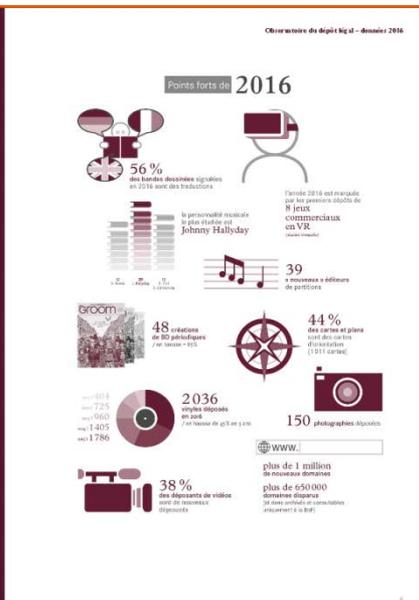
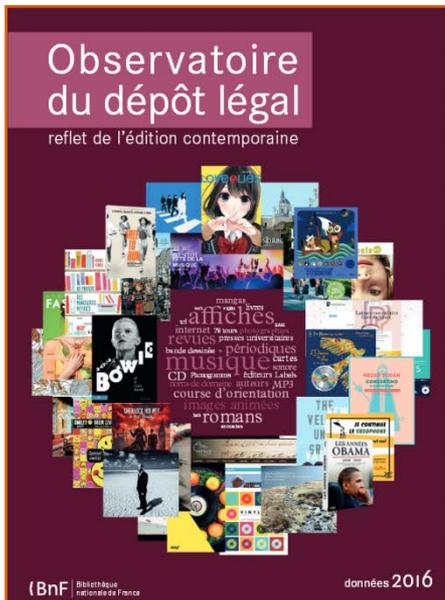
# QUELS CORPUS ?



Gallica + Gallica intra muros  
4.5 millions de documents numériques  
<http://gallica.bnf.fr>



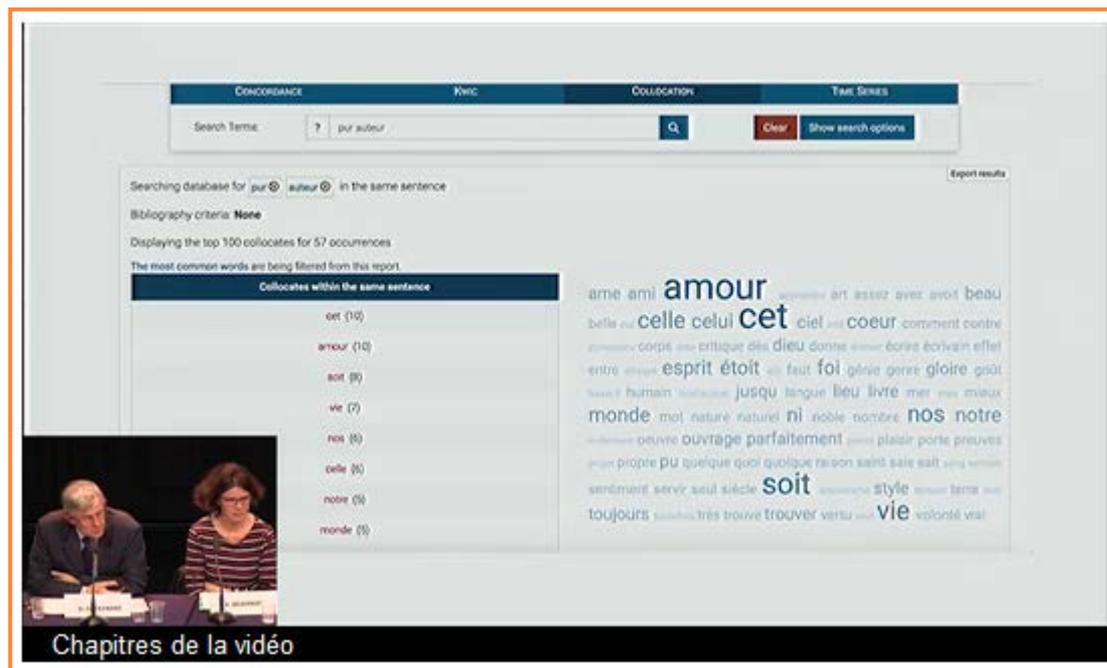
Archives de l'Internet  
793 To



Métadonnées  
+ de 20 millions de notices  
<http://data.bnf.fr>

# CORPUS ET MÉTADONNÉES : EXEMPLES

- Convention avec le laboratoire OBVIL :  
fourniture de 135 000 fichiers ALTO +  
métadonnées à des fins de fouille de texte



The screenshot displays the OBVIL search interface. At the top, there are tabs for 'CONCORDANCE', 'KWIC', 'COLLOCATION', and 'TIME SERIES'. The search term 'pur' is entered in the search box. Below the search bar, it indicates 'Searching database for pur in the same sentence' and 'Bibliography criteria: None'. The results show 'Displaying the top 100 collocates for 57 occurrences'. A list of collocates is shown on the left, including 'cet (10)', 'amour (10)', 'soit (9)', 'vie (7)', 'nos (6)', 'elle (6)', 'notre (5)', and 'monde (5)'. The main area displays a list of words: 'ame ami amour art assez avec avoit beau belle sur celle celui cet ciel coeur comment contre corps des dieu donne écrire écrivain effet entre esprit étoit fait foi génie gloire goût human jusque jusqu' langue lieu livre mer mieux monde mot nature naturel ni noble nombre nos notre oeuvre ouvrage parfaitement plaisir porte preuves propre pu quelque quoi quelque raison saint saie sait sang serment serment servir seul siècle soit style temps terre toujours très trouve trouver vertu vie volonté voir'. The word 'soit' is highlighted in blue. In the bottom left corner, there is a small video thumbnail showing two people at a table, with the text 'Chapitres de la vidéo' below it.

Présentation de Didier  
Alexandre, directeur du  
Labex OBVIL, lors de la  
table ronde sur la fouille  
de données aux JPAC  
2016

<http://c.bnf.fr/fuK>

Site OBVIL

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

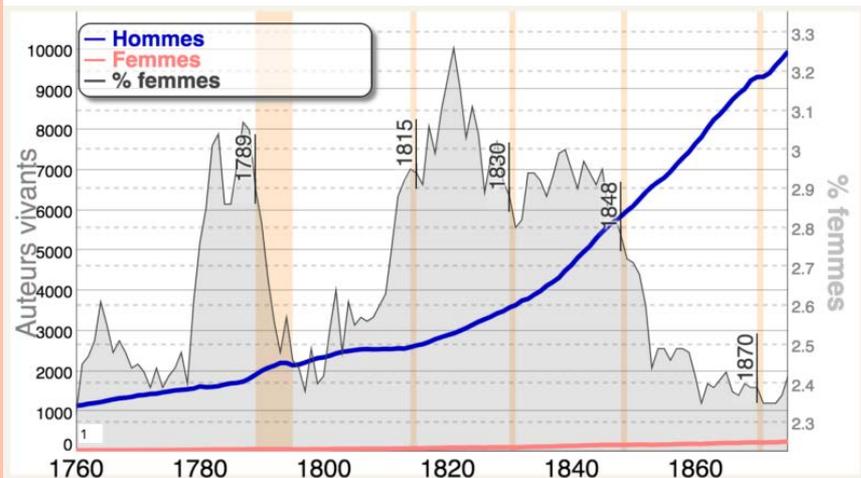
Bientôt sur

<http://api.bnf.fr>

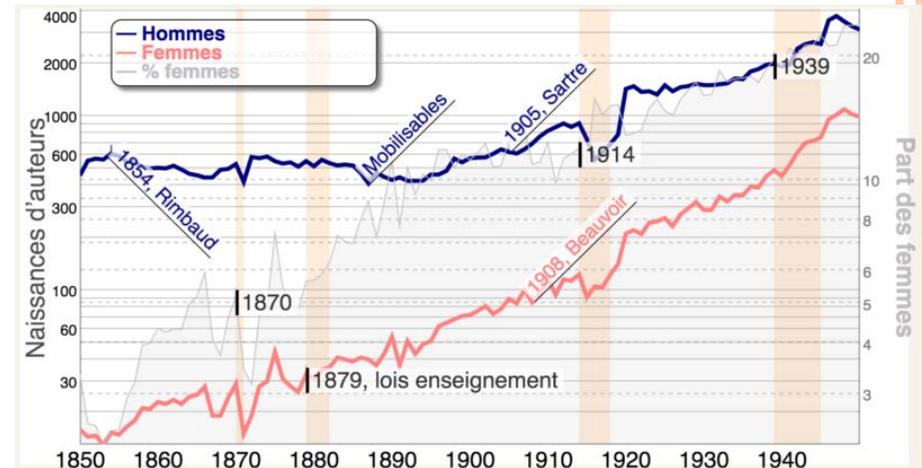


# CORPUS ET MÉTADONNÉES : EXEMPLES

- Travaux de Frédéric Glorieux sur data.bnf.fr -> étude sur la démographie des femmes de lettre  
<https://resultats.hypotheses.org/1048>



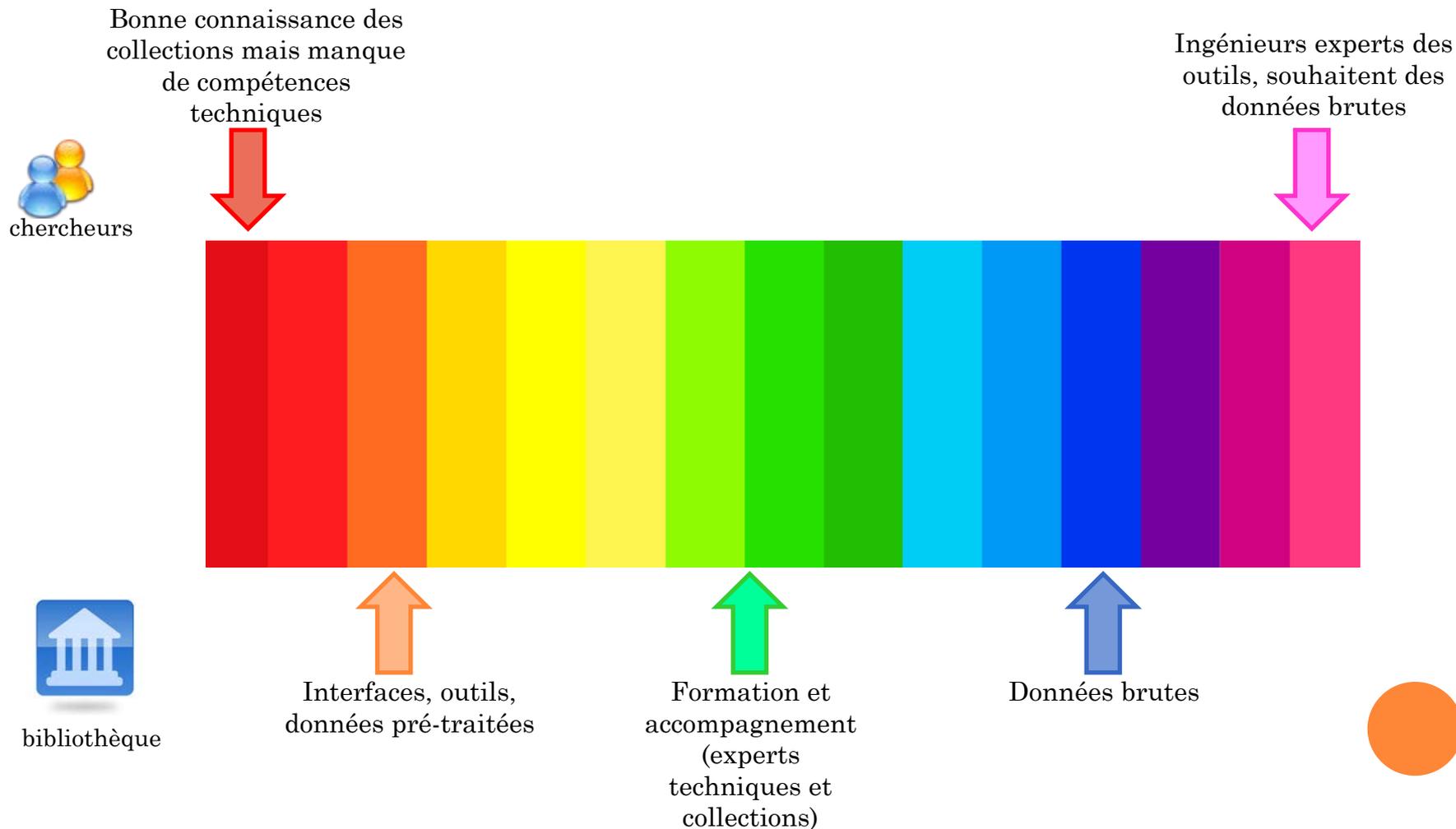
Databnf, 1760–1880,  
population des auteurs  
vivants, variation du ratio  
sexuel



Databnf, natalité des  
auteurs (pour chaque année,  
nombre de naissance au  
catalogue)



# UNE GRANDE DIVERSITÉ DE SITUATIONS...



# CONSTRUIRE L'AVENIR ?

## 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

## Loges

pour satisfaire le besoin de

TRAVAIL INDIVIDUEL  
LOCAUX

## Espaces du personnel

pour satisfaire le besoin de

ACCOMPAGNEMENT PERMANENT  
SERVICE PUBLIC  
RENDEZ-VOUS



## Collections

Nu­mé­ri­sées ou  
na­ti­ve­ment nu­mé­ri­ques

## Zone détente

pour satisfaire le besoin de

CONVIVALITÉ  
ÉCHANGE

## Infrastructure numérique

pour satisfaire le besoin de

ACCÈS À DISTANCE  
MOBILITÉ  
CONTRÔLE FLUX DE TRAVAIL  
ORDINATEUR PERSONNEL

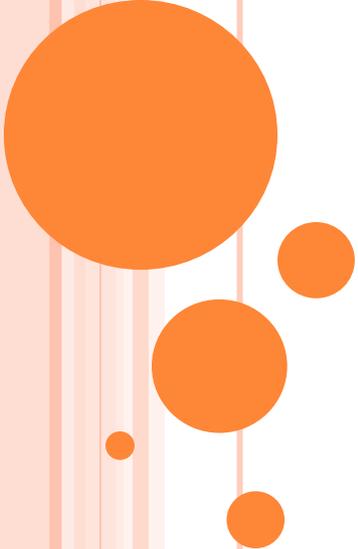
Offrir aux chercheurs, dans les emprises de la Bibliothèque, des outils de fouille et d'exploration de textes et de données sur des corpus numériques de la BnF

► 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





**MERCI !**

[emmanuelle.bermes@bnf.fr](mailto:emmanuelle.bermes@bnf.fr)

 [@figoblog](https://twitter.com/figoblog)