

# Intelligence Artificielle

## un outil au service de l'investigation

Nelly Barret, Simon Ebel, Théo Galizzi, Ioana Manolescu, Madhulika Mohanty

Inria Saclay & Institut Polytechnique de Paris



# Qui sommes nous ?

Nous faisons partie de l'**équipe de recherche CEDAR** (Rich Data Analytics at Cloud scale) à **Inria Saclay**.



**Nelly Barret**  
Doctorante  
Inria & IPP



**Simon Ebel**  
Ingénieur  
recherche  
Inria



**Théo Galizzi**  
Ingénieur  
recherche  
Inria

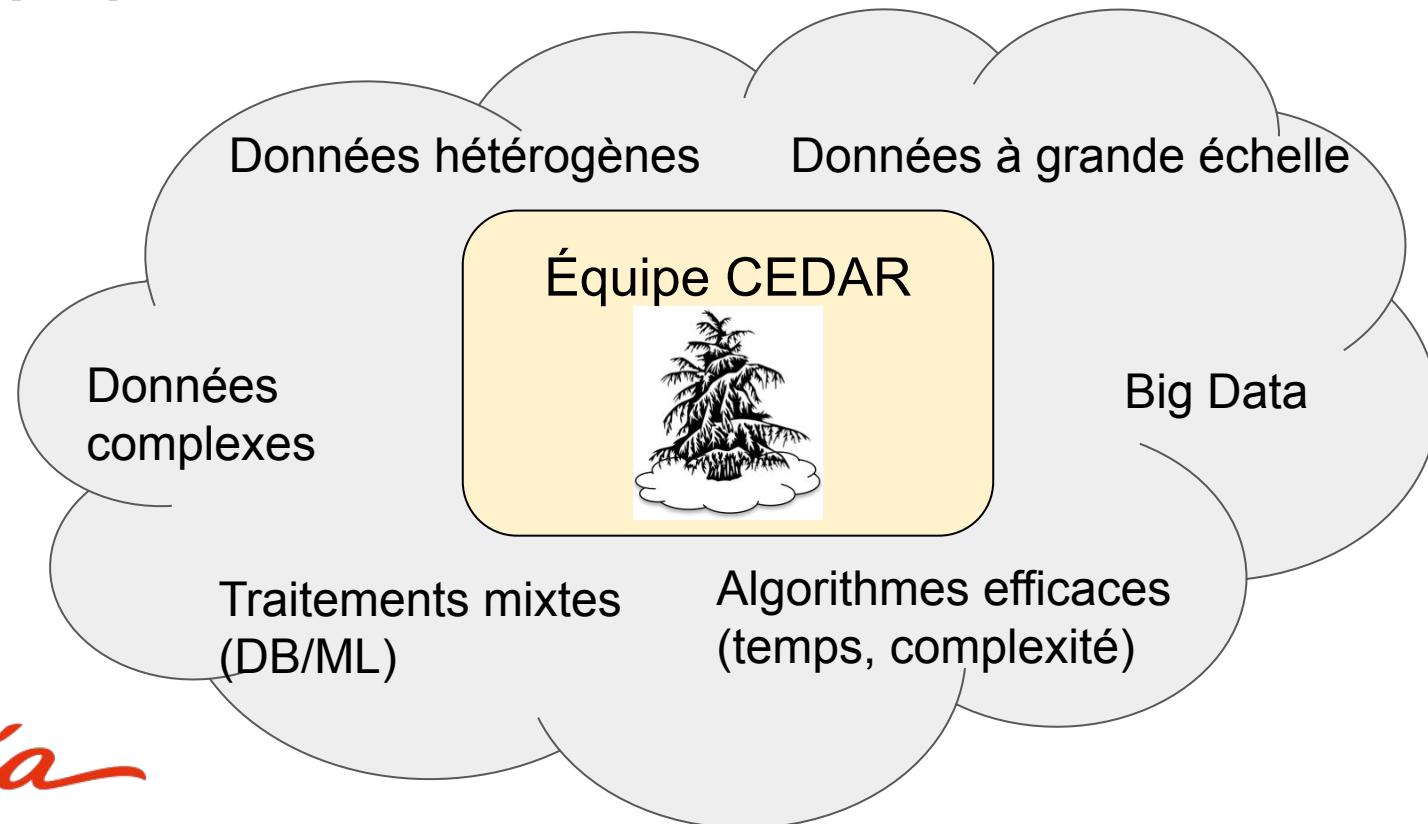


**Ioana Manolescu**  
Chercheure et chef  
d'équipe  
Inria & IPP

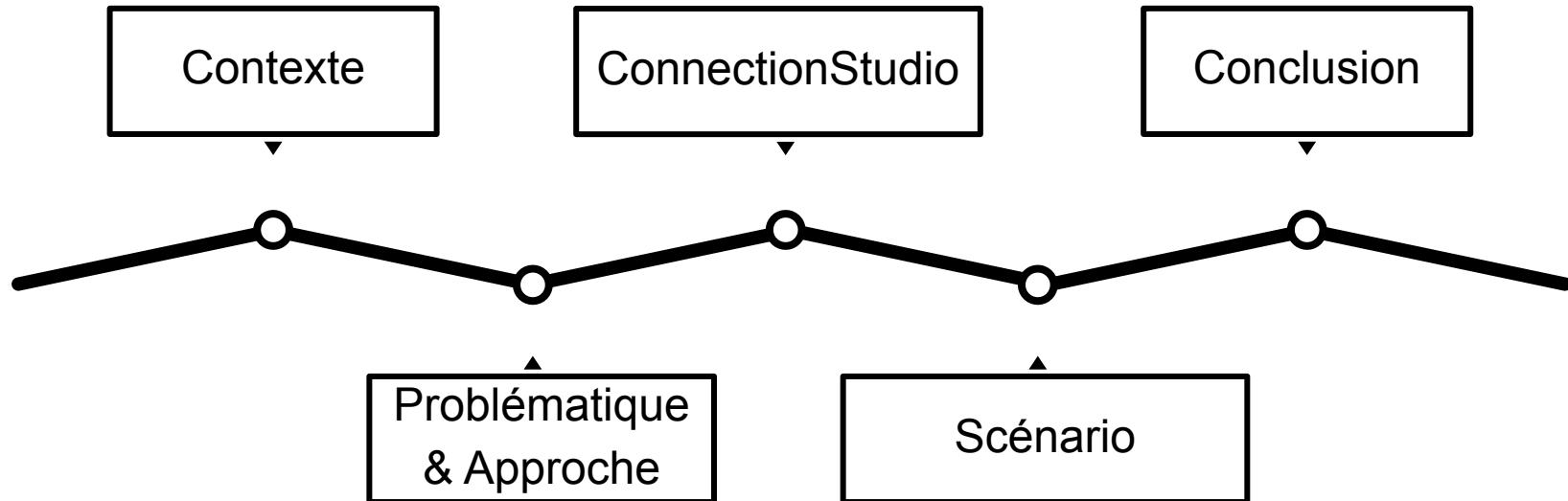


**Madhulika  
Mohanty**  
Post-doctorante  
Inria

# L'équipe CEDAR



# Organisation de l'atelier



# Contexte

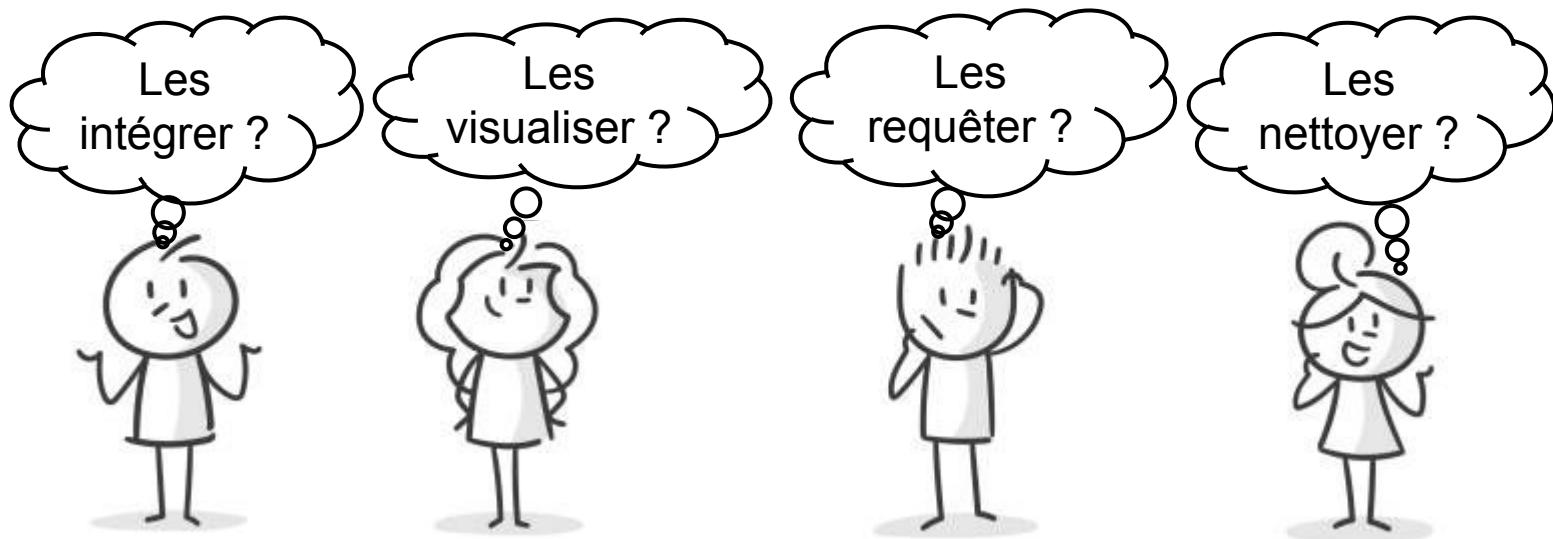
- Les “faits” proviennent de **différentes sources** de données, souvent **larges et complexes**
- Les sources sont de **formats très variés** :

Structurées	Semi-structurées	Non-structurées
<ul style="list-style-type: none"><li>- Bases de données</li><li>- Tables (Excel, CSV)</li></ul>	<ul style="list-style-type: none"><li>- JSON (Web)</li><li>- XML (Web)</li><li>- HTML (Web)</li></ul>	<ul style="list-style-type: none"><li>- Texte</li><li>- PDF</li><li>- RDF (<i>Open Data</i>)</li><li>- Neo4j (<i>Graphe</i>)</li></ul>

- Chaque source vient d'un *producteur* → **pas d'uniformisation ni de schéma**
- Les données ne sont pas produites en pensant à l'**utilisateur final**

# Problématique & approche

**Comment faciliter l'investigation de sources très hétérogènes ?**



# Problématique & approche

**Comment faciliter l'investigation de sources très hétérogènes ?**

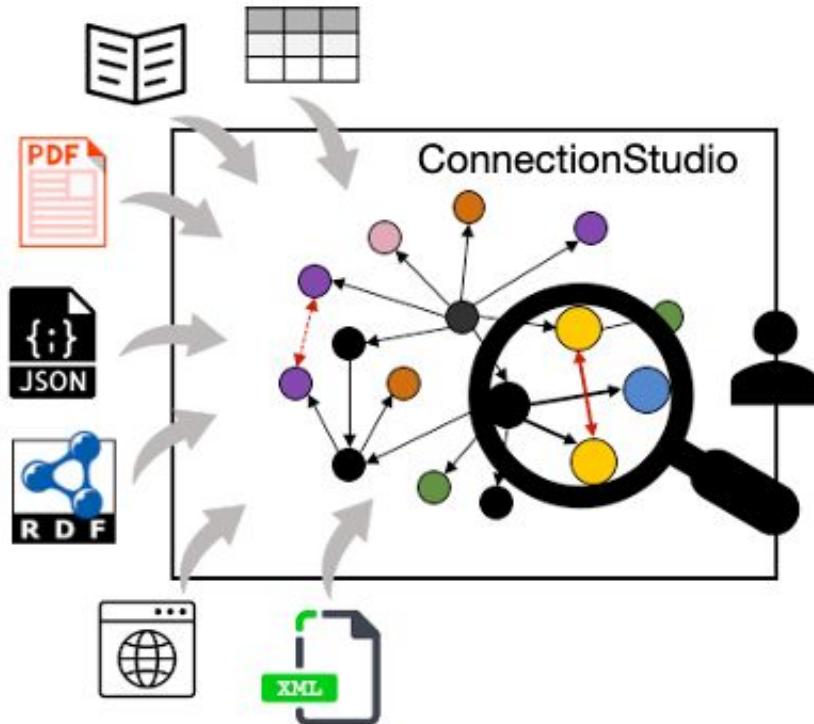
ConnectionStudio !



# ConnectionStudio

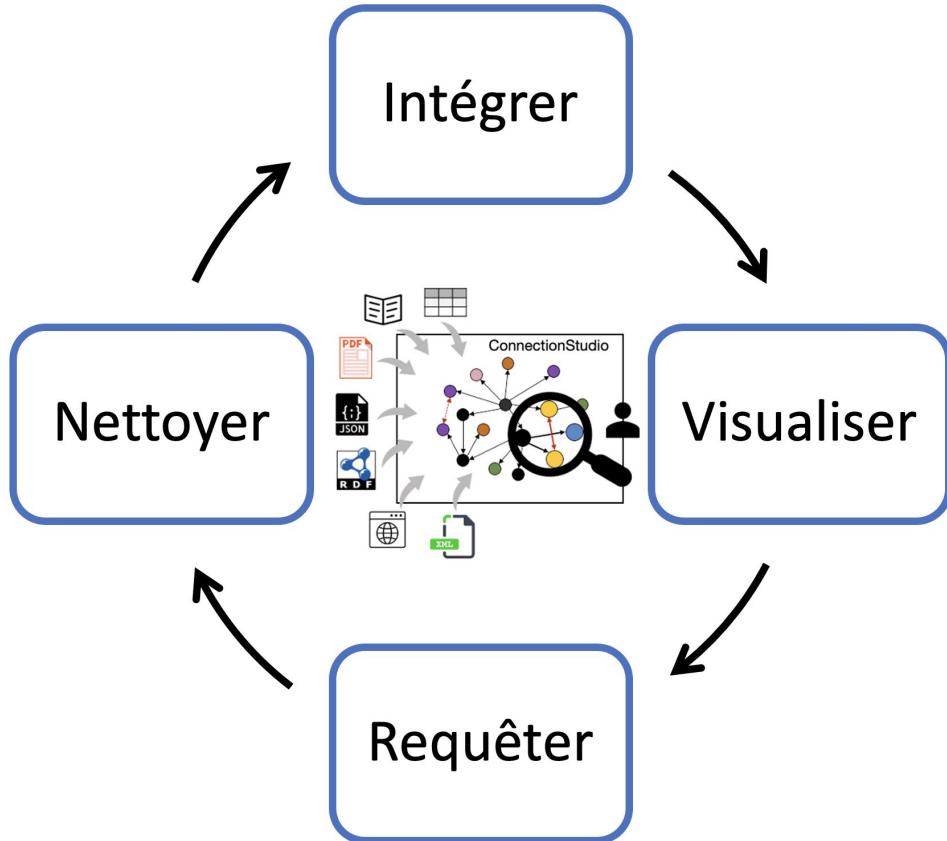
ConnectionStudio est un outil qui :

- Intègre des données hétérogènes de manière unifiée
- Identifie les entités : lieux, personnes, entreprises, ...
- Permet de visualiser et interagir avec les données



# Problématique & approche

**Comment faciliter  
l'investigation de  
sources très  
hétérogènes ?**



# ConnectionStudio

## Connection Studio Projets



Trier par  
Nom du projet ▾

TERMINER LA SESSION

CRÉER UN PROJET

### Projet Cac40

1 fichiers  
Créé le : 2023-07-05 16:12:38  
Dernier ajout : 2023-07-05 16:12:38

GÉRER

### Projet Hatvp Cac40

2 fichiers  
Créé le : 2023-07-05 15:46:07  
Dernier ajout : 2023-07-05 16:25:52

GÉRER

### Projet Hatvpsmall

Pas encore de fichier chargé, ajoutez-en !

GÉRER

### Projet Pubmed

1 fichiers  
Créé le : 2023-07-05 09:46:07  
Dernier ajout : 2023-07-05 09:46:07

GÉRER

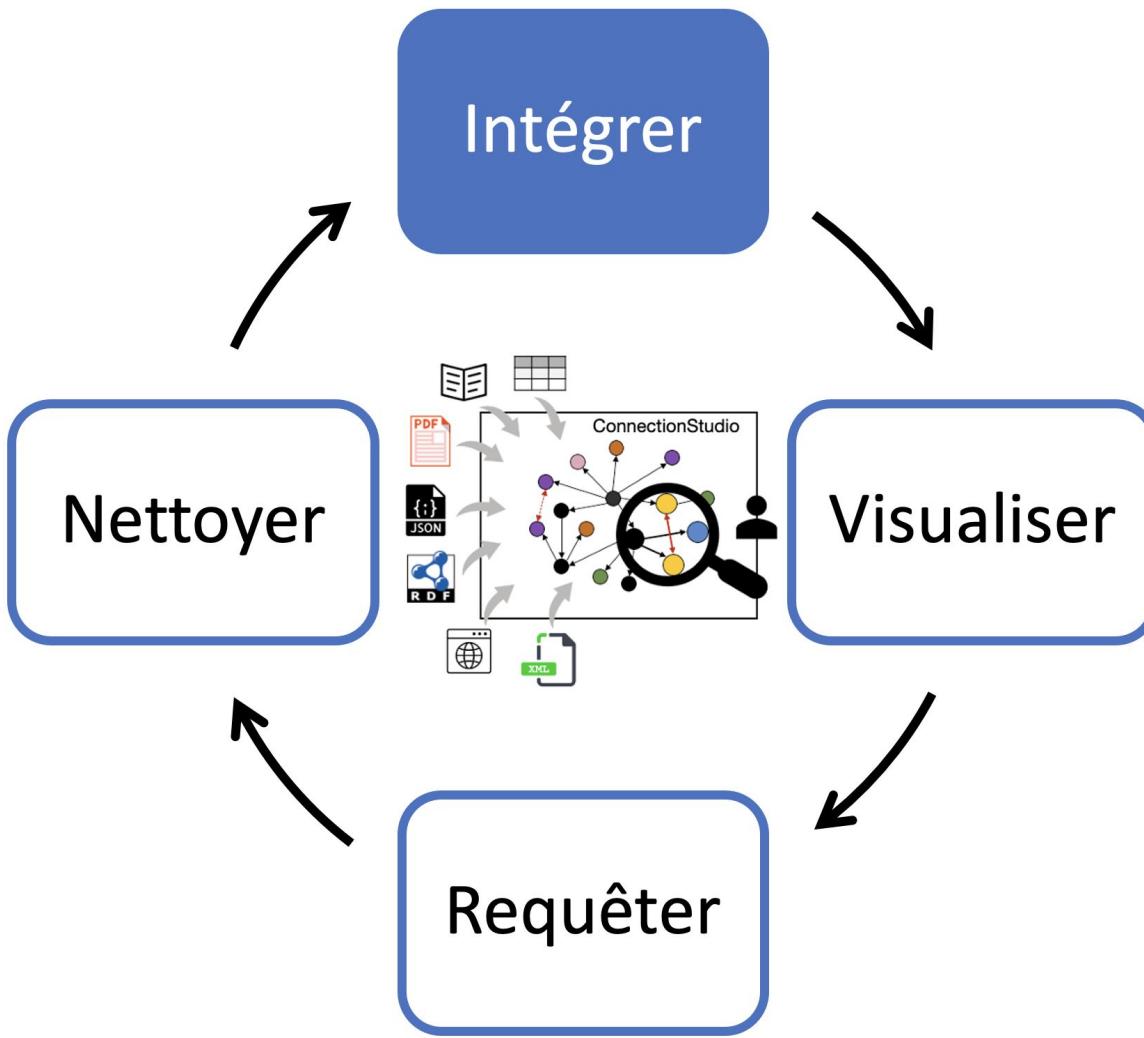
Les projets déjà créés

Un projet = tous les fichiers concernant la même thématique.

# Introduction aux données : CAC40

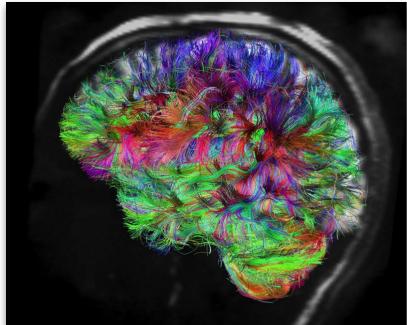
**CAC40** : 40 sociétés, de différentes branches d'activités, qui reflètent la tendance globale de l'économie des grandes entreprises françaises.

	A	B	C
1	numero	nom_entreprise	description_wikipedia
2	1	AIR LIQUIDE	Air liquide (AL), autrefois connue sous le nom L'Air liquide, est un groupe industriel français d'envergure internationale, spécialisé dans les gaz industriels, c'est-à-dire des gaz pour l'industrie, la santé, l'environnement et la recherche.
3	2	AIRBUS	Airbus est un constructeur d'avion franco-allemand et une coopération industrielle internationale présente dans le secteur aéronautique et spatial civil et militaire. Ses activités les plus importantes sont la construction d'avions de ligne, d'hélicoptères et d'avions militaires. À travers la participation à diverses entreprises, le groupe est engagé dans les lanceurs spatiaux, les satellites artificiels, les missiles et les avions de combat.
4	3	ALSTOM	Alstom est une multinationale française, aujourd'hui spécialisée dans le secteur des transports, principalement ferroviaires (trains, tramways et métros). Alstom faisait partie du groupe Alcatel-Alsthom, nouveau nom de

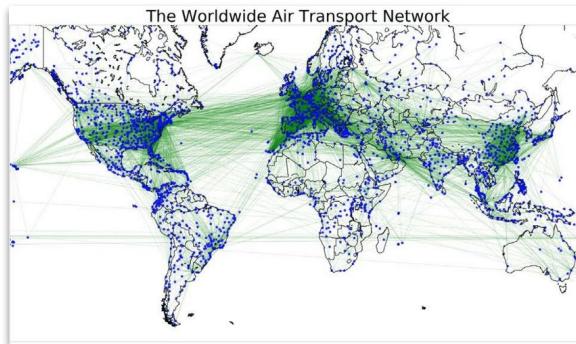


# Graphe

- Le **paradigme “graphe”** décrit :
  - Des objets ( *noeuds*)
  - Connectés par des liens (*arêtes*)
- Beaucoup utilisé car très **flexible**



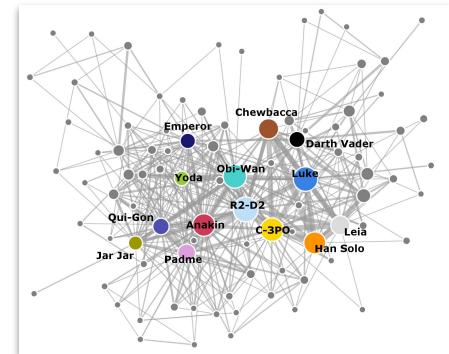
Neurones du cerveau humain



Vols internationaux



Panama papers (données ICIJ)



Personnages dans “Star Wars”

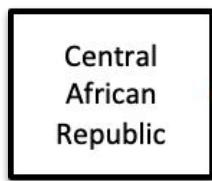
# Extraction d'entités

- Extraction d'entités simples :
  - Email, date, site Web, ...
- Extraction d'entité nommées
  - Expression linguistique référentielle: personnes, lieux, organisations, ...
  - Complexe à identifier → fait par des modèles de langage

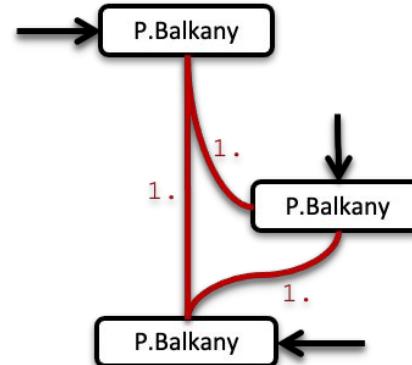
In December 1903 **DATE** the Royal Swedish Academy of Sciences **ORG** awarded  
Marie **PERSON** and Pierre Curie **PERSON**, along with Henri Becquerel **PERSON**,  
the Nobel Prize in Physics **WORK\_OF\_ART**.

# Annotation sémantique

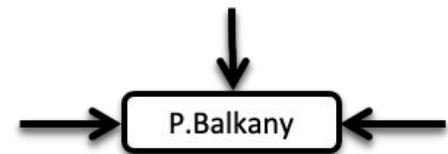
- Connexion entre les entités nommées quand elles sont *similaires*
- Plusieurs similarités : exactes, approximatives
- Unification possible quand similarité totale → - de noeuds et + de connexions



Similarité approximative



Similarité exacte

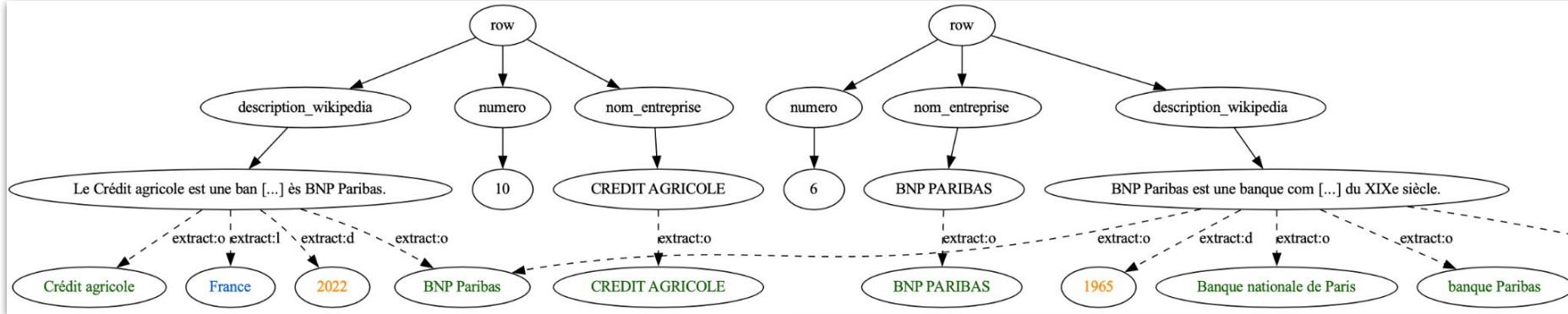


Fusion des entités identiques

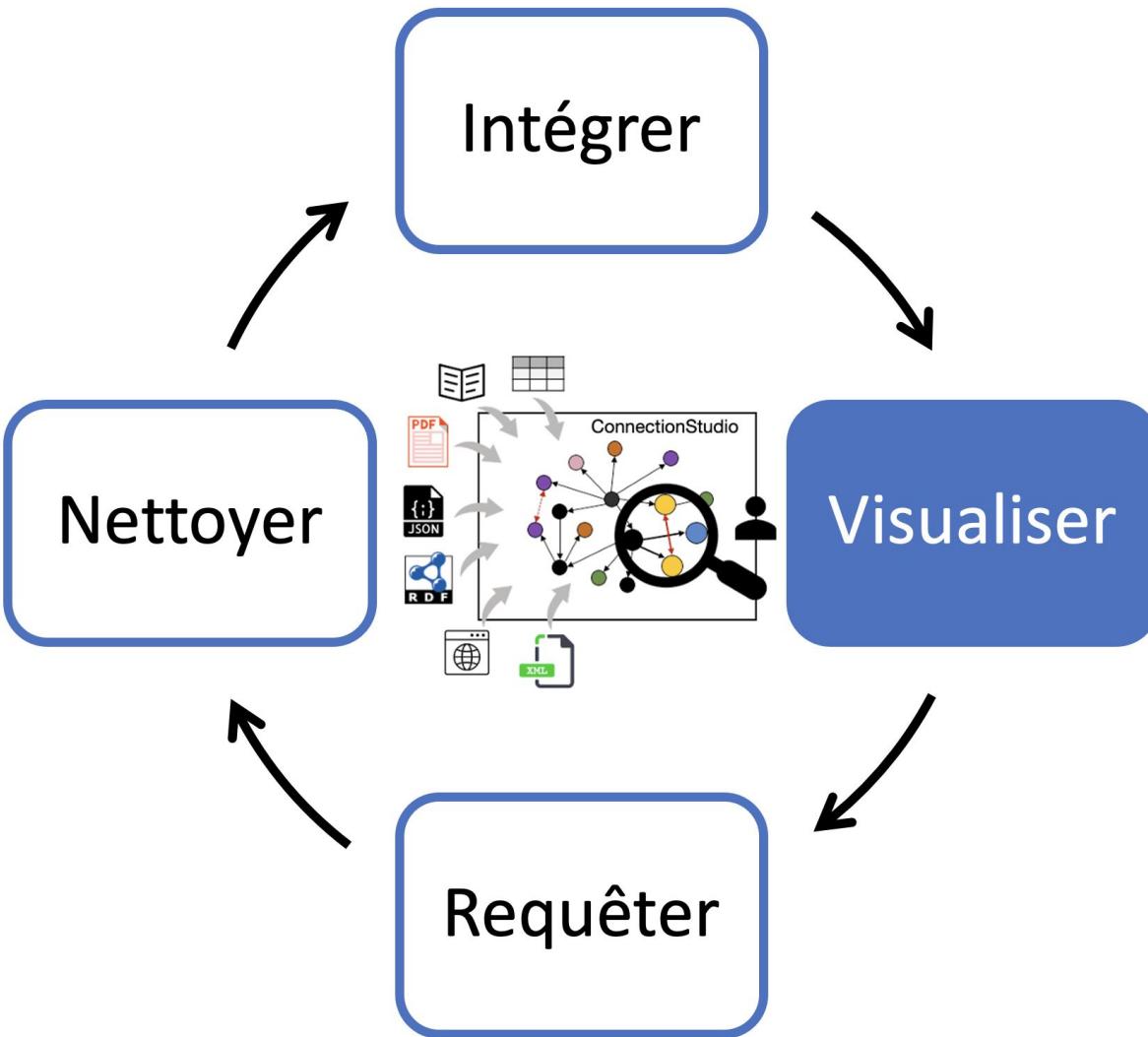
# Chargement des sources de données

- Les données sont modélisées selon le paradigme “graphe”:
  - Un **noeud** est une **petite portion d'information**
  - Les **liens** connectent les noeuds pour reconstituer la **sémantique logique** du fichier
- L'extraction d'entités nommées est appliquée sur chaque noeud valeur
  - 2 modèles linguistiques : celui de Stanford et celui de FLAIR
  - 2 langues : français ou anglais
- Les noeuds qui contiennent des informations identiques sont fusionnés → connexions entre les sources

# Introduction aux données : CAC40



Graphe de données correspondant au CSV du CAC40



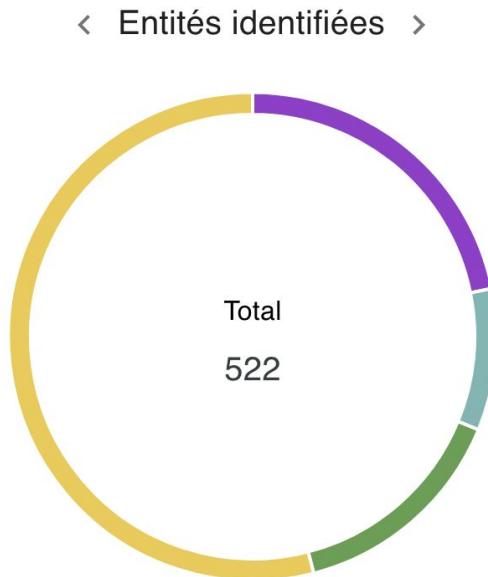
# Visualisation 1 : statistiques

Comment avoir une bonne idée des entités détectées ?

- Distribution d'entités par types
- Nuage des 100 entités les plus fréquentes (*tag cloud*)
- Distribution des types d'entités par fichier
- Liste des 100 entités les plus partagées parmi le plus de sources
- Tous les graphiques/tableaux sont exportables

# Visualisation 1 : statistiques

Distribution d'entités par type



- Nombre de dates
- Nombre de Personnes
- Nombre de Places
- Nombre d'Organisations

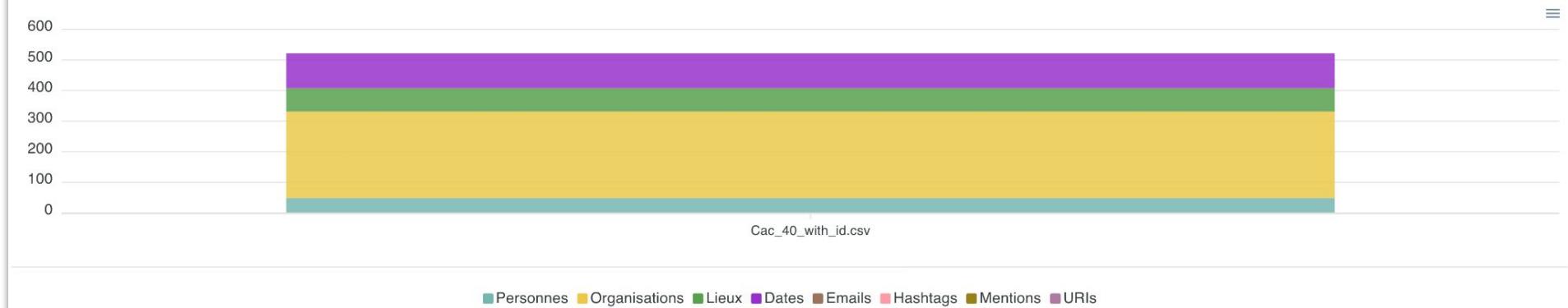
# Visualisation 1 : statistiques

NUAGE D'ENTITÉS



# Visualisation 1 : statistiques

Distribution des types d'entités par fichier



# Visualisation 1 : statistiques

## Entités communes

COLONNES				FILTRES	DENSITÉ	EXPORTER
Label	Type	Frequency ↓	Datasets			
Paris	Lieu	11	Cac_40_with_id.csv			
France	Lieu	9	Cac_40_with_id.csv			
Bourse de Paris	Organisation	6	Cac_40_with_id.csv			
Europe	Lieu	5	Cac_40_with_id.csv			
Luxembourg	Lieu	3	Cac_40_with_id.csv			
Orange	Lieu	2	Cac_40_with_id.csv			
La Défense	Lieu	2	Cac_40_with_id.csv			
Rueil-Malmaison	Lieu	2	Cac_40_with_id.csv			

# Visualisation 2 : structure d'un fichier

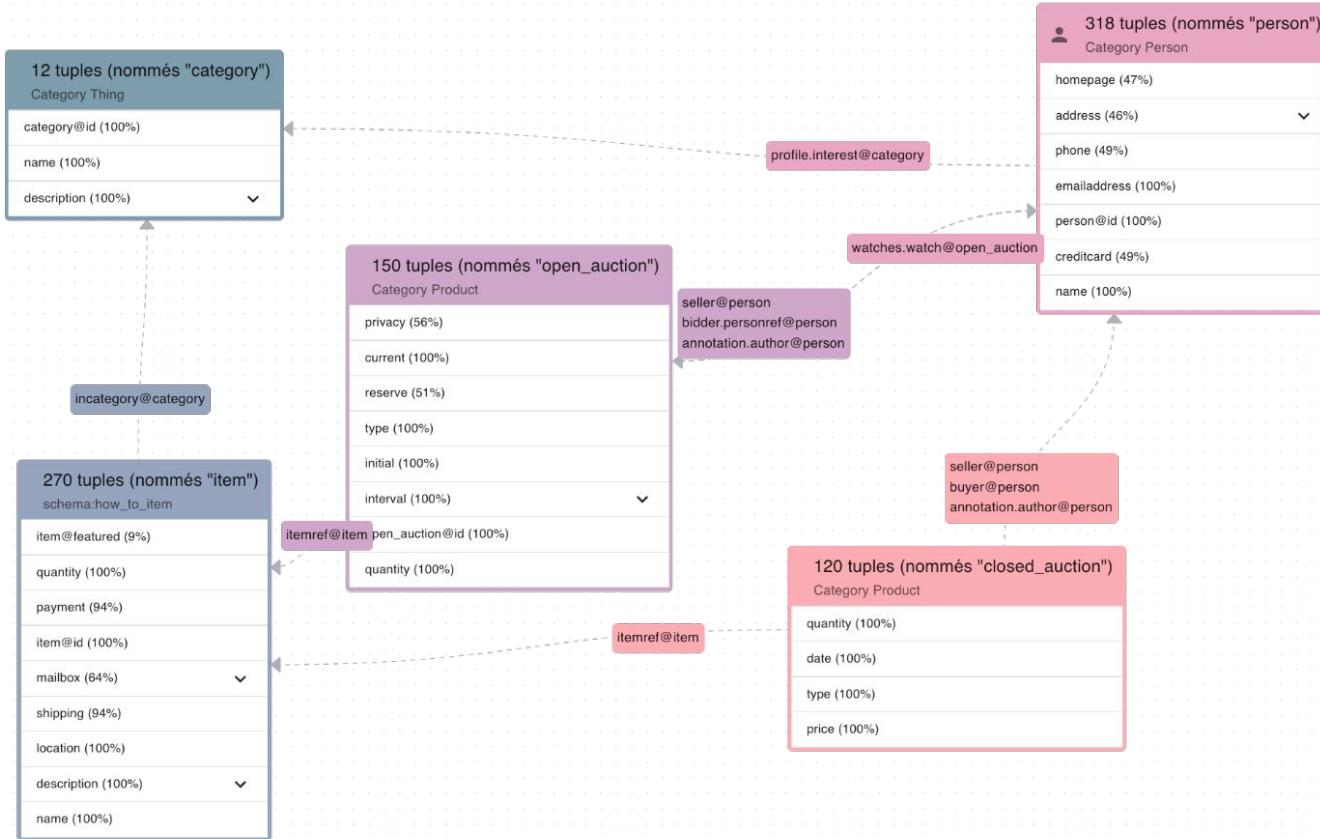
Comment avoir une bonne idée de la structure d'un fichier ?

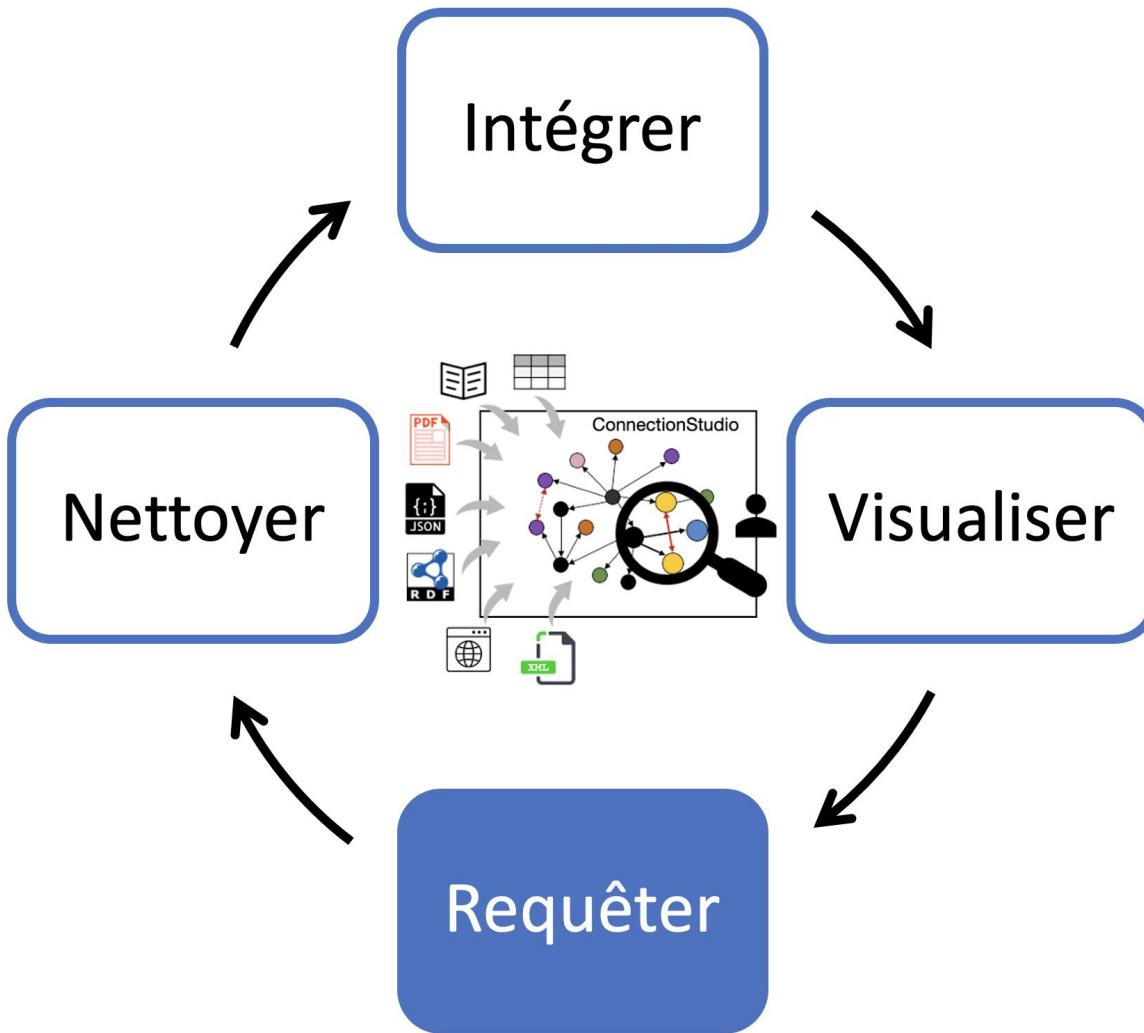
- Une source de données peut être vue comme :
  - Un ou plusieurs ensembles d'entités similaires
  - Un ensemble de relations qui les connectent
- Résultat : une **description orientée utilisateur**

# Visualisation 2 : structure d'un fichier

The screenshot shows the Connection Studio interface, specifically the 'Abstraction de données' (Data Abstraction) section. The title bar reads 'Connection Studio Abstraction de données'. On the left, there's an 'Explorer' sidebar and a 'FICHIERS' (Files) panel where 'Cac\_40\_with\_id.csv' is selected from a list of files (JSON, XML, CSV, RDF). In the center, there's a large grid representing the data structure. A tooltip box highlights the first row of the grid with the text: '40 tuples (nommés "row") Category Thing'. Below this, three columns are listed: 'description\_wikipedia (100%)', 'nom\_entreprise (100%)', and 'numero (100%)'. The bottom of the grid has icons for zooming in and out, and for saving as SVG or PNG.

# Visualisation 2 : structure d'un fichier





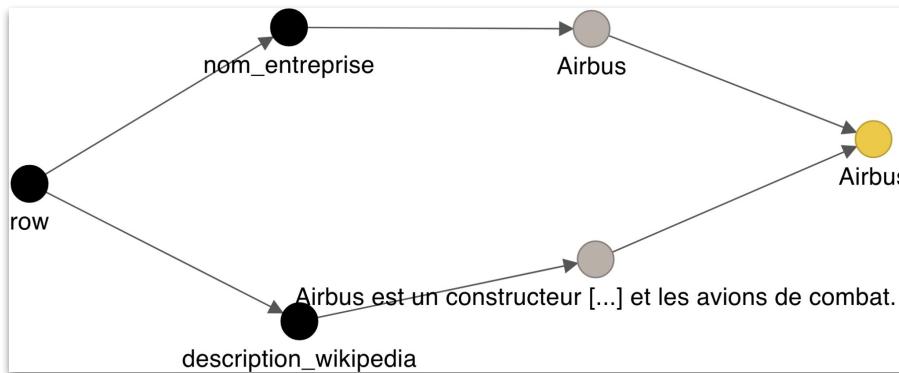
# Requêtes

- **Requête** : interrogation des données pour en récupérer un sous-ensemble
- 2 moyens :

	Par <b>mot(s)-clé(s)</b>	Avec le <b>langage SQL</b>
Requiert	un ou plusieurs mots clés	une requête SQL
Donne	parties du graphe qui contiennent les mots clés	tableau de résultats
Adapté quand/pour	on connaît les mots-clés qui nous intéressent	explorer de grandes données

# Requêtes

	Par <b>mot(s)-clé(s)</b>	Avec le <b>langage SQL</b>
Requiert	un ou plusieurs mots clés	une requête SQL
Donne	parties du graphe qui contiennent les mots clés	tableau de résultats
Adapté quand/pour	on connaît les mots-clés qui nous intéressent	explorer de grandes données



Requête “Airbus” sur le graphe de données

```
cl_cac40=# SELECT id, label  
FROM nodes  
WHERE label LIKE 'Airbus%';  
+-----+-----+  
| id   | label |  
+-----+-----+  
| 20   | Airbus |  
| 24   | Airbus est un constructeur... |  
| 628  | Airbus |  
(3 rows)
```

Requête “Airbus” en SQL

# Requête 1 : Connexions entre entités

Comment trouver comment sont connectées les entités ?

Identifie toutes les connexions entre deux types d'entités :

- Personnes ↔ Personnes
- Personnes ↔ Lieux
- Entreprises ↔ Personnes
- ...

# Requête 1 : Connexions entre entités

Résultats

X

COLONNES	FILTRES	DENSITÉ	EXPORTER	ETENDRE LE TEXTE
nom_entreprise#val	nom_entreprise	row	description_wikipedia	description_wikipedia#val ▾
ArcelorMittal	nom_entreprise	row	description_wikipedia	Europe
ArcelorMittal	nom_entreprise	row	description_wikipedia	Luxembourg
Engie	nom_entreprise	row	description_wikipedia	Bruxelles
Engie	nom_entreprise	row	description_wikipedia	Luxembourg
Engie	nom_entreprise	row	description_wikipedia	Paris
Eurofins	nom_entreprise	row	description_wikipedia	Luxembourg
Eurofins	nom_entreprise	row	description_wikipedia	Nantes

# Conflits d'intérêts dans le bio-médical

The screenshot shows the Connection Studio interface, specifically the 'Exploration de chemins' (Pathway Exploration) section. The title bar includes 'Explorer Connection Studio Exploration de chemins' and 'Projet: Pubmed'. The main area is titled 'Résultats' (Results) and displays a table of 2454 rows. The columns are labeled: Name#val, Name, Author, AuthorList, PubmedArticle, CoiStatement, and CoiStatement#val. The first two rows for 'A. Coates' are highlighted in green boxes. The last two rows for 'A. Nishiyama' are also highlighted in green boxes. The 'CoiStatement' column contains several yellow callout boxes with organization names: 'CP', 'Helperby Therapeutics', 'Japan Society for the promotion of Science', 'Grant-in-Aid for Scientific Research B of Japan Society for the Promotion of Science', and 'Grant-in-Aid for Scientific Research C of Japan Society for the Promotion of Science'. At the bottom, there are pagination controls: 'Lignes par page : 10', '1-10 sur 2454', and navigation arrows.

Name#val	Name	Author	AuthorList	PubmedArticle	CoiStatement	CoiStatement#val
A. Coates	Name	Author	AuthorList	PubmedArticle	CoiStatement	CP
A. Coates	Name	Author	AuthorList	PubmedArticle	CoiStatement	Helperby Therapeutics
A. Nagano	Name	Author	AuthorList	PubmedArticle	CoiStatement	Japan Society for the promotion of Science
A. Nishiyama	Name	Author	AuthorList	PubmedArticle	CoiStatement	Grant-in-Aid for Scientific Research B of Japan Society for the Promotion of Science
A. Nishiyama	Name	Author	AuthorList	PubmedArticle	CoiStatement	Grant-in-Aid for Scientific Research C of Japan Society for the Promotion of Science

Lignes par page : 10 ▾ 1-10 sur 2454 < >

Navigation icons at the bottom: Affiliation#val, Affiliation, Author, AuthorList, PubmedArticle, CoiStatement, CoiStatement#val, (252 chemins)

# Requête 2 : recherche par mots-clés

Comment explorer les données via des mots-clés ?

- Chaque mot-clé → un noeud dans le graphe de données
- Résultats → connexions entre ces noeuds
- Paramètres

Adapté pour :

- Chercher certains mots-clés
- Explorer le graphe sans en connaître la structure

# Requête 2 : recherche par mots-clés

Quoi ?	Comment ?	Résultat
Exactement un mot	exact:Airbus	Un noeud exactement labellisé <b>Airbus</b>
Un mot	Airbus / "Airbus"	Un noeud qui contient : - <b>Airbus</b> (non-sensible à la casse) - " <b>Airbus</b> " (sensible)
Exploration par mot-clé	explore Capgemini	Les noeuds labellisés <b>Capgemini</b> ainsi que leurs voisins
Exactement plusieurs mots	exact:CREDIT+AGRICOLE	Les noeuds labellisés exactement et uniquement <b>CREDIT AGRICOLE</b>
Plusieurs mots	"Airbus Engie"	Les noeuds qui contiennent <b>Airbus</b> et <b>Engie</b> (dans n'importe quel ordre)
Exploration par type	type:Person	Tous les noeuds de type Personne

# Requête 2 : recherche par mots-clés

Information du Noeud X

Airbus

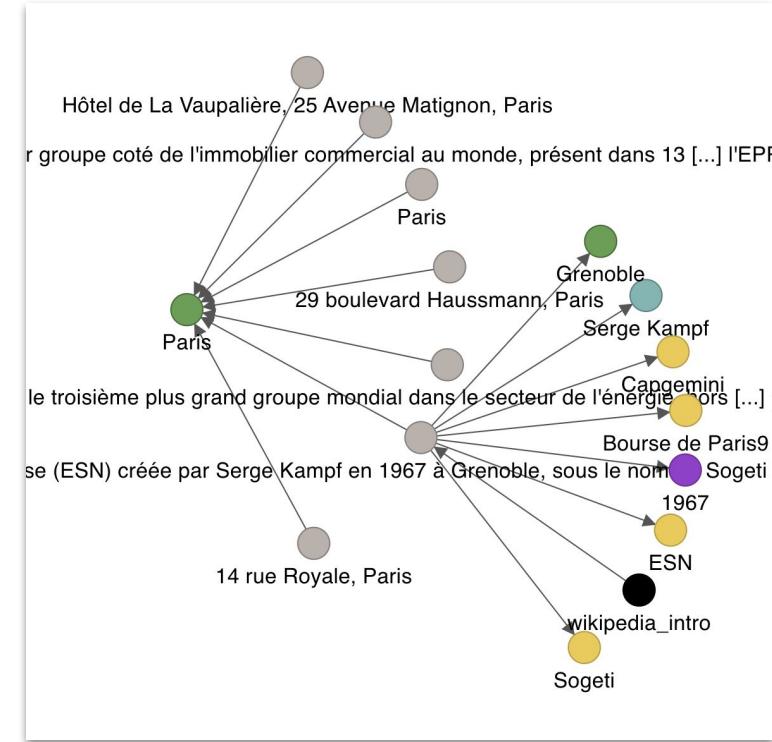
Noeud **Airbus** de type **Organisation** extrait depuis le fichier **Cac\_40\_with\_id\_actionnaires\_and\_filiales.csv**

Identifiant interne: 2073

Un noeud et ses informations



Les noeuds “Airbus”



Exploration manuelle

# Requête 2 : recherche par mots-clés

CREDIT AGRICOLE

## Information du Noeud



Noeud **CREDIT AGRICOLE** de type  
**Organisation** extrait depuis le fichier  
**Cac\_40\_with\_id\_actionnaires\_and\_filiales.csv**  
Identifiant interne: 2618

Un noeud et ses informations

Des noeuds de type "Personne"

Ernest Mercier

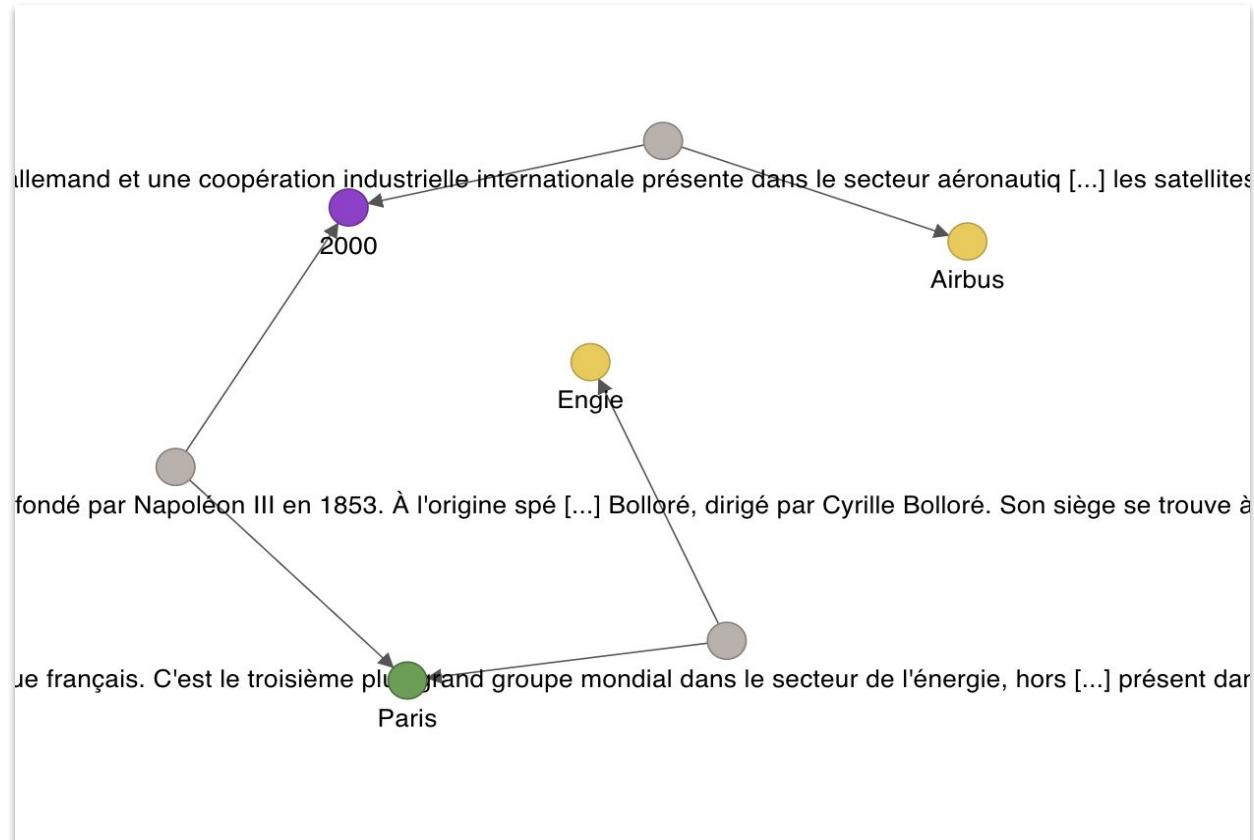
Vinci

Badinter

Hermès

# Requête 2 : recherche par mots-clés

Connexions entre les noeuds labellisés "Airbus" et "Engie" respectivement



# Requête 3 : Requêtes utilisateur

Comment requêter les données de manière personnalisée ?

- **Explorer la base de données sans compétences en SQL**
  - SQL : langage utilisé pour interagir avec une base de données
  - Syntaxe spécifique à apprendre → bloquant pour les utilisateurs
- Liste de fragments de données (↔ SELECT de SQL)
- Interconnectables entre eux de manière personnalisée (↔ JOIN de SQL)

```
SELECT magasin, pays,  
COUNT(vente)  
FROM magasins, ventes  
WHERE date='12/07/2023'  
GROUP BY magasin, pays  
HAVING COUNT(vente) > 1000€
```

Exemple de requête SQL

# Requête 3 : Requêtes utilisateur

The screenshot shows the Connection Studio interface with the following details:

- Header:** Connection Studio Vue des données
- File Selection:** Sélectionner un fichier: Cac\_40\_with\_id.csv
- Path Selection:** Sélectionner un chemin: row.nom\_entreprise.#val
- Request Bar:** Afficher la requête, EVALUER LA REQUÊTE (highlighted in blue), SAUVER LES MODIFICATIONS
- Request Editor:** Chemin 1: row.nom\_entreprise.#val
- Toolbar:** COLONNES, FILTRES, DENSITÉ, EXPORTER
- Data Preview:** A table showing the results of the query, with columns var0 and var1. The data includes:

var0	var1 ↑
1	air liquide
14	airbus
27	alstom
40	arcelorittal
53	axa
66	bnp paribas
79	bouygues
92	capgemini
105	carrefour
118	crédit agricole

# Requête 3 : Requêtes utilisateur

Connection Studio    localhost:3000/cl\_cac40/view    100%    Projet: Cac40

## Connection Studio Vue des données

Explorer

Sélectionner un fichier  
Cac\_40\_with\_id.csv

Sélectionner un chemin  
row.nom\_entreprise.#val.extract:o

Chemin 1  
row.description\_wikipedia.#val

Chemin 2  
row.description\_wikipedia.#val.extract:o

Chemin 3  
row.nom\_entreprise.#val.extract:o

Afficher la requête    EVALUER LA REQUÊTE    SAUVER LES MODIFICATIONS

Variable de départ ligne    Variable d'arrivée texte

Jointure Requis    Optionnel

Variable de départ ligne    Variable d'arrivée entreprise

Jointure Requis    Optionnel

Variable de départ ligne2    Variable d'arrivée entreprise

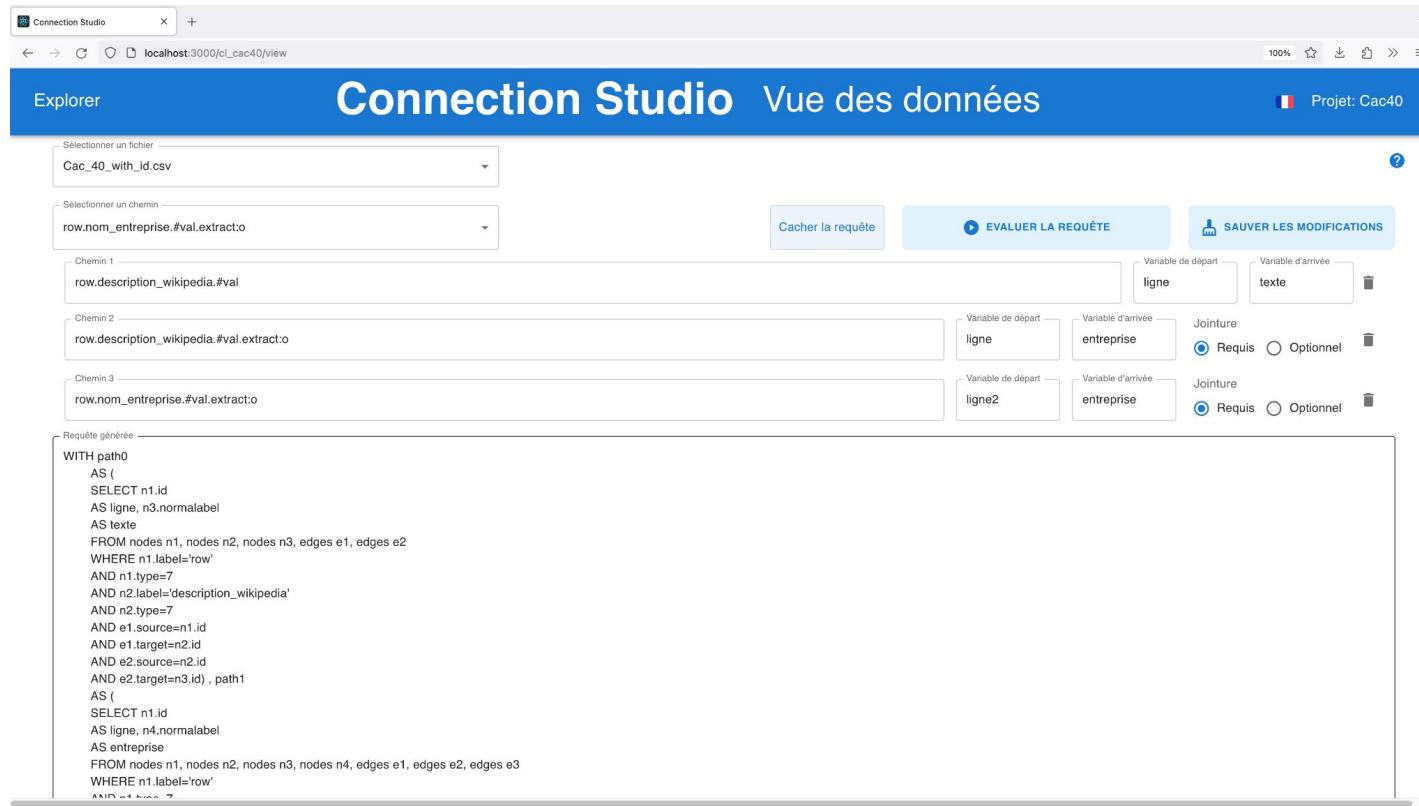
Jointure Requis    Optionnel

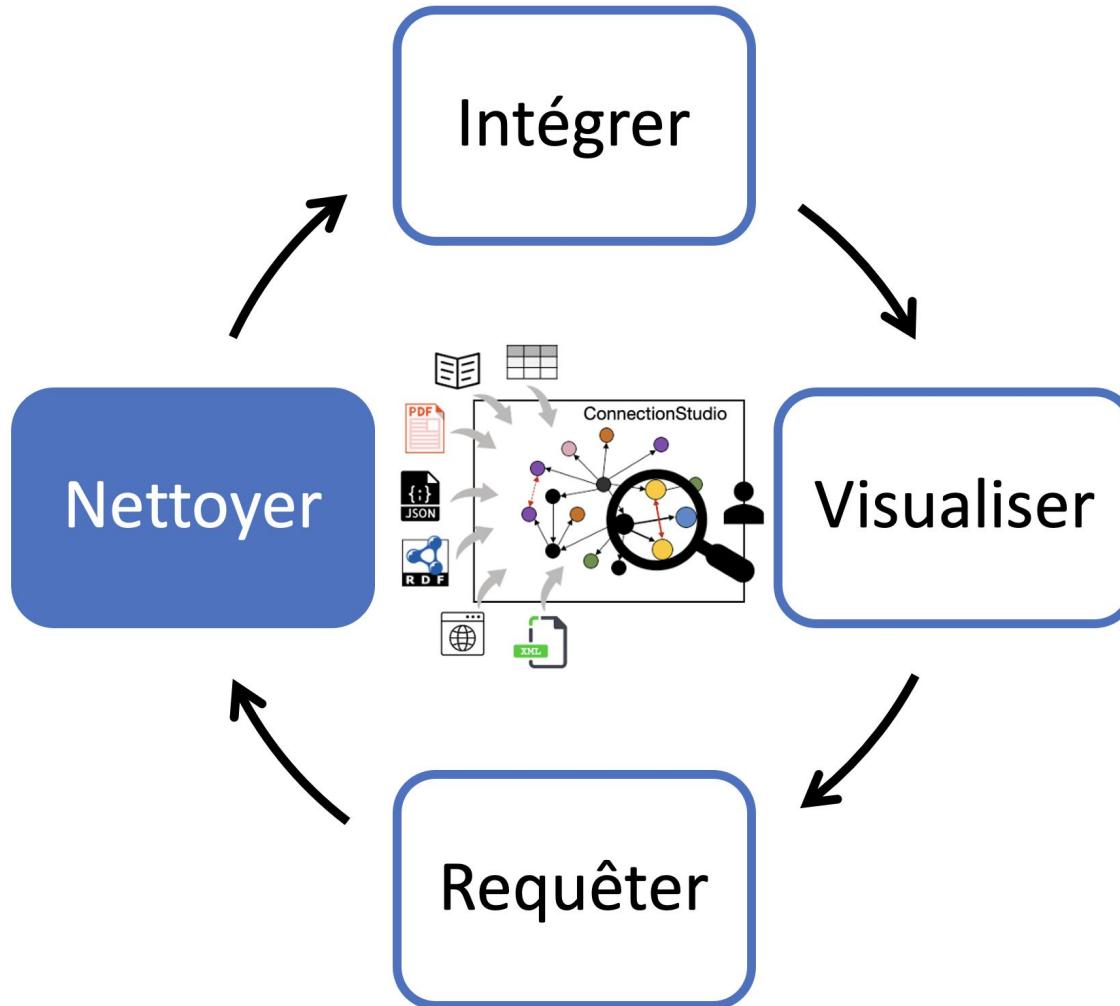
COLONNES FILTRES DENSITÉ EXPORTER

ligne	entreprise	ligne2
66	bpn paribas est une banque commerciale française elle est la première banque européenne par son activité et sa rentabilité avec 3 080 milliards \$ d'actifs et le 8e groupe bancaire international présent dans 65 pays il est coté au premier marché d'euronext paris et fait partie de l'indice cac 40 au 31 décembre 2022 le bénéfice net part du groupe s'élève à 10 2 milliards d'euros en hausse de 7 5 % par rapport à 202110 avec 193 000 employés en février 2023 la banque est organisée selon trois grands domaines d'activités : services bancaires pour particuliers et pour commerçants services d'investissement et de protection services bancaires pour entreprises et institutions le groupe est issu de la fusion en mai 2000 entre la banque nationale de paris banque née en 1965 de la fusion de l'ancienne banque nationale de crédit et du comptoir national d'escompte de paris et de la banque paribas établissement né au cours du xixe siècle	66
	crédit agricole est une banque française proposant des services financiers à l'échelle internationale le groupe possédant le plus grand réseau de banques coopératives et mutualistes au monde6 dont trente neuf caisses régionales de crédit agricole en france en 2022 le crédit agricole est la dixième plus grande banque au monde en termes d'actifs la deuxième française après bnp paribas	

crédit agricole est une banque française proposant des services financiers à l'échelle internationale le groupe possédant le plus grand réseau de banques coopératives et mutualistes au monde6 dont trente neuf caisses régionales de crédit agricole en france en 2022 le crédit agricole est la dixième plus grande banque au monde en termes d'actifs la deuxième française après bnp paribas

# Requête 3 : Requêtes utilisateur





# Nettoyage : politiques d'extraction

Comment nettoyer et uniformiser les données ?

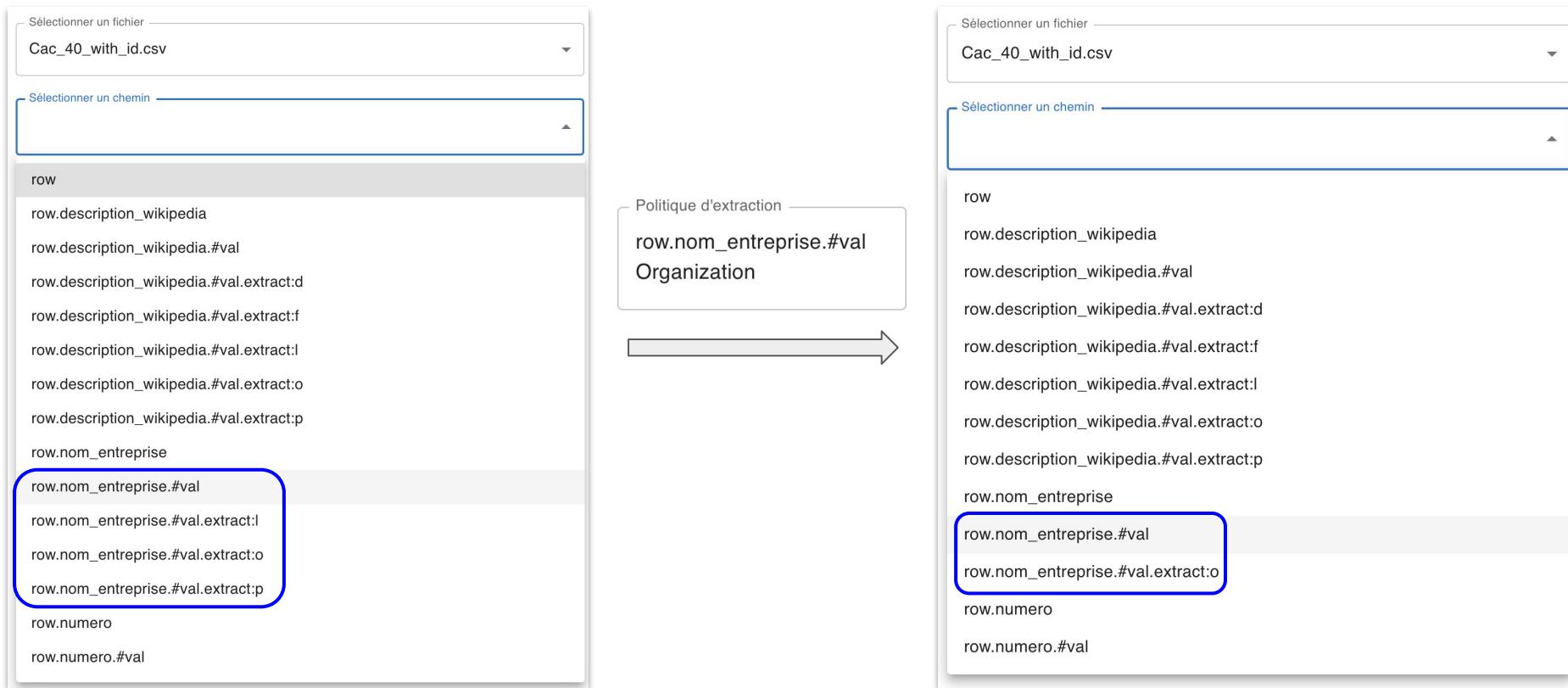
- Nettoyage des données semi-automatique :
  - Nettoyage lors de l'ingestion (i.e. *normalisation* des valeurs)
  - Nettoyage par l'utilisateur après chargement

**Politique d'extraction** : association d'un chemin et d'un type d'entité pour aider l'extraction d'entités

The screenshot shows a user interface for data cleaning. At the top, there is a dropdown menu labeled "Sélectionner un fichier" containing the file path "Cac\_40\_with\_id.csv". Below this is another dropdown menu labeled "Sélectionner un chemin". A list of extraction policies is displayed below the second dropdown, each consisting of a row identifier followed by a dot and a specific extraction rule. The policies listed are: row, row.description\_wikipedia, row.description\_wikipedia.#val, row.description\_wikipedia.#val.extract:d, row.description\_wikipedia.#val.extract:f, row.description\_wikipedia.#val.extract:l, row.description\_wikipedia.#val.extract:o, row.description\_wikipedia.#val.extract:p, row.nom\_entreprise, row.nom\_entreprise.#val, row.nom\_entreprise.#val.extract:l, row.nom\_entreprise.#val.extract:o, row.nom\_entreprise.#val.extract:p, row.numero, and row.numero.#val.

- Sélectionner un fichier — Cac\_40\_with\_id.csv
- Sélectionner un chemin
- row
- row.description\_wikipedia
- row.description\_wikipedia.#val
- row.description\_wikipedia.#val.extract:d
- row.description\_wikipedia.#val.extract:f
- row.description\_wikipedia.#val.extract:l
- row.description\_wikipedia.#val.extract:o
- row.description\_wikipedia.#val.extract:p
- row.nom\_entreprise
- row.nom\_entreprise.#val
- row.nom\_entreprise.#val.extract:l
- row.nom\_entreprise.#val.extract:o
- row.nom\_entreprise.#val.extract:p
- row.numero
- row.numero.#val

# Nettoyage : politiques d'extraction



# Nettoyage : modification des valeurs

Comment nettoyer et uniformiser les données ?

- Nettoyage des données semi-automatique :
  - Nettoyage lors de l'ingestion (i.e. *normalisation* des valeurs)
  - Nettoyage par l'utilisateur après chargement
- Possibilité de nettoyer les valeurs à la main
- Modifie les valeurs dans la base de données

# Nettoyage : modification des valeurs

The screenshot shows the Connection Studio interface for a project titled "Projet: Hatvp Cac". The main area displays a data cleaning process:

- Sélectionner un fichier:** Cac40.csv
- Sélectionner un chemin:** row.description\_wikipedia.#val.extract:o
- Chemin 1:** row.description\_wikipedia.#val.extract:o
- Actions bouton:** Afficher la requête, EVALUER LA REQUÊTE, SAUVER LES MODIFICATIONS

Below the interface, a preview table shows the results of the cleaning step:

COLONNES	FILTRES	DENSITÉ	EXPORTER
var5	var6		
5186881	absolut		
5186842	agache sca		
5186595	air liquide		
5186595	air liquide al		
5186608	airbus		
5186621	alcatel alsthom		

# Scénario : HATVP & CAC40

- HATVP : Haute Autorité pour la Transparence de la Vie Publique
- CAC40 : liste des 40 entreprises françaises “les plus influentes”

HATVP → fichier XML de 1,5 millions de lignes

CAC40 → fichier CSV de 40 lignes



**Haute Autorité  
pour la transparence  
de la vie publique**



# Scénario : HATVP & CAC40

Connection Studio x +

localhost:3000/cl\_hatvp\_cac/view

90% ⭐ ⌂ ⌃ ⌄ ⌅ ⌆ ⌇ ⌈ ⌉

## Connection Studio Vue des données

Projet: Hatvp Cac

Chemin 1

declaration.general.declarant.nom#val

Variable de départ: decla

Variable d'arrivée: nom

Chemin 2

declaration.general.mandat.label#val

Variable de départ: decla

Variable d'arrivée: typeMandat

Jointure: Requis

COLONNES	FILTRES	DENSITÉ	EXPORTER
decla	nom	typemandat	
237676	abbassia hakem	elu local ou membre d'un établissement public de coopération intercommunale	
1836220	abdalatif ammar	elu local ou membre d'un établissement public de coopération intercommunale	
3156530	abdallah hassani	député ou sénateur	
2987788	abdel madjid sadi	elu local ou membre d'un établissement public de coopération intercommunale	
558126	abdelaziz hamida	elu local ou membre d'un établissement public de coopération intercommunale	
4894366	abdelghani ghezali	elu local ou membre d'un établissement public de coopération intercommunale	

# Scénario : HATVP & CAC40

The screenshot shows a web-based application window titled "Connection Studio" with a blue header bar. The header includes tabs for "Explorer" and "Vue des données". On the right side of the header, there is a French flag icon and the text "Projet: Hatvp Cac". Below the header, there are several input fields and buttons: "declaration:general:mandat:label#val", "decla", "typeemandat", and two radio buttons labeled "Requis" and "Optionnel".

The main content area is a table with the following data:

typemandat	nbpersonnes
élu local ou membre d'un établissement public de coopération intercommunale	4470
député ou sénateur	911
député européen	77
titulaire d'une fonction soumise uniquement au dépôt d'une déclaration d'intérêts	46
membre du gouvernement	42
membre ou dirigeant d'une autorité administrative indépendante	13
dirigeant du secteur public (sociétés publiques offices publics de l'habitat epic de l'état)	2
membre d'un organe chargé de la déontologie parlementaire	1

At the bottom of the table, there are pagination controls: "Lignes par page : 20", "1-8 sur 8", and navigation arrows.

# Scénario : HATVP & CAC40

Connection Studio    JSON Online Validator and Form

localhost:3000/cl\_hatvp\_cac/view

90% ⭐ ↴ 🔍

## Explorer Connection Studio Vue des données

Projet: Hatvp Cac

Chemin 1  
row.nom\_entreprise.#val.extract:o

Chemin 2  
declaration.participationFinanciereDto.items.item.nomSociete#val.extract:o

Chemin 3  
declaration.participationFinanciereDto.items.item.nombreParts#val

Variable de départ ligneNum

Variable d'arrivée société

Jointure  Requis  Optionnel

Variable de départ declaNum

Variable d'arrivée société

Jointure  Requis  Optionnel

Variable de départ declaNum

Variable d'arrivée nbParts

Jointure  Requis  Optionnel

COLONNES FILTRES DENSITÉ EXPORTER

lignenum	societe	declanum	nbparts ↓
5186647	axa	556656	998
5186725	danone	556656	998
5186751	engie	556656	998
5186933	saint gobain	556656	998
5186946	sanofi	556656	998
5186972	societe generale	556656	998
5197000	vivendi	556656	998

# Conclusion



ConnectionStudio :

- intègre des données très hétérogènes de manière unifiée,
- en extrait des entités nommées (personnes, lieux, ...),
- permet d'inspecter, requêter et connecter ses données de différentes façons

Site web : <https://project.inria.fr/connectionstudio/fr/>

Contact : [ioana.manolescu@inria.fr](mailto:ioana.manolescu@inria.fr)

SCAN  
ME!

