

# L'enquête de terrain dans tous ses états: le web et les réseaux sociaux

Séminaire M2 "Préparation de l'enquête de terrain (focus terrains difficiles)"

**Jean-François Legrain**

Aix Marseille Université

CNRS/IREMAM

(Institut de Recherches et d'Études sur le Monde Arabe  
et Musulman)

<http://www.iremam.cnrs.fr/legrain>

Sciences Po, Aix-en-Provence, 26 octobre 2016



# Enquêter dans les mondes numériques

- Approcher le web comme **source primaire** ou comme **source secondaire** constitue deux démarches de construction de **corpus** qui impliquent :
  - des **méthodes communes** (collecter, vérifier, valider, sauvegarder, analyser)
  - des **outils appropriés**

# A – Une méthode pour construire un corpus



- Processus **itératif** et **cumulatif**
- **Collecter, vérifier, valider, sauvegarder et analyser**

# Construire le corpus

Collecter, vérifier, valider : à partir d'où?



Maîtriser le reflexe Google et préciser sa question

Privilégier la quête "à la source"

Dans le cas du web classique, exploitation des

- rubriques "sites amis" ou équivalent sur un site de référence
- liens sortants (*outbound links*) du site et liens entrants (*backlinks*), ie ceux qui dirigent vers lui de l'extérieur

[Opérateurs de recherche sur les moteurs](http://iremam.cnrs.fr/): <link:http://iremam.cnrs.fr/ >

[BacklinkWatch](http://www.backlinkwatch.com/) (http://www.backlinkwatch.com/)

[Alexa](http://www.alexa.com/) (http://www.alexa.com/)

**BACKLINK  
WATCH!**



Dans le cas des réseaux sociaux, exploitation des

- amis, abonnés, abonnements, followers, followings, etc.

# Construire le corpus

## Collecter, vérifier, valider : qui?



**Attention : le papier ne refuse pas l'encre pas plus que le numérique ne refuse l'octet**

- **Pour un site web classique :**
  - Complexité de la notion d'auteur
  - Rubriques "à propos", mentions légales, etc.
  - URL et whois: [DomainTools](http://whois.domaintools.com/) (<http://whois.domaintools.com/>)
  - Métadonnées (ctrl+U)
- **Pour les réseaux sociaux :**
  - Absence de whois
  - Rubriques "à propos"
  - Multiplication des comptes, "pages officielles" ou supposées telles
  - Exploitation des listes d'amis ou des mentions "j'aime"
- **Témoignage direct comme la meilleure garantie**
- **Vérification avec les liens entrants ou sortants des sites ou comptes**

# Construire le corpus

Collecter, **vérifier**, valider : **quand?**



- **Pour un site web classique :**
  - Complexité de la notion de date
  - Le site: whois [DomainTools](http://whois.domaintools.com/) (<http://whois.domaintools.com/>)
  - La page: contenu, métadonnées, ‘informations sur la page’
  - Les éléments: ‘afficher l’image’, ‘ouvrir le lien dans un nouvel onglet’
  - La recherche par date des moteurs
- **Pour les réseaux sociaux :**
  - Facebook et la possibilité de mentir

# Construire le corpus

Collecter, vérifier, **valider**, sauvegarder



**Le document répond à la requête. Il est ce qu'il dit être.**

La logique traditionnelle de la diplomatique s'applique aux documents analogiques comme aux documents numériques

# Construire le corpus

Collecter, vérifier, valider, sauvegarder



- **Pour un site web classique**
  - ‘Travailler hors connexion’
  - Aspirateur de sites comme [WinHTTrack](http://www.httrack.com/) Website Copier (<http://www.httrack.com/>)



- **Pour les réseaux sociaux**
  - L'archivage privé d'un compte Facebook ou Twitter passe par l'ouverture de toute la période recherchée suivie d'un copier/coller sous Word

# **B – Des outils pour construire et contextualiser un corpus**



**Profiter de l'expertise des autres**

# Profiter de l'expertise des autres

## Moteurs généralistes en 2<sup>e</sup> intention

Les moteurs ignorent le web invisible des bases de données (à la différence des répertoires/signets) Chaque moteur range ses résultats selon un algorithme propre

**Filtrer** est mieux que recouper avec métamoteur: utiliser les **fonctions avancées**



[Google](https://www.google.fr/advanced_search) (https://www.google.fr/advanced\_search)

[Yahoo](https://fr.search.yahoo.com//web/advanced) (https://fr.search.yahoo.com//web/advanced)

[Bing](http://www.bing.com/) (http://www.bing.com/)

# Profiter de l'expertise des autres

## Répertoires, guides et signets

[Dmoz](http://www.dmoz.org/) (<http://www.dmoz.org/>)

[Intute](http://www.intute.ac.uk/) (pour Internet Tutorial) (<http://www.intute.ac.uk/>)



[Les Signets de la Bibliothèque nationale de France](http://signets.bnf.fr/) (<http://signets.bnf.fr/>)

[Signets-universités](http://www.signets-universites.fr/fr/) (<http://www.signets-universites.fr/fr/>)

[Signets de la France contemporaine](http://www.bpi.fr) (<http://www.bpi.fr>)

[Databases: Library of Congress Electronic Resources Online Catalog](http://eresources.loc.gov/)  
(<http://eresources.loc.gov/>)

[Internet pour les journalistes \(IPLJ\)](http://www.cyberjournalisme.net/) (<http://www.cyberjournalisme.net/>)



[Zotero](http://www.zotero.org/) (<http://www.zotero.org/>)



# Profiter de l'expertise des autres

## Moteurs dédiés aux productions académiques

[BiblioSHS](http://biblioshs.inist.fr.gate3.inist.fr/) (http://biblioshs.inist.fr.gate3.inist.fr/) (réservé au CNRS)

[Environnement numérique de Travail \(ENT\)](http://ent.univ-amu.fr/) de l'AMU (http://ent.univ-amu.fr/)

[ISIDORE \(Intégration de Services, Interconnexion de Données de la Recherche et de l'Enseignement\)](http://www.rechercheisidore.fr/) (http://www.rechercheisidore.fr/)

[scanR, moteur de la recherche et de l'innovation](https://scanr.enseignementsup-recherche.gouv.fr/)

(https://scanr.enseignementsup-recherche.gouv.fr/)



[OpenAIRE](https://openaire.eu/search/find) (https://openaire.eu/search/find)

[OpenGrey](http://www.opengrey.eu/) (http://www.opengrey.eu/)

[BASE](http://www.base-search.net/) (Biefefeld Academic Search Engine) (http://www.base-search.net/)

[Infomine](http://infomine.ucr.edu/) (http://infomine.ucr.edu/)



[Google Scholar](http://scholar.google.fr/) (http://scholar.google.fr/)

[Microsoft Academic Search](http://academic.research.microsoft.com/) (http://academic.research.microsoft.com/)



# Profiter de l'expertise des autres

## Bibliothèques, catalogues et métacatalogues

[Karlsruhe Virtual Catalog \(KVK\)](http://www.ubka.uni-karlsruhe.de/) (<http://www.ubka.uni-karlsruhe.de/>)



[Catalogue collectif de France \(CCFr\)](http://ccfr.bnf.fr/portailccfr/) (<http://ccfr.bnf.fr/portailccfr/>)

[British Library](http://www.bl.uk/) (<http://www.bl.uk/>)

[Library of Congress](http://catalog2.loc.gov/) (<http://catalog2.loc.gov/>)



[WorldCat](http://oaister.worldcat.org/) (<http://oaister.worldcat.org/>)



# Profiter de l'expertise des autres

## Thèses et autres mémoires

[Thèses produites en France](http://www.theses.fr/) (<http://www.theses.fr/>)

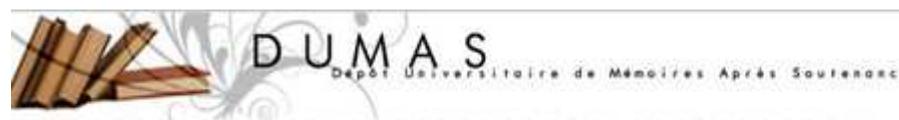
[TEL \(Thèses-en-ligne\)](http://tel.archives-ouvertes.fr/) (<http://tel.archives-ouvertes.fr/>)

[Atelier National de Reproduction des Thèses](http://www.diffusiontheses.fr/) (<http://www.diffusiontheses.fr/>)

[DUMAS \(Dépôt Universitaire de Mémoires Après Soutenance\)](http://dumas.ccsd.cnrs.fr/)

(<http://dumas.ccsd.cnrs.fr/>)

[Travaux universitaires inédits de science politique](http://tuisp.online.fr/) (<http://tuisp.online.fr/>)



[Bictel](http://www.bictel.be/) (<http://www.bictel.be/>)

[EThOS \(Electronic Theses Online Service\)](http://ethos.bl.uk/) (<http://ethos.bl.uk/>)

[DART \[Digital Access to Research Theses\]-Europe](http://www.dart-europe.eu/) (<http://www.dart-europe.eu/>)

[Networked Digital Library of Theses & Dissertations \(NDLTD\)](http://www.ndltd.org/) (<http://www.ndltd.org/>)

[OCLC WorldCat Dissertations and Theses](http://www.oclc.org/) ([http://www.oclc.org/...](http://www.oclc.org/))



# Profiter de l'expertise des autres

## Bibliothèques numériques

[Gallica](http://gallica.bnf.fr/) (http://gallica.bnf.fr/)

[Europeana](http://www.europeana.eu/) (http://www.europeana.eu/)

[Directory of Open access Books \(DOAB\)](http://www.doabooks.org/doab) (http://www.doabooks.org/doab)

[Directory of Open Access Journals \(DOAJ\)](http://www.doaj.org/) (http://www.doaj.org/)

[European Reference Index for the Humanities \(ERIH\)](https://www2.esf.org/...) (https://www2.esf.org/...)

[British Library](http://www.bl.uk/onlinegallery/index.html) (http://www.bl.uk/onlinegallery/index.html)

[E-Resources Online Catalog](http://eresources.loc.gov) (http://eresources.loc.gov)

[HathiTrust](http://www.hathitrust.org/) (http://www.hathitrust.org/)

[Universal Library Project](http://www.ulib.org/) (http://www.ulib.org/)

[Bibliothèque numérique mondiale \(BNL\)](http://www.wdl.org/) (http://www.wdl.org/)

[Google Books](http://books.google.fr/) (http://books.google.fr/)

[Internet Archive](http://www.archive.org/) (http://www.archive.org/)

[Bibliothèques numériques](http://signets.bnf.fr/html/categories/c_011textes_num.html)

(http://signets.bnf.fr/html/categories/c\_011textes\_num.html)

[List of digital library projects](http://en.wikipedia.org/wiki/List_of_digital_library_projects)

(http://en.wikipedia.org/wiki/List\_of\_digital\_library\_projects)

# Profiter de l'expertise des autres

## Archives ouvertes et réseautage

[HAL-SHS \(Sciences de l'Homme et de la Société\)](http://halshs.archives-ouvertes.fr/) (<http://halshs.archives-ouvertes.fr/>)



[Registered Data Providers](http://www.openarchives.org/Register/BrowseSites) (<http://www.openarchives.org/Register/BrowseSites>)

[Registry of Open Access Repositories \(ROAR\)](http://roar.eprints.org/) (<http://roar.eprints.org/>)

[OpenDOAR](http://www.opendoar.org/) [Directory of Open Access Repositories] (<http://www.opendoar.org/>)



Registry of Open Access Repositories

*OpenDOAR*

[Academia.edu](http://www.academia.edu/) (<http://www.academia.edu/>)

[Social Science Research Network \(SSRN\)](http://www.ssrn.com/) (<http://www.ssrn.com/>)

[ResearchGate](https://www.researchgate.net/) (<https://www.researchgate.net/>)

[Mendeley](http://www.mendeley.com/) (<http://www.mendeley.com/>)

[MyScienceWork](http://www.mysciencework.com/) (<http://www.mysciencework.com/>)



# Accéder au web passé

## Via les "archives" publiques



- Critique de la notion d'archive du web
- Les fonds
  - [Wayback Machine](http://www.archive.org/) de l'Internet Archive (<http://www.archive.org/>) revendique l'archivage de quelque 510 milliards de pages
  - [Archive.is](http://archive.is/) (<http://archive.is/>) ou [WebCite](http://webcitation.org/) (<http://webcitation.org/>), lequel affiche un objectif académique
  - Les archives nationales (France, UK, USA, etc.)
  - Facebook, aucune archive publique. Twitter, dépôt à la Library of Congress



[International Internet Preservation Consortium \(IIPC\)](http://netpreserve.org/) (<http://netpreserve.org/>)  
[List of Web archiving initiatives](http://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives)  
([http://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](http://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives))

## **Jean-François Legrain**

Aix Marseille Université

CNRS/IREMAM

(Institut de Recherches et d'Études sur le Monde Arabe et Musulman)

<http://www.iremam.cnrs.fr/legrain>