

# HAL API

Création de réseaux à partir d'une archive en ligne

# HAL

- “Hyper Article en Ligne”
- Dépôt d’articles scientifiques en ligne
- Développée par le CCSD (Centre pour la communication scientifique directe) du CNRS

# HAL

- Répertoires de Dépôt:
  - HAL (France), ArXiv, biorXiv, ...
  - Chercheurs déposent des “preprints” (version soumise d’articles sans le travail fait par l’éditeur de publication), libres d’accès.
  - HAL et ArXiv sont liées : dépôt commun possible
- Sur HAL, on trouve aussi simplement des méta-données:
  - Titres, auteurs, affiliations, résumé, etc., mais pas l’article lui-même.

# HAL

- Position particulière en France:
  - La majorité des chercheurs français ont l'obligation de déposer au moins les méta-données de tous leurs articles publiés dans HAL
  - Cela permet un suivi quantitatif de la recherche française:
    - Combien de papiers publiés par laboratoire, université, voire chercheur...
- IdHAL: identifiant unique d'un chercheur (homonymies...)

# HAL

- <https://hal.archives-ouvertes.fr>
- <https://hal-univ-lyon1.archives-ouvertes.fr>
- <https://liris.cnrs.fr/publications/publications-liris>
- <https://liris.cnrs.fr/page-membre/remy-cazabet>

# API HAL

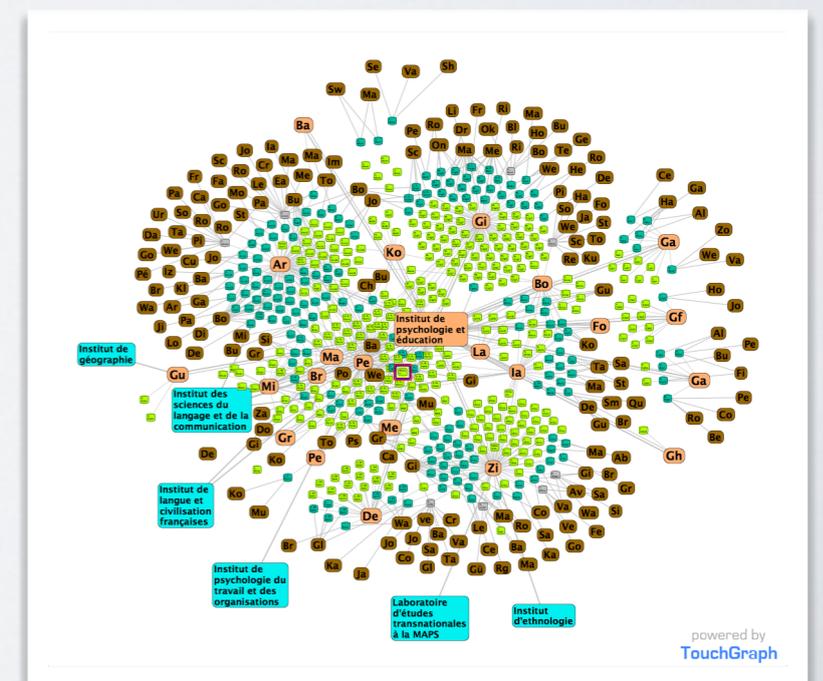
- API: Application Programming Interface
  - Un moyen offert par un logiciel, une base de données, un site web, etc. De communiquer avec lui
- HAL permet à n'importe qui de lui faire des requêtes, et renvoie le résultat
  - Faire une requête : simplement visiter une url (adresse web)
  - Ex : <https://api.archives-ouvertes.fr/search/?q=cazabet&wt=xml>
  - Documentation : <https://api.archives-ouvertes.fr/docs>

# API HAL

- Possible de récupérer les données tout simplement en .csv
- [https://api.archives-ouvertes.fr/search/?q=cazabet&wt=csv&rows=6&fl=title\\_s,publicationDate\\_s,keyword\\_s,structName\\_s](https://api.archives-ouvertes.fr/search/?q=cazabet&wt=csv&rows=6&fl=title_s,publicationDate_s,keyword_s,structName_s)
- Plus qu'à copier-coller le résultat dans un fichier et l'ouvrir avec un tableur

# API->RÉSEAU

- Nous voulons créer des réseaux:
  - ▶ Réseau de co-auteurs:
    - Comment est organisée la recherche autour d'un sujet, dans un laboratoire, etc.
  - ▶ Réseau d'institutions:
    - Coopération autour d'un domaine de recherche, entre laboratoires, universités, villes...
  - ▶ On pourrait imaginer d'autres possibilités : réseaux de journaux , de mots-clés...
- Objectif:
  - ▶ Construire une cartographie de l'état actuel
  - ▶ Analyser l'évolution
  - ▶ Identifier des acteurs clefs
  - ▶ Etc. (Seule limite est votre imagination ...)





# API->RÉSEAU

- Problème : passer de données tabulaires (liste d'auteurs, d'institutions, etc.) à un réseau
  - Lorsque 2 personnes/institutions apparaissent dans un même article, on considère qu'elles sont connectées
  - Éventuellement avec des seuils limites, pour ne garder que les plus importants
    - Combien de travaux commun pour être un "collaborateurs" ?
- Solution : utiliser un peu de code informatique
  - Python

# CODE PYTHON

- Je vous fournis un code
  - L'objectif est de comprendre ce qu'il fait et comment il fonctionne
  - L'objectif n'est pas de savoir le réécrire vous-même
  - Pour ceux avec un peu d'expérience en programmation, vous devez pouvoir le modifier pour faire des réseaux un peu différent (journaux...)

# CODE PYTHON

- Python:

- ▶ Language de programmation considéré comme l'un des plus "simple"
- ▶ Enseigné aujourd'hui au lycée
- ▶ Majoritaire dans l'analyse de données et l'intelligence artificielle
- ▶ Language orienté bibliothèques(library) : on installe des "outils" que l'on peut utiliser dans notre code (manipulation de graphes...)

- Notebooks:

- ▶ Outil pour écrire du code de manière interactive : Chaque ligne peut être exécutée au fur et à mesure
- ▶ Sur vos machines : installer <https://www.anaconda.com>  
(python+notebook+toutes librairies standard)

- Google colab (<http://colab.research.google.com>)

- ▶ Notebook en ligne qui permet d'exécuter du code sans installation

# PROJET

- Par groupes de 2 ou 3
- Une analyse réseau d'un domaine d'étude scientifique :
  - Des données extraites de HAL
  - Une visualisation dans Gephi
  - Une analyse selon les outils de l'analyse de réseau
- Rendu : un rapport (PDF) contenant obligatoirement
  - 1 ou plusieurs illustrations de réseaux
  - 1 ou plusieurs requêtes faites à l'API (sauf si sources externes)
  - Analyses (quantitatives/qualitatives/interprétations libres...)
- Prochaine séance=>Je serai votre support informatique

# CODE

- To the code !

<http://cazabetremy.fr/Teaching/IMST/IMST-EISO.html>

# EXEMPLE D'ANALYSE

- <https://evaluation-de-la-recherche.com/reseaux-coauteurs/>
- [https://images.cnrs.fr/photo/20210091\\_00025](https://images.cnrs.fr/photo/20210091_00025)