



Présentation du projet faite au séminaire « Science des données et SHS » (Cédric, Cnam), séance du 28 septembre 2021.

(Visuel, affiche du film L'Exorciste (1973), pochette de DV  
cf. <https://www.ecranlarge.com/films/840887-exorciste-l/photos>)



Contexte : réaction politique à la pandémie de la covid 19 → confinement de la population et fermeture des lieux publics, dont les bibliothèques

Confinement 1 → 17 mars 2020 – 11 mai 2020

Confinement 2 → 1<sup>er</sup> novembre 2020 – 15 décembre 2020

Visuel, © Bpi :

Tweet de la Bibliothèque publique d'information du Centre Pompidou

(@Bpi\_Pompidou) du 13 mars 2020 :

[△ FERMETURE] Suite aux mesures gouvernementales de prévention du coronavirus, la Bpi sera fermée à partir du samedi 14 mars et jusqu'à nouvel ordre. Merci pour votre compréhension.

## Projet éditorial

Ouvrage à paraître aux Presses de l'ENSSIB, coll. *La Numérique*, en décembre 2021

*Direction :*

- Clément Bert-Erboul
- Sylvie Fayet
- Louis Wiart

*Édition :*

- Muriel Amar, Paris Nanterre, ENSSIB

*Préface / Postface :*

- Christophe Evans, Bpi
- Olivier Zerbib, U. Grenoble

*Contributeurs :*

- Clément Bert-Erboul, ULB
- Romane Coutanson, MESRI
- Grégoire Clemencin, consultant
- Sylvie Fayet, BU La Rochelle
- Julien Hage, Paris Nanterre
- Geneviève de Maupeou, Bpi
- Cécile Touitou, Sc Po Paris
- Maryline Vallez, Bpi
- Louis Wiart, ULB

 <https://presses.enssib.fr/catalogue/lombre-des-bibliotheques>



Commande émise par l'ENSSIB (école de formation de bibliothécaires)

Recherche d'un éclairage sur l'(in)activité des bibliothèques au cours des confinements de l'année 2020

Constitution d'une équipe pluridisciplinaire pour apporter des éléments de réponse

Démarrage du projet : janvier 2021

Finalisation de la rédaction du livre : août 2021

→ délais particulièrement serrés !

## Questionnement

Qu'ont fait les usagers des bibliothèques lorsque celles-ci ont fermé leurs portes ?

- Comment l'ont-ils fait ?
- Qu'en ont-ils retenu ?

*(Et qu'en auront retenu les bibliothèques elles-mêmes ?)*

Focus sur le groupe Facebook  
« Bibliothèque Solidaire du confinement »

Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques



Le projet est une commande ==> la démarche de recherche est « située », nous travaillons « après coup »

Q. « *qu'en auront retenu les bibliothèques ?* »

R. Le projet et le livre qui en résulte sont un élément de réponse que l'ENSSIB se propose d'apporter à la communauté

(Visuel porte close : <https://500px.com/p/gerago>)



Groupe créé le 16 mars 2020

À partir d'une idée initiée sur Twitter

Groupe d'abord public, devenu privé le 14 avril 2020

Compte aujourd'hui près de 70 000 membres inscrits

## Étudier un groupe Facebook ?

### Appréhender l'objet d'étude

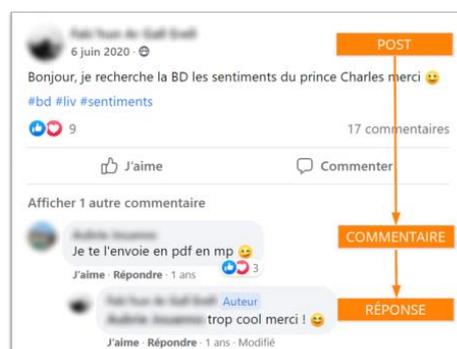
- un type de « ressource » du Web
- un outil d'expression cadré

### Un flux de messages datés et signés

- des commentaires à ces messages
- des réponses à ces commentaires
- des réactions, ...

### Un groupe « privé »

- contenu réservé aux membres inscrits
- membres, modérateurs, administrateurs



Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques

6

Objet d'étude privilégié dans le cadre du projet : un « groupe » Facebook.

C'est une « ressource » spécifique du Web (identifiée par un URL)

Un outil mis gratuitement à disposition d'un ensemble d'utilisateurs, mais développé par une entreprise commerciale : on n'y fait que ce que FB veut bien qu'on y fasse. (Q. peut-on faire une bibliothèque dans le cadre d'un groupe FB ?)

En pratique, c'est une suite de messages, articulés selon une hiérarchie particulière

Le groupe est « privé » → les messages sont réservés aux membres (ce n'était pas le cas lors du 1<sup>er</sup> mois d'activité du groupe)

Tous les membres ne sont pas égaux  
→ 10 administrateurs / modérateurs

→ Il a fallu « apprendre » l'objet étudié

## Démarche

### 1/ Enquêtes qualitatives

- Entretiens avec des membres volontaires du groupe
- Analyses spécifiques de certains contenus textuels

### 2/ Analyse quantitative

- Que peut-on factueliser de l'activité au sein du groupe ?
  - Caractérisation des membres
  - Caractérisation des messages
  - Dimension temporelle
- Contraintes de délais
  - Contraintes éditoriales

Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques

7

Comprendre ce qui s'est fait (et pas fait) dans le cadre de la BSc  
càd dans le cadre des fonctionnalités offertes par le groupe FB

d'une part, en questionnant les acteurs

- entretiens
- zoom sur des extraits de messages

d'autre part, en observant leurs *traces* dans le groupe BSc

- qui a produit quoi, quand, comment

Le tout avec des contraintes de délai importantes.

## Matériaux – Qualitatif

- 11 entretiens individuels
  - semi-directifs
  - réalisés à distance (1h30 – 2h00)
  - auto-désignation des personnes interrogées
  - sélection en fonction de leur activité
- 1 entretien collectif
  - avec les gestionnaires du groupe
- Enregistrement & retranscription



Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques

8

Entretiens individuels :

5 font l'objet de portraits spécifiques, les 6 autres exploités via des extraits de verbatims

Outils informatiques de retranscription

→ remarques sur leur capacités, points forts, points faibles...

→ en l'état, décorrélés des outils de traitement des messages du groupe

Visuels :

- Duras-Godard <https://www.theparisreview.org/blog/2020/10/27/cinema-hardly-exists-duras-and-godard-in-conversation/>
- Flickr, <https://www.flickr.com/photos/65359853@N00/16917470567>

## Matériaux – Quantitatif

### A. Source Facebook

- Fichier Excel
- Chiffres quotidiens
  - nb membres
  - nb messages
- Chiffres cumulés
  - âge, sexe
  - ville, pays
  - jour et heure de connexion

### B. Source Web

- Extraction (*scraping*) de la page
- Balisage de composants
  - message
    - contenu textuel
    - hiérarchie
  - horodatage
  - signataires
- Stockage en base de données

Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques

9

Deux sources :

- Fichier Excel produit par FB
- Données brutes extraites de la page Web du groupe

Mais questions sous-jacentes :

- que compte-t-on, que mesure-t-on exactement ?
- comment interpréter des résultats quantifiés ?

Exemple simple : « les X utilisateurs du groupe publient en moyenne Y messages par mois »

→ à quoi raccrocher cette information ?



# Collecte & exploitation de données

Acquisition, structuration & restitution

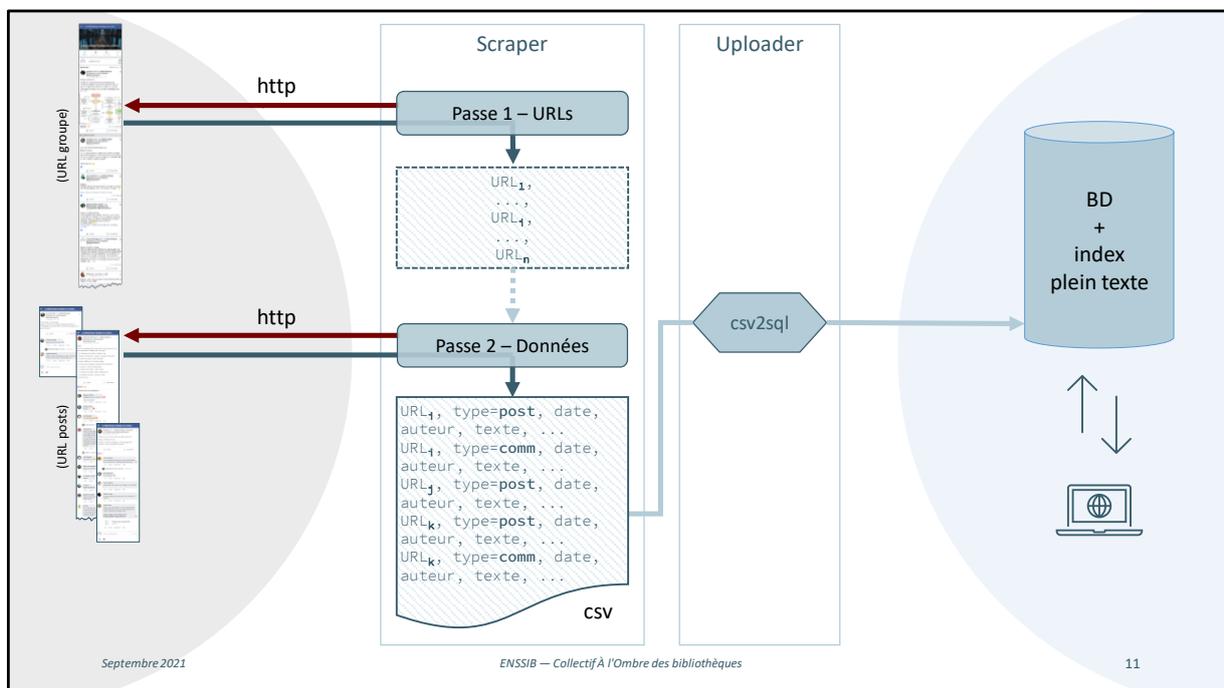
Septembre 2021

ENSSIB — Collectif À l'Ombre des bibliothèques

10

Zoom sur les aspects des travaux « quanti » menés dans le cadre du projet

(Visuel, BNP Paribas <https://histoire.bnpparibas/dossier/petite-histoire-de-linformatique-bancaire/>)



### **Schéma d'acquisition des données depuis la page Facebook du groupe BSc**

**Scraper** – Un robot parcourt la page « Discussion » du groupe Facebook en deux temps :

- passe 1 : identification et extraction des URLs des messages de 1<sup>er</sup> niveau (posts)
- passe 2 : pour chaque URL de post, extraction du contenu, dont les messages de 2<sup>e</sup> niveau (commentaires)

(Les associations posts/comm se font sur l'URL du post.)

Il écrit les données extraites avec leur structuration interne dans un fichier CSV.

Techno utilisée : extension Web Scraper de Chrome

**Uploader** – Le fichier CSV est réinterprété et le contenu normalisé pour alimenter une base de données (baptisée BScSnap) selon un schéma prédéfini.

La base est intégrée à un applicatif d'interrogation exploitant notamment un index plein texte.

Techno utilisée : Python + SQLite (et module FTS)



Un exemple extrait du fil de discussion du groupe BSc (1/2)

Et un extrait du CSV produit par le scraper à partir de cette source.

À noter le lien implicite entre post et commentaire via la mention du même URL, et le typage de chaque message.

N.B. Pour qq exemples des problèmes qui surgissent dans la récupération de la matière textuelle des messages, voir Annexes, « Données textuelles brutes »

The image shows a screenshot of a Facebook discussion thread with overlaid CSV data and annotations. The CSV data is as follows:

```

"11";"https://m.facebook.com/groups/bibliothequesolidaire/permalink/FBID_A/"FALSE;"posts";"USER_A";"16
mars 2020, 15:44";"Chères toutes, chers tous, L'initiative est super. Je n'ai pas les moyens de
photographier toutes les couvertures de ma bibliothèque, mais [...]";"14554245-
111705";"J'aimeJ'adoreHaha26";"26";"FALSE";18;99;0;0;0;0;FALSE;"user_a";FALSE;6,08068315665489;2020-
03-16;" 15:44";"2020-03-01";"male";"1_lockdown";1

"16";"https://m.facebook.com/groups/bibliothequesolidaire/permalink/FBID_A/"FALSE;"comments";"USER_B";
;01 février 2020,";"Les gens pourraient probablement parler de leur sujet de recherche/thèse/mémoire
en commentaire [...]";"1614554245-
111697";"FALSE";"FALSE";"FALSE";"null";18;31;0;0;0;0;FALSE;"user_b";";2020-02-01";"2020-02-
01";"no_lockdown";1

```

Annotations in the image include circles around "FBID\_A" and "posts" in the first row, and "FBID\_A" and "comments" in the second row. Arrows point from these annotations to the corresponding fields in the Facebook interface below.

The Facebook interface shows a post with 26 likes and 17 comments. A comment is visible with the text: "Les gens pourraient probablement parler de leur sujet de recherche/thèse/mémoire en commentaire sur un post et citer les ouvrages les plus importants, et les personnes les ayant pourraient leur répondre directement 😊".

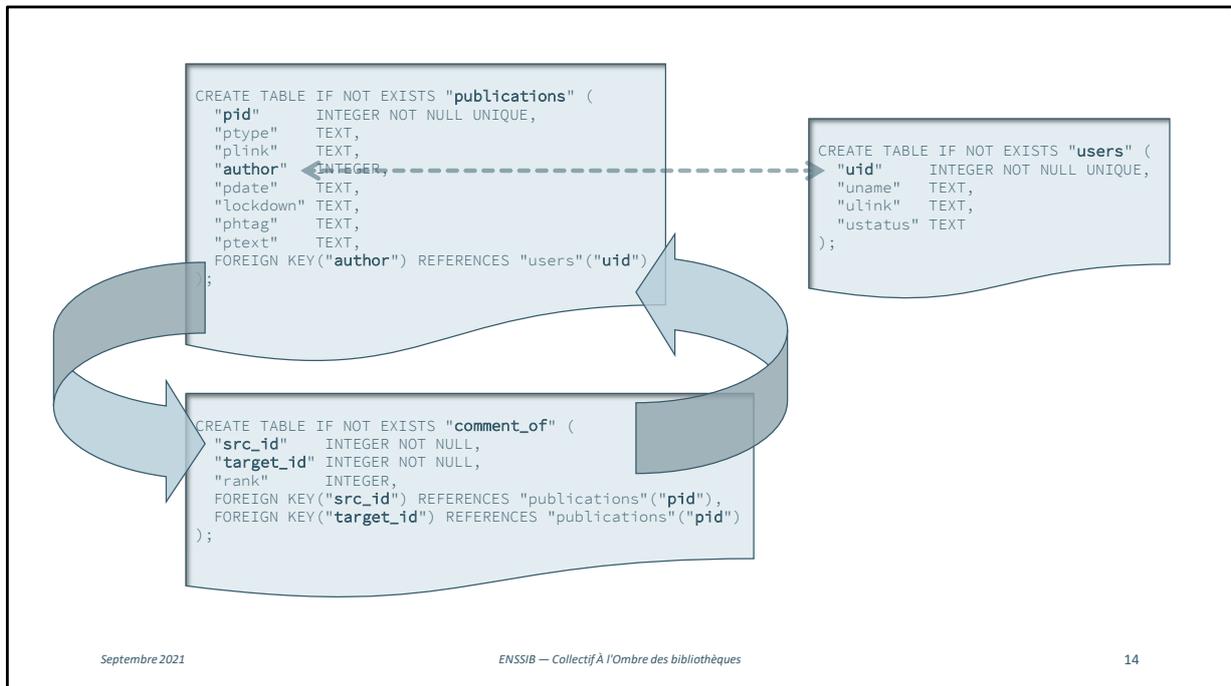
Page footer: Septembre 2021 | ENSSIB — Collectif À l'Ombre des bibliothèques | 13

Un exemple extrait du fil de discussion du groupe BSc (2/2)

Et un extrait du CSV produit par le scraper à partir de cette source.

À noter le lien implicite entre post et commentaire via la mention du même URL, et le typage de chaque message.

N.B. Pour qq exemples des problèmes qui surgissent dans la récupération de la matière textuelle des messages, voir Annexes, « Données textuelles brutes »



Structuration de l'information dans la base de données BScSnap reconstruite à partir du CSV.

Différentes entités ont été dégagées :

- les publications
- les auteurs de publications
- les mots-dièse apparaissant dans les textes des publications
- ...

Ces entités entretiennent diverses associations

- directes par le partage d'attributs (identifiants), cf. les auteurs
- indirectes par la mise en relation explicite via une table d'association, cf. les posts et leurs commentaires

## Fiche signalétique de la BSc

	Source Facebook	Source BScSnap
Membres inscrits	66 395	
Membres contributeurs		23 037
Total messages	218 933	140 796
- posts	27 100	19 665
- commentaires		59 313
- réponses		61 818
- comm. + réponses	191 833	121 131
- réactions	393 835	

Données au 18 janvier 2021

À remarquer : les « trous » dans les infos

Q méthodologique de base :

- comment « saisir » un groupe FB ?
- quelle en est la référence ?
- et bien sûr : quels sont les bons indicateurs à observer ?

+ variations FB / BScSnap sur les données comparables