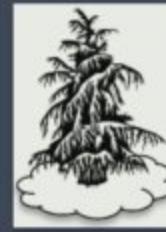




# Teasing journalistic findings out of heterogeneous sources: a data/AI journey



Ioana Manolescu  
Inria Saclay-Île-de-France, Institut Polytechnique de Paris



INSTITUT  
POLYTECHNIQUE  
DE PARIS

# Outline

1. Motivation: why journalism?
2. Integrating heterogeneous data using NLP
3. Fact-checking for RadioFrance



# Motivation: Why journalism?

## Bad memories: Romania, 1989



## Bad memories: Romania, 1989



## Bad memories: Romania, 1989



Ceaușescu re-elected  
at the 14th congress!

He had held power  
since 1965.

## Bad memories: Romania, 1985



## 1990: things got better



... kind of



## Democratic societies crucially need the press

- To debate and express dissent



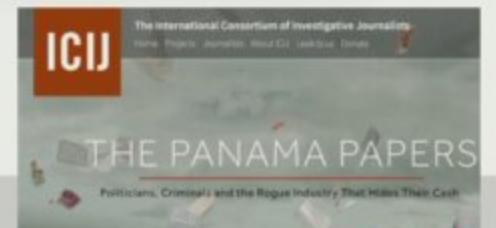
Socialist Romania, 1984

- To analyze, confirm or refute public statements

Fact-checking

(Data) journalism

- To expose and explain society functioning



Inria

# Fact-checking

Not everyone agrees, however, that Democrats are not flip-flopping on the issue.

Mark Krikorian, executive director of the Center for Immigration Studies, a think tank that advocates for lower immigration, said that because the public doesn't know exactly what border barriers the Trump administration wants to build, Mulvaney's statement is not an "exact" comparison. But, he said, to dismiss it simply on that basis would be "terribly literal."

"The fact is that, other than the 'Mexico will pay for it' stuff, Trump is simply channeling the 2006 Secure Fence Act, and Schumer et al. who voted for it out of political calculation are indeed hypocrites for opposing the attempt to finally bring that law to fruition," Krikorian told us via email.

At the surface level, it is true in a broad sense that Democrats including Schumer, Obama and Clinton have in the past supported border fencing. All three voted for the Secure Fence Act of 2006, and all three supported the 2013 Senate immigration overhaul that passed the Senate, which called for tougher border security including some additional fencing. But to claim that those measures are the same as what Trump is proposing is a stretch.

## Share The Facts



**Mick Mulvaney**

Director, Office of Management and Budget

**MISLEADING**

FACTCHECK.ORG

"We don't understand why the Democrats are so wholeheartedly against [President Trump's border wall]. They voted for it in 2006."

Fox News Sunday – Sunday, April 23, 2017

SHARE READ MORE

www.factcheck.org/2017/04/did-democrats-support-border-wall/

FACTCHECK.ORG A Project of The Annenberg Public Policy Center

HOME ARTICLES ASK A QUESTION VIRAL SPOTLIGHT ARCHIVES ABOUT US SEARCH MORE

THE WIRE

## Did Democrats Once Support Border Wall?

By Robert Parley Posted on April 26, 2017

White House Office of Management and Budget Director Mick Mulvaney made an apples-to-oranges comparison when he said he couldn't understand why Democrats opposed supplemental funding for a border wall since many of them were for it back in 2006.

Mulvaney is referring to the Secure Fence Act of 2006, which called for construction of 700 miles of fencing and enhanced surveillance technology, such as unmanned drones, ground-based sensors, satellites, radar coverage and cameras. Sen. Chuck Schumer and then-Sens. Barack Obama and Hillary Clinton were among a bipartisan majority that voted in favor of the legislation, and it was signed into law by President George W. Bush.

In a very general sense, the Democrats named by Mulvaney supported a bill to build more

### ASK FACTCHECK

Ask FactCheck

Ask FactCheck

Ask FactCheck

Q: Did the Supreme Court rule that public schools cannot teach students about Islam?

A: No. That false claim was spread by a network of fake news websites.



# Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)



# Fact-checking

Not everyone agrees, however, that Democrats are not flip-flopping on the issue.

Mark Krikorian, executive director of the Center for Immigration Studies, a think tank that advocates for lower immigration, said that because the public doesn't know exactly what border barriers the Trump administration wants to build, Mulvaney's statement is not an "exact" comparison. But, he said, to dismiss it simply on that basis would be "terribly literal."

"The fact is that, other than the 'Mexico will pay for it' stuff, Trump is simply channeling the 2006 Secure Fence Act, and Schumer et al. who voted for it out of political calculation are indeed hypocrites for opposing the attempt to finally bring that law to fruition," Krikorian told us via email.

At the surface level, it is true in a broad sense that Democrats including Schumer, Obama and Clinton have in the past supported border fencing. All three voted for the Secure Fence Act of 2006, and all three supported the 2013 Senate immigration overhaul that passed the Senate, which called for tougher border security including some additional fencing. But to claim that those measures are the same as what Trump is proposing is a stretch.

**Share The Facts**



**Mick Mulvaney**  
Director, Office of Management and Budget

MISLEADING

*"We don't understand why the Democrats are so wholeheartedly against [President Trump's border wall]. They voted for it in 2006."*

Fox News Sunday – Sunday, April 23, 2017

[SHARE](#)   [READ MORE](#)

[HOME](#)   [ARTICLES](#)   [ASK A QUESTION](#)   [VIRAL SPOTLIGHT](#)   [ARCHIVES](#)   [ABOUT US](#)   [SEARCH](#)   [MORE](#)

[Facebook](#)   [Twitter](#)   [RSS](#)   [DONATE](#)

**THE WIRE**

## Did Democrats Once Support Border Wall?

By Robert Parley | Posted on April 26, 2017

White House Office of Management and Budget Director Mick Mulvaney made an apples-to-oranges comparison when he said he couldn't understand why Democrats opposed supplemental funding for a border wall since many of them were for it back in 2006. Mulvaney is referring to the Secure Fence Act of 2006, which called for construction of 700 miles of fencing and enhanced surveillance technology such as unmanned drones, ground-based sensors, satellites, radar coverage and cameras. Sen. Chuck Schumer and then-Sens. Barack Obama and Hillary Clinton were among a bipartisan majority that voted in favor of the legislation, and it was signed into law by President George W. Bush.

In a very general sense, the Democrats named by Mulvaney supported a bill to build more



TRUE  
POLITFACT  
TRUTH-O-METER™



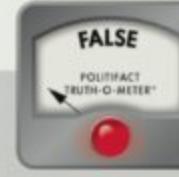
MOSTLY TRUE  
POLITFACT  
TRUTH-O-METER™



HALF TRUE  
POLITFACT  
TRUTH-O-METER™



MOSTLY FALSE  
POLITFACT  
TRUTH-O-METER™



FALSE  
POLITFACT  
TRUTH-O-METER™



PANTS ON FIRE!  
POLITFACT  
TRUTH-O-METER™

Maria

# Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)



# Projects and collaborations

**Google Award** (2015) with U. Paris Sud

**ANR ContentCheck** (2016-2020) with  
Les Décodeurs (Le Monde) <https://contentcheck.inria.fr>

**Inria Associated Team WebClaimExplain** (2017-2019),  
with AIST Japan

Collaboration with H. Galhardas (University of Lisbon),  
A. Anadiotis, O. Balalau, E. Pietriga (Inria, Polytechnique)

**ANR SourcesSay AI Chair** (2020-2023),  
with Le Monde and WeDoData <https://sourcessay.inria.fr>

Fact-checking for **RadioFrance** (2022-2023)



**Le Monde**

**LES DÉCODEURS**  
VENONS-EN AUX FAITS

**Inria** **AIST**

**TÉCNICO**  
LISBOA

**radiofrance**

**Inria**

## Project discussed with Les Décodeurs: fake news detection and propagation on Twitter

**Online fact-checks:** (semi)structured data sources (JSON, XML) listing

- Link to claim (media, social network etc.), **claim author**
- **Fact-check**, containing: analysis (details), final assessment, fc author, date, institution



Among the first published: <https://www.lemonde.fr/webservice/decodex/updates>

Years later: **ClaimReview** by Google and others (<https://www.claimreviewproject.com/>)



## Project discussed with Les Décodeurs: fake news detection and propagation on Twitter

**Online fact-checks:** (semi)structured data sources (JSON, XML) listing

- Link to claim (media, social network etc.), **claim author**
- **Fact-check**, containing: analysis (details), final assessment, fc author, date

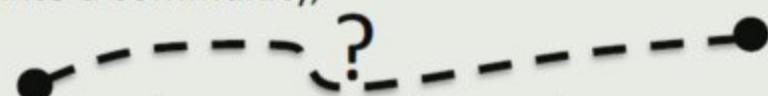


Décodeurs' database of French public figures (Excel)

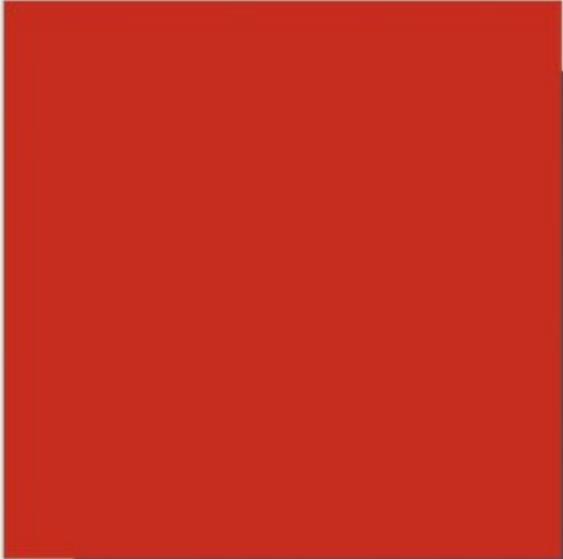
- First name, last name, Twitter ID, position, political party when known



**Question:** When does a fake news post first cross into a community,  
e.g., members of the Parliament?



- Looking for tweets *connected to* a fake news author, *and to a community member*;  
arbitrary paths (chains of author/likes/retweets/inParty/...)



# Integrating heterogeneous journalistic datasets with ConnectionLens

<https://team.inria.fr/cedar/connectionlens/>

## Graph-based integration of heterogeneous data sources

- ❑ The sources are **not RDF**. They can be **(semi)structured**, or **unstructured** (text).
- ❑ The sources may be very **dynamic** (projects started and abandoned as per news cycle and data availability).
- ❑ There is no schema. Data producers often uncollaborative.
- ❑ For most journalists, databases do not come naturally, and IT support is limited. They know keyword-based search...



**Integrate heterogeneous sources within a graph, query w/ keywords**

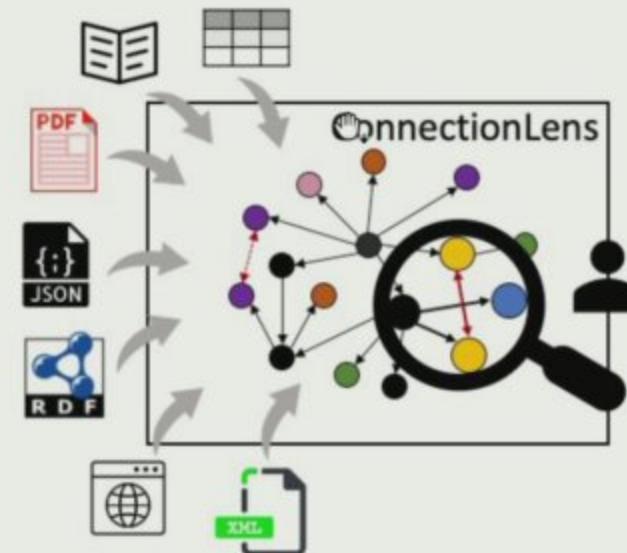
# ConnectionLens: graph-based integration of heterogeneous data sources

<https://team.inria.fr/cedar/connectionlens/>

**Joint work with:** J. Leblay (AIST Japan),  
H. Galhardas and C. Conceição (U. Portugal),  
A. Anadiotis, O. Balalau, N. Barret, T. Bouganim,  
F. Chimienti, M.-Y. Haddad, T. Merabti,  
P. Upadhyay (CEDAR) + interns

S.Horel (Le Monde, European Press Prize  
“Investigative Reporting Award 2018”)

Ongoing work in ANR/DGA AI Chair SourcesSay  
(<https://sourcessay.inria.fr>), DIM RFSI



# ConnectionLens principles [VLDBdemo2018]

Integrate **any kind of data** into a **graph**

**Extract entities** from any text node (regardless of the model of the data source where the text comes from)

- ❑ Same entity in two different text nodes = link among the text nodes (*densification* of the graph)



The graph is **heterogeneous** and **irregular** →

Query it through **keywords**: find trees that connect 1 node matching each kwds

- ❑ Closely related to the Group Steiner Tree Problem (GSTP)

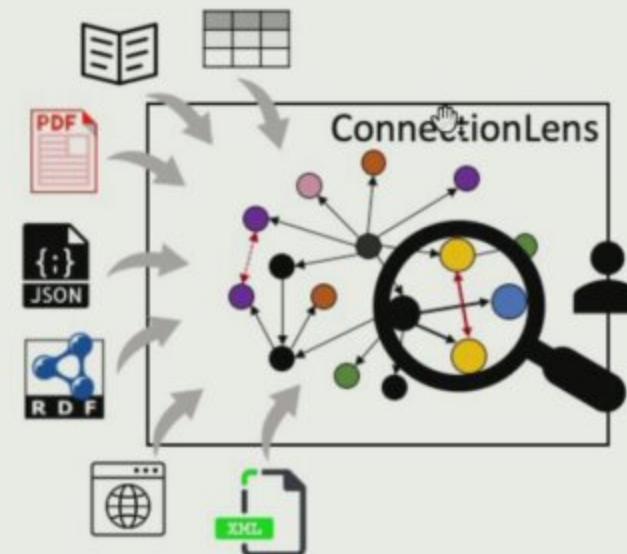
# ConnectionLens: graph-based integration of heterogeneous data sources

<https://team.inria.fr/cedar/connectionlens/>

**Joint work with:** J. Leblay (AIST Japan),  
H. Galhardas and C. Conceição (U. Portugal),  
A. Anadiotis, O. Balalau, N. Barret, T. Bouganim,  
F. Chimienti, M.-Y. Haddad, T. Merabti,  
P. Upadhyay (CEDAR) + interns

S.Horel (Le Monde, European Press Prize  
“Investigative Reporting Award 2018”)

Ongoing work in ANR/DGA AI Chair SourcesSay  
(<https://sourcessay.inria.fr>), DIM RFSI



## Graph-based integration of heterogeneous data sources

- ❑ The sources are **not RDF**. They can be **(semi)structured**, or **unstructured** (text).
- ❑ The sources may be very **dynamic** (projects started and abandoned as per news cycle and data availability).
- ❑ There is no schema. Data producers often uncollaborative.
- ❑ For most journalists, databases do not come naturally, and IT support is limited. They know keyword-based search...



**Integrate heterogeneous sources within a graph, query w/ keywords**

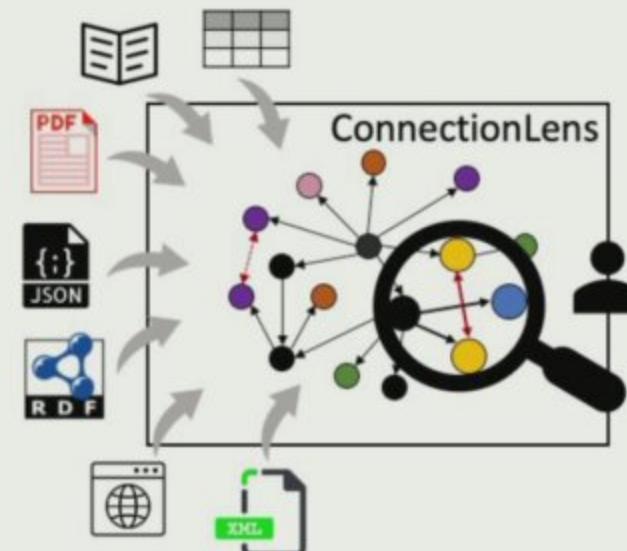
# ConnectionLens: graph-based integration of heterogeneous data sources

<https://team.inria.fr/cedar/connectionlens/>

**Joint work with:** J. Leblay (AIST Japan),  
H. Galhardas and C. Conceição (U. Portugal),  
A. Anadiotis, O. Balalau, N. Barret, T. Bouganim,  
F. Chimienti, M.-Y. Haddad, T. Merabti,  
P. Upadhyay (CEDAR) + interns

S.Horel (Le Monde, European Press Prize  
“Investigative Reporting Award 2018”)

Ongoing work in ANR/DGA AI Chair SourcesSay  
(<https://sourcessay.inria.fr>), DIM RFSI



# ConnectionLens principles [VLDBdemo2018]

Integrate **any kind of data** into a **graph**

**Extract entities** from any text node (regardless of the model of the data source where the text comes from)

- ❑ Same entity in two different text nodes = link among the text nodes (*densification* of the graph)

The graph is **heterogeneous** and **irregular** →

Query it through **keywords**: find trees that connect 1 node matching each kwds

- ❑ Closely related to the Group Steiner Tree Problem (GSTP)

# ConnectionLens principles [VLDBdemo2018]

Integrate any kind of data into a **graph**

Extract entities from any text node (regardless of the model of the data source where the text comes from)

- ❑ Same entity in two different text nodes = link among the text nodes (*densification* of the graph)

The graph is **heterogeneous** and **irregular** →

Query it through **keywords**: find trees that connect 1 node matching each kwds

- ❑ Closely related to the Group Steiner Tree Problem (GSTP)

Rest of the talk based on current state of the project:

[CIKMdemo2021, Elsevier InfSys 2022, IEEE DataEngBull 2021, ongoing works]

# The Balkany and their African connections

The screenshot shows a news article from Libération.fr. The title is "Troisième villa saisie pour les Balkany" (Third villa seized for the Balkany family). The article discusses the couple being tried for tax evasion, specifically mentioning a villa in Marrakech. Below the title is a photograph of Isabelle Balkany. The URL in the browser bar is https://www.liberation.fr/france/2015/06/16/troisieme-villa-saisie-pour-les-balkany/1582215.

**ENQUÈTE**

## Troisième villa saisie pour les Balkany

Par Emmanuelle Farinet — 13 mai 2015 à 14:08

Villas à Marrakech, fonds « occultes »... : les époux Balkany jugés lundi

Soupçonnés d'avoir dissimulé 13 millions d'euros d'avois au fisc, les édiles de Levallois-Perret comparaissent pour fraude fiscale et blanchiment.

Source AFP

Précédent | Suivant

De somptueuses villas à Marrakech et dans les Caraïbes, des fonds « occultes » transitant par le Panama ou Singapour... Soupçonnés d'avoir dissimulé plus de 13 millions d'euros d'avois au fisc, les édiles de Levallois-Perret Patrick et Isabelle Balkany sont jugés à partir de

**Le Point** Politique

Actualité > Politique

**EN CONTINU**

- SI438 L'Assemblée se penche sur un énième débat de la liste grise...
- SI439 Un nouveau traité européen, pourquoi pas ?
- SI440 Fabrice Le Dé - Maisons, le commandant cache-t-il...
- SI441 Les élus votent le plan de relance, pourtant sur la crise lancière
- SI442 Présidentielle : Marine Le Pen prend pour un paravent de l'Ue...

[Voir toute l'actualité en continu](#)

**x qonto**

Ouvrez un compte pro en ligne, contrôlez vos finances simplement.

[Découvrez les offres](#)

## The Balkany and their African connections

The screenshot shows a web browser with two open tabs. The left tab displays a Le Monde article titled "Troisième villa saisie pour les Balkany" (Third villa seized for the Balkany family), dated 13 mai 2010. The right tab displays a Le Point Politique article titled "Villas à Marrakech, fonds « occultes »... : les époux Balkany jugés lundi" (Villas in Marrakech, « hidden funds »... : the Balkany couple tried on Monday). Both articles mention the couple's alleged tax evasion and the seizure of their properties.

Sample integration scenario

## The Balkany and their African connections

Troisième villa saisie pour les Balkany

Villas à Marrakech, fonds « occultes »... : les époux Balkany jugés lundi

3 RÉSULTATS CORRESPONDANT À VOTRE RECHERCHE

RESPONSABLES PUBLICS

REPRÉSENTANTS D'INTÉRÊTS

ACTIVITÉS DE REPRÉSENTATION D'INTÉRÊTS

Perte de l'autorité morale : demande d'Anticor au Président de la République pour que soient révoqués un maire et son adjointe qui ont avoué avoir fraudé l'administration fiscale

ANTICOR

Inria

## The Balkanys and their African connections

Public officials transparency high authority (CSV)

Name	Owner	Location	Type
Dar Gyucy	P. Balkany	Marrakech	Real Estate
Moulin Cossy	I. Balkany	Giverny	Real Estate

## The Balkanys and their African connections

Public officials transparency high authority (CSV)

Name	Owner	Location	Type
Dar Gyucy	P. Balkany	Marrakech	Real Estate
Moulin Cossy	I. Balkany	Giverny	Real Estate

National Directory of Elected Officials (JSON)

```
[ {  
    name: "Levallois-Perret",  
    mayor: "P. Balkany",  
    city-council: [  
        {name: "I. Balkany"},  
        ...  
    ]  
}, ...]
```

## The Balkanys and their African connections

Public officials transparency high authority (CSV)

Name	Owner	Location	Type
Dar Gyucy	P. Balkany	Marrakech	Real Estate
Moulin Cossy	I. Balkany	Giverny	Real Estate

dbpedia.org (RDF)

```
{
  dbr:Marrakech
    dbr:name      "Marrakech"
    rdf:type     dbo:City ;
    dbo:country  dbr:Morrocco .
  dbr:Morocco
    dbr:name      "Morocco"
    rdf:type     dbo:Country
    dbo:locatedIn dbr:Africa .
  dbr:CentralAfricanRepublic
    dbr:name      "Central African Republic"
    dbo:locatedIn dbr:Africa .
}
```

National Directory of Elected Officials (JSON)

```
[ {
  name: "Levallois-Perret",
  mayor: "P. Balkany",
  city-council: [
    {name: "I. Balkany"}, ...
  ]
}, ...]
```

## The Balkanys and their African connections

Public officials transparency high authority (CSV)

Name	Owner	Location	Type
Dar Gyucy	P. Balkany	Marrakech	Real Estate
Moulin Cossy	I. Balkany	Giverny	Real Estate

dbpedia.org (RDF)

```
{
  dbr:Marrakech
    dbr:name      "Marakech"
    rdf:type      dbo:City ;
    dbo:country   dbr:Morrocco .
  dbr:Morocco
    dbr:name      "Morocco"
    rdf:type      dbo:Country
    dbo:locatedIn dbr:Africa .
  dbr:CentralAfricanRepublic
    dbr:name      "Central African Republic"
    dbo:locatedIn dbr:Africa .
}
```

National Directory of Elected Officials (JSON)

```
[ {
  name: "Levallois-Perret",
  mayor: "P. Balkany",
  city-council: [
    {name: "I. Balkany"}, ...
  ],
  ...
}, ...]
```

Libération – Nov. 13, 2014 (Text)

### Balkany mineur de fonds

L'élu de Levallois-Perret est soupçonné d'avoir touché 5 millions de dollars de commission en 2009 grâce à son rôle d'intermédiaire entre Areva et la Centrafrique dans le dossier Uramin. [...]

## How is Levallois-Perret connected to Africa and "real estate"?

Public officials transparency high authority (CSV)

Name	Owner	Location	Type
Dar Gyucy	P. Balkany	Marrakech	Real Estate
Moulin Cossy	I. Balkany	Giverny	Real Estate

dbpedia.org (RDF)

```
{
  dbr:Marrakech
    dbr:name      "Marrakech"
    rdf:type     dbo:City ;
    dbo:country  dbr:Morrocco .
  dbr:Morocco
    dbr:name      "Morocco"
    rdf:type     dbo:Country
    dbo:locatedIn dbr:Africa .
  dbr:CentralAfricanRepublic
    dbr:name      "Central African Republic"
    dbo:locatedIn dbr:Africa .
}
```

National Directory of Elected Officials (JSON)

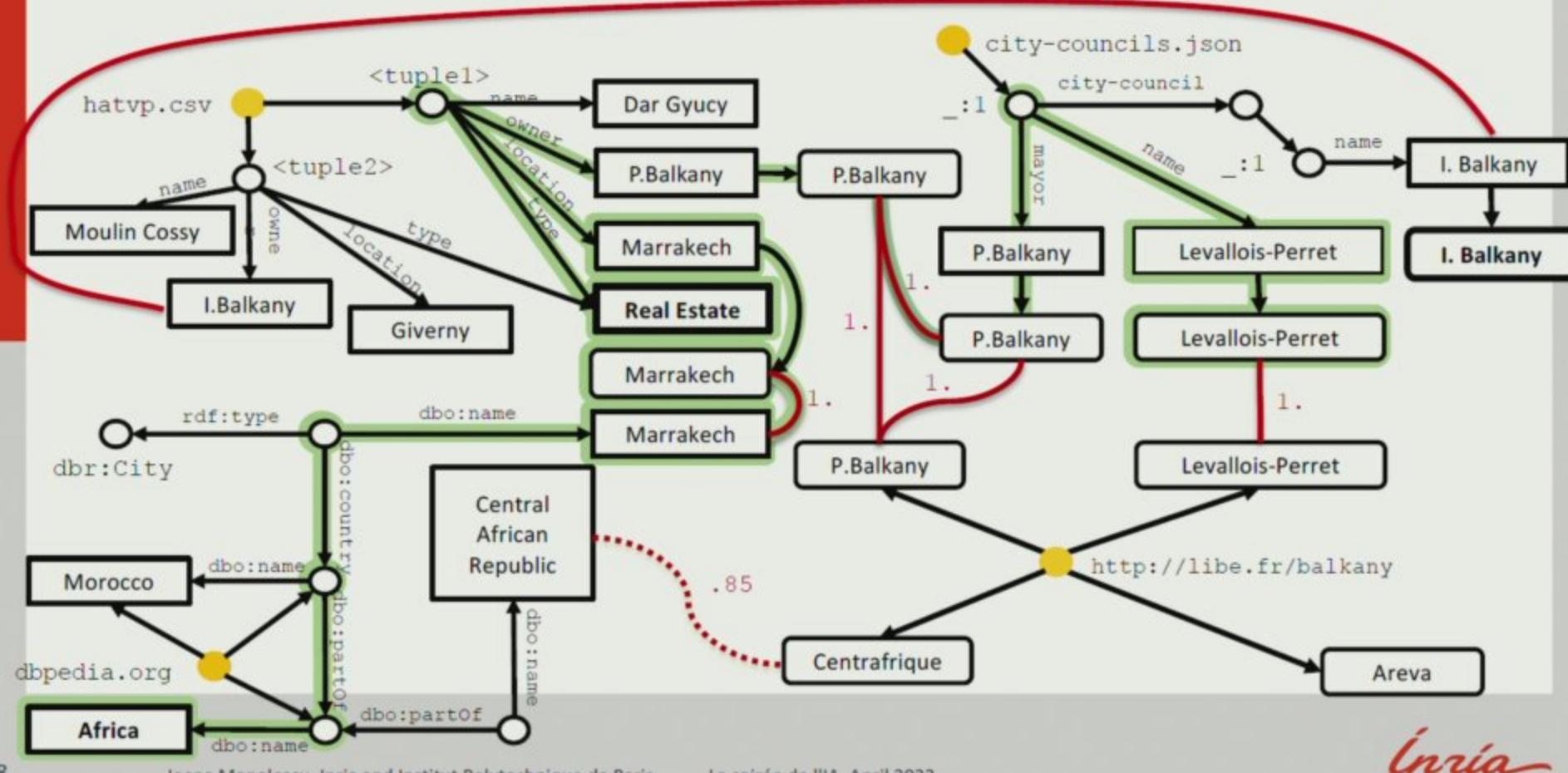
```
[ {
  name: "Levallois-Perret",
  mayor: "P. Balkany",
  city-council: [
    {name: "I. Balkany"}, ...
  ],
  ...
}, ...]
```

Libération – Nov. 13, 2014 (Text)

### Balkany mineur de fonds

L'élu de Levallois-Perret est soupçonné d'avoir touché 5 millions de dollars de commission en 2009 grâce à son rôle d'intermédiaire entre Areva et la Centrafrique dans le dossier Uramin. [...]

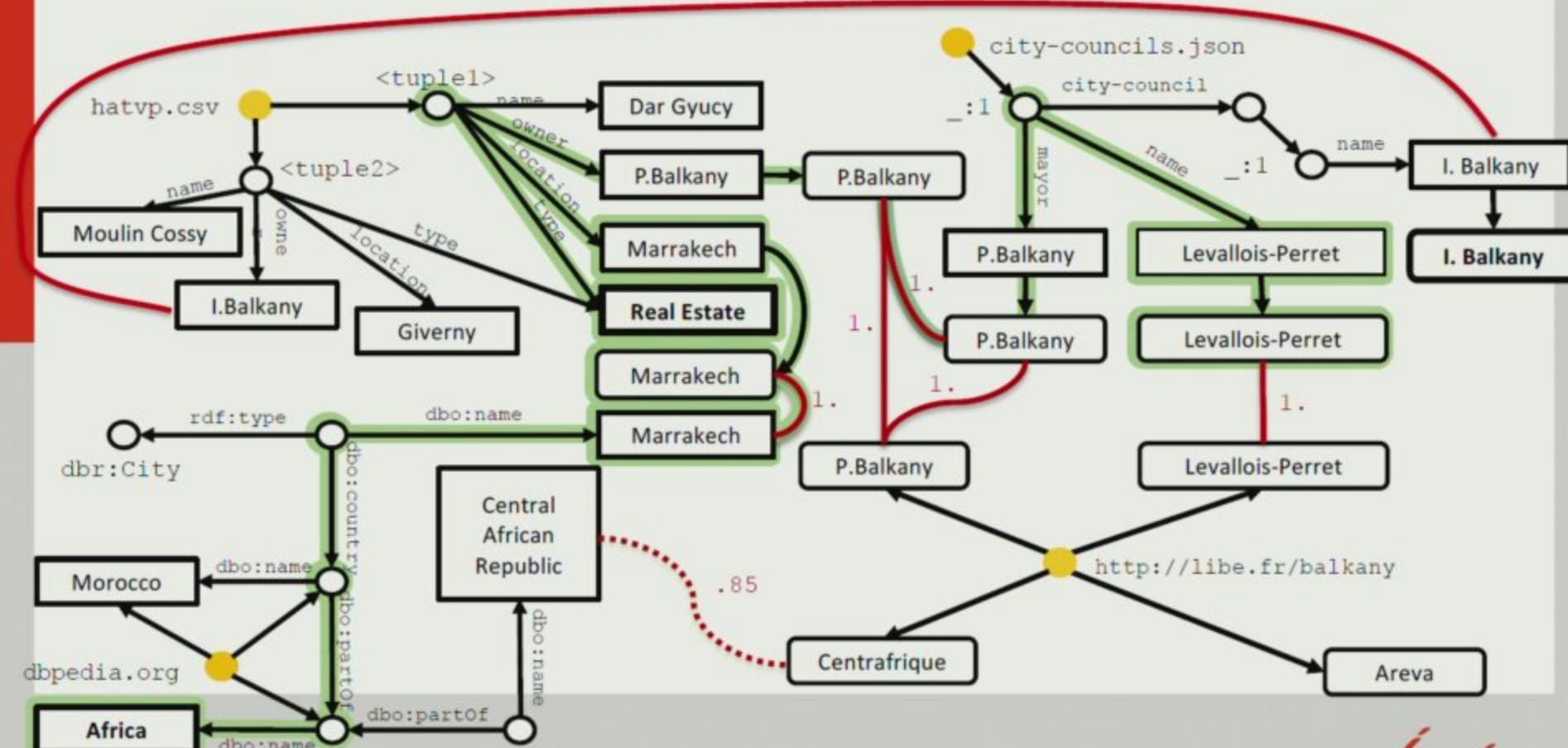
## Idea: integrate **all** data sources into a **heterogeneous graph**



## Querying problem statement

- ❑ Given the graph  $G = (N, E)$  built out of the datasets D and a query  $Q = \{w_1, \dots, w_m\}$ , return the **k highest-score minimal answer trees**.
- ❑ An answer tree is a set of edges which (i) form a tree (ii) contain at least one node whose label matches each keyword  $w_i$ .
- ❑ We are interested in **minimal answer trees**, that is:
  - ❑ Removing an edge from the tree should make it lack some keyword(s).
  - ❑ If a keyword matches more than one nodes in the answer tree, then all these matching nodes must be equivalent.

## Idea: integrate **all** data sources into a **heterogeneous graph**



## Querying problem statement

- ❑ Given the graph  $G = (N, E)$  built out of the datasets D and a query  $Q = \{w_1, \dots, w_m\}$ , return the **k highest-score minimal answer trees**.
- ❑ An answer tree is a set of edges which (i) form a tree (ii) contain at least one node whose label matches each keyword  $w_i$ .
- ❑ We are interested in **minimal answer trees**, that is:
  - ❑ Removing an edge from the tree should make it lack some keyword(s).
  - ❑ If a keyword matches more than one nodes in the answer tree, then all these matching nodes must be equivalent.

## Search space and complexity

- ❑ Problem related to the **(Group) Steiner Tree Problem**
  - ❑ Given graph  $G$ , and nodes  $n_1, \dots, n_m$ , the Steiner Tree Problem (STP) requires the smallest tree in  $G$  that connects all the nodes. Known NP-hard problem in  $|G|$
  - ❑ Group STP: start with  $m$  groups of nodes
- ❑ **Differences with our problem:**
  - ❑ Each edge can be taken in **both directions**: exponential increase in search space size
  - ❑ We need the  **$k$  smallest-cost trees**, not just one.
  - ❑ Support arbitrary score functions
- ❑ **Large literature (30+)** on **kwd search in data graphs**.
  - ❑ Differ in search space and/or make limitative assumptions on score
- ❑ **Our approach: enumerate solutions until time-out or max number of solutions reached.**
  - ❑ Return best  $k$  solutions found, for given score function

## Application: conflicts of interest in the biomedical domain [IEEE DataEngBull 2021, CIKM2021]

Collaboration with Stéphane Horel (Le Monde)

The screenshot shows a web browser window with the URL <https://cl-linkingcois.inria.fr/>. The page title is "CL-LinkingCOIS". The search bar contains "org:Novartis" and the search button is labeled "Search". Below the search bar, it says "219 results". The results are displayed in a table with two columns: "CoIStatement" and "PubmedLink". The "CoIStatement" column contains text about competing interests for Richard L. Banerji and Mamidipudi Thirumala Krishna, and a conflict of interest statement for Benjamin Waschki, Christian Herr, Christina Magrussen, Christoph Sinner, Claus F. Vogelmeier, Henrik Watz, and Johannes T. Neumann. The "PubmedLink" column contains links to view the pubmed paper for each row.

CoIStatement	PubmedLink
Competing interests : <a href="#">Richard L. Banerji</a> reports grants , personal fees and honoraria for lectures from <a href="#">ThermoFisher</a> , <a href="#">Novartis</a> and <a href="#">ALK Abello</a> outside the submitted work. <a href="#">Mamidipudi Thirumala Krishna</a> received honoraria for lectures from <a href="#">Thermo Fisher</a> and <a href="#">ALK Abello</a> , outside the submitted work.	<a href="#">view the pubmed paper</a>
Conflict of interest : <a href="#">Benjamin Waschki</a> has nothing to disclose. Conflict of interest : <a href="#">Christian Herr</a> has nothing to disclose. Conflict of interest : <a href="#">Christina Magrussen</a> has nothing to disclose. Conflict of interest : <a href="#">Christoph Sinner</a> has nothing to disclose. Conflict of interest : <a href="#">Claus F. Vogelmeier</a> reports grants and personal fees from <a href="#">AstraZeneca</a> , <a href="#">Boehringer Ingelheim</a> , <a href="#">GlaxoSmithKline</a> , <a href="#">Grifols</a> and <a href="#">Novartis</a> , personal fees from <a href="#">CSL Behring</a> , <a href="#">Chiesi</a> , <a href="#">Menarini</a> , <a href="#">Mundipharma</a> , <a href="#">Teva</a> and <a href="#">Cipla</a> , grants from <a href="#">Bayer - Schering</a> , <a href="#">MSD</a> and <a href="#">Pfizer</a> , outside the submitted work. Conflict of interest : <a href="#">Henrik Watz</a> reports personal fees from <a href="#">AstraZeneca</a> , <a href="#">Boehringer Ingelheim</a> , <a href="#">GlaxoSmithKline</a> , <a href="#">BerlinChemie</a> , <a href="#">Chiesi</a> , <a href="#">Novartis</a> and <a href="#">Roche</a> , outside the submitted work. Conflict of interest : <a href="#">Johannes T. Neumann</a> reports personal fees from <a href="#">Abbott Diagnostics</a> and <a href="#">Siemens</a> , outside the submitted work. Conflict of	<a href="#">view the pubmed paper</a>

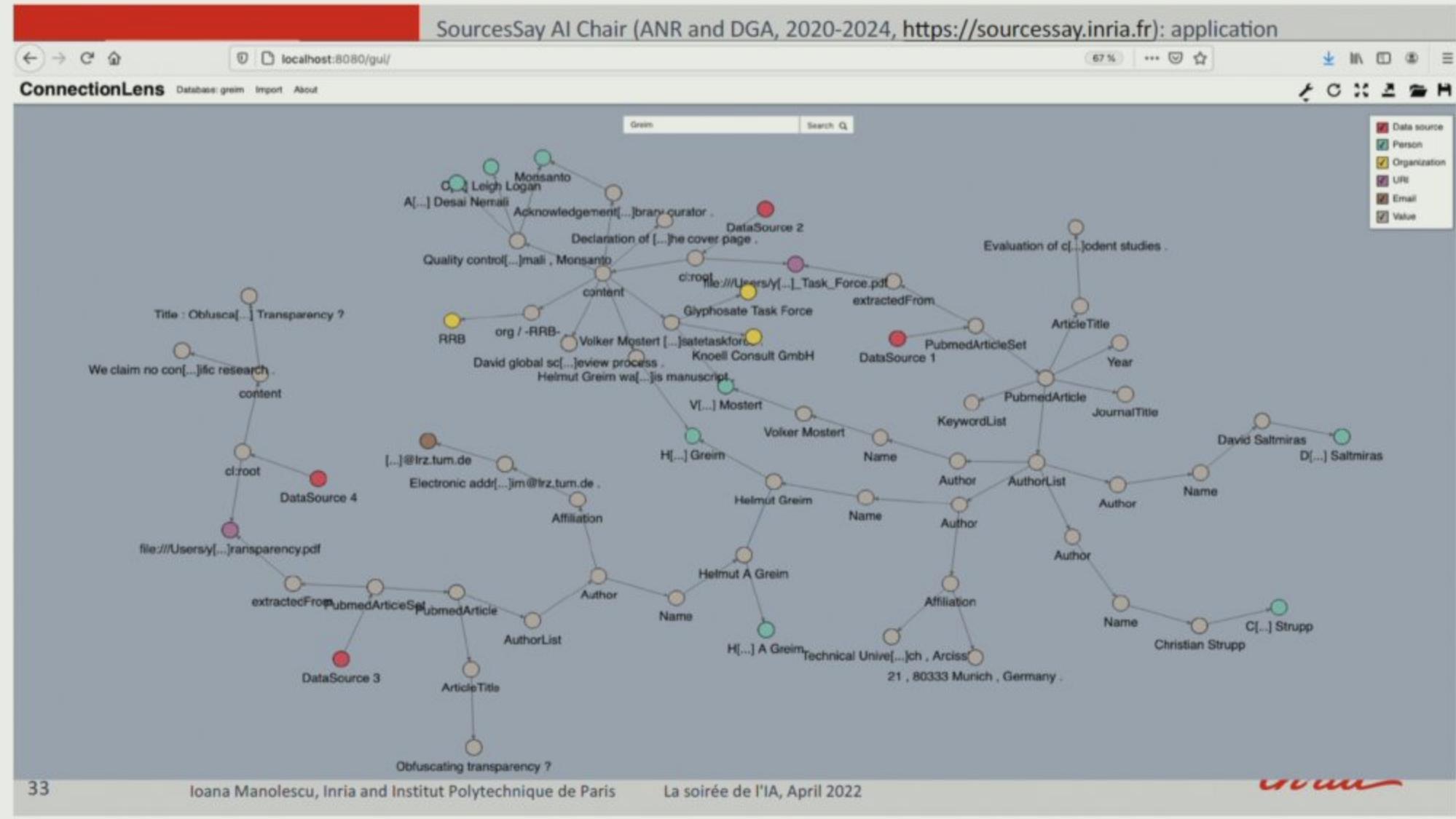
## Application: conflicts of interest in the biomedical domain [IEEE DataEngBull 2021, CIKM2021]

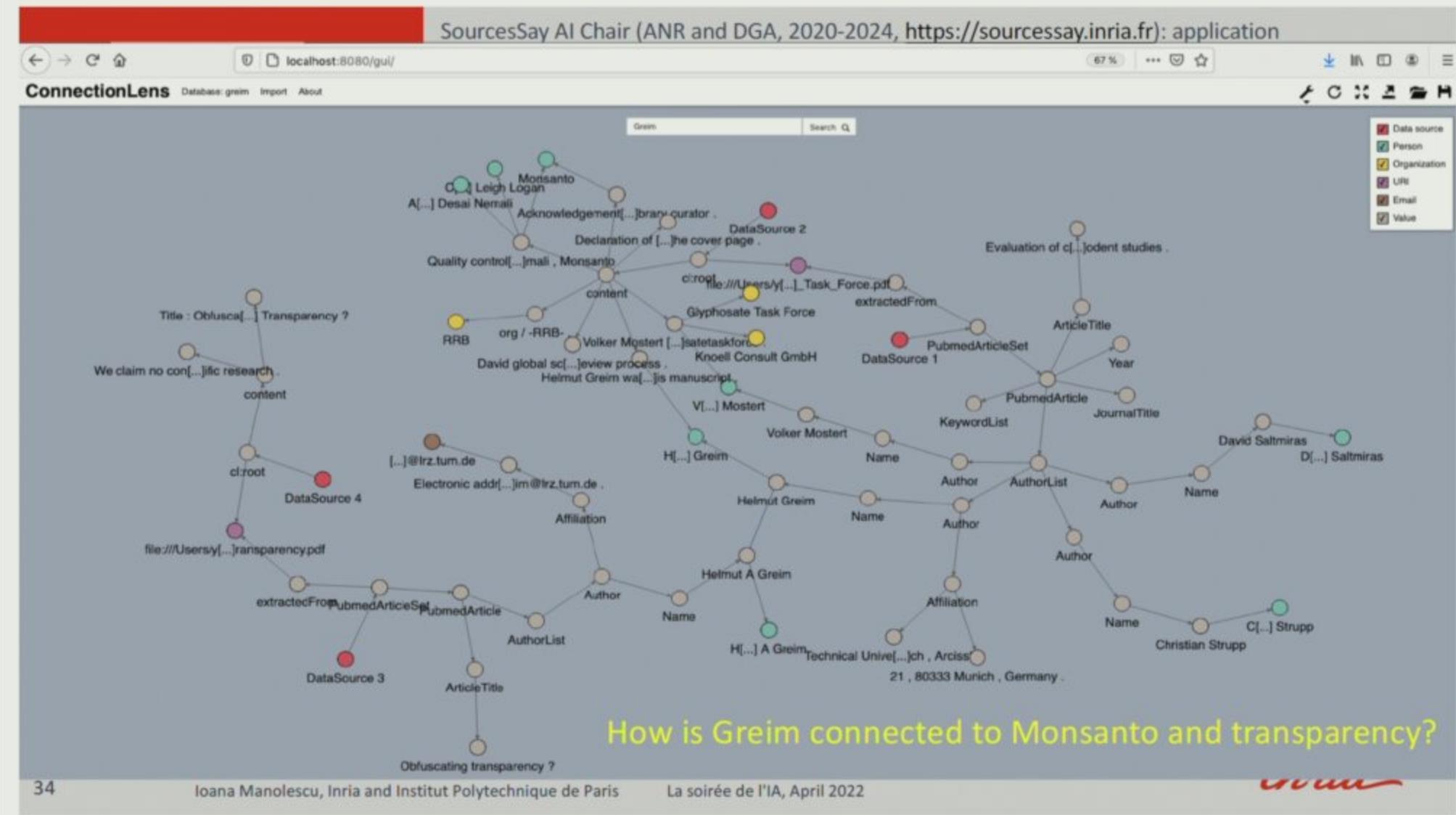
Collaboration with Stéphane Horel (Le Monde)

Data: XML, PDF→JSON, HTML

$ N $	$ E $	$ N $	$ N_P $	$ N_O $	$ N_L $
XML	32,028,429	19,851,904	1,483,631	584,734	126,629
JSON	1,025,307	432,303	75,297	7,320	4,139
HTML	246,636	185,479	3,726	7,227	320
Total	33,300,372	20,469,686	1,562,654	665,167	131,088

**Table 3: Statistics on Conflict of Interest application graph.**





# Sample query answers

The screenshot shows the ConnectionLens application running in a browser window. The title bar says "ConnectionLens" and the address bar shows "localhost:8080/gui". The main interface has a search bar with the placeholder "Search" and a magnifying glass icon. Below the search bar, the text "briand tonolli ruffin" is displayed, followed by a colorful circular icon and a right-pointing arrow.

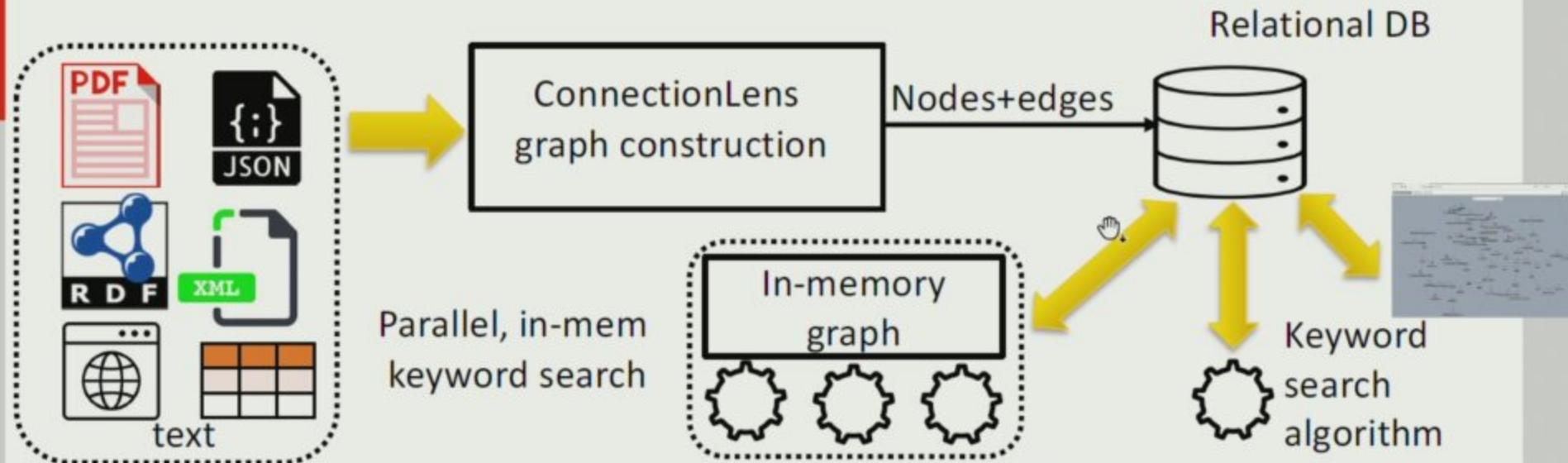
The search results are listed as follows:

- Rank: #1 Size: 10 Score: 0.28 2 sources
- Rank: #2 Size: 9 Score: 0.28 1 source
- Rank: #3 Size: 9 Score: 0.28 1 source
- Rank: #4 Size: 9 Score: 0.28 1 source
- Rank: #5 Size: 10 Score: 0.27 1 source
- Rank: #6 Size: 10 Score: 0.26 1 source
- Rank: #7 Size: 11 Score: 0.25 1 source
- Rank: #8 Size: 10 Score: 0.23 1 source

To the right of the search results is a network graph visualization. The nodes represent entities and their connections. Nodes include "Julie[...] Briand", "Mme Julie Briand", "collaborateurs", "Francois[ruffin] ruffin", "M. Angelo Tonolli", and "Angelo[...] Tonolli". Edges represent relationships like "collaborateurs". A red Inria logo is visible in the bottom right corner.

# Implementation [InfSyst2022]

- ❑ Java (230 classes/46K LOC), Python (24 classes/2800 LOC), JS + CSS
- ❑ Available online: <https://gitlab.inria.fr/cedar/connectionlens>



# Fact-checking with RadioFrance: Heterogeneous data integration: Leveraging Information Extraction, NLP

# Fact-checking and enriching debates

Coordinated ANR ContentCheck (2016-2020): **Le Monde**

- ❑ Fact-checking pipeline to check statistical statements against all INSEE publications (→ graph database)
- ❑ Project restarted in 2022, adding EuroStat, FB, etc.

10/01/2022 14:49

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil | Les Echos

Avec le Covid-19, la campagne électorale se jouera dans les médias, plus que dans les meetings, et nous sommes prêts. En 2021, le paysage de l'information a continué à être bouleversé : l'économie de l'attention et de l'émotion domine avec, pour effets, le buzz, la polarisation et une viralité exponentielle des infos et des infox.

Pour nous, l'enjeu est d'être dans la nuance, dans le vrai plutôt que dans l'annonce. Et surtout de lutter contre la désinformation. Pour cela, nous engageons un partenariat avec Polytechnique et l'Inria pour développer l'intelligence artificielle dans nos rédactions dès cette campagne électorale : la technologie nous permet de réagir vite en traçant la circulation de l'information en ligne et en croisant des bases de données pour la vérifier.

INTERVIEW

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil



La PDG de Radio France se félicite que l'application Radio France ait dépassé le million d'utilisateurs chaque mois. Sibyle Veil explique aussi pourquoi son groupe veut lancer un marché international du podcast. Elle veut aussi garder la main libre sur la succession de Laurence Bloch à la tête de France Inter.

**radiofrance**



Recherche Analyse Evaluation

Afficher le menu

**26 avril 2022**

● NOUVEAU

- + Adrien Quatennens, le 26 avril 2022 à 09h54, [lien du post](#):

Un **ENT** salarié de 2022 produit **3 fois** plus de richesse qu'il y a 50 ans. On peut donc financer plus de retraites avec moins de travail à condition de mieux partager la richesse produite. Ce que Macron refuse de faire. Or l'espérance de vie en bonne santé stagne depuis 10 ans. BFMTV
- + Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ENT** Pfizer pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs **ENT** enfants (**5%** jusqu'ici) ?! (Données ministère de la Santé)
- + Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ENT** Pfizer pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 **CONT\_201** ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs enfants (**5%** jusqu'ici) ?! (Données ministère de la Santé)

**25 avril 2022**

- + François Asselineau, le 25 avril 2022 à 02h12, [lien du post](#):

**ENT** Marine Le Pen obtient + de **60%** des voix à **ENT** Mayotte où **95%** de la **ENT** population est musulmane, Mais aussi Guadeloupe Martinique Guyane où la majorité des populations est d'origine africaine ou amérindienne. COMMENT LES MÉDIAS VONT-ILS CONTINUER À AFFIRMER QU'ELLE EST RACISTE ?
- + François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

MAYOTTE : LE PEN OBTIENT PRÈS DE 60% DES VOIX DANS LE DÉPARTEMENT LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des **17** **ENT** communes et obtient plus de **70%** des voix dans 5 d'entre elles.
- + François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

**ENT** MAYOTTE : LE PEN OBTIENT PRÈS DE **60%** **ENT** DES VOIX DANS LE **ENT** DÉPARTEMENT **ENT** LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des 17 communes et obtient plus de 70% des voix dans 5 d'entre elles.

## Fact-checking for RadioFrance

The screenshot shows a search interface with a blue header bar containing 'Recherche', 'Analyse', and 'Evaluation' tabs. Below the header is a search bar with the query 'chomage auvergne rhône alpes/2019'. The main area displays two search results:

Rang	Lien	Date de publication	Score	Cellule de donnée
0	<p>Figure 1 - Taux de chômage en 2019 (en %) - Lecture : en 2019, le taux de chômage en Auvergne-Rhône-Alpes est de 7,3 %. C'est 0,4 point de moins qu'en 2018. Le taux de chômage des femmes y est de 7,3 %, en %. <a href="https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1">https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1</a> Consulter le site pour voir le tableau tl_path : /data/insee/insee_00/4901601/tabclesse-figure1.xls</p>	03/07/2020	2089.0000	<b>Auvergne-Rhône-Alpes 7,3</b>
1	<p>Fichier Locatifs Social et Fiscal (FLoSf) - Année 2019 Distribution des revenus déclarés par unité de consommation et composition du revenu déclaré - Ménages ayant pour source principale de revenus déclarés des indemnités de chômage Mise en ligne le 17/01/2022 Géographie au 01/01/2020 <a href="https://insee.insee.fr/fr/statistiques/8339912/indic-struct-distrib-revenu-2019-SUPRA.xls">https://insee.insee.fr/fr/statistiques/8339912/indic-struct-distrib-revenu-2019-SUPRA.xls</a> Consulter la feuille n° 25 du fichier "FLO2019_DEC_RES" tl_path : /data/insee/insee_00/601601/indic-struct-distrib-revenu-2019-SUPRA/multiple_FLO2019_DEC_RES/24.xls</p>	17/01/2022	1163.0000	<b>Auvergne-Rhône-Alpes 79,5</b>



## Fact-checking for RadioFrance

Document title: chomage auvergne rhône alpes/2019

Rang	Lien	Date de publication	Score	Cellule de donnée
0	<p>Figure 1 - Taux de chômage en 2019 (en %) - Lecture : en 2019, le taux de chômage en Auvergne-Rhône-Alpes est de 7,9 %. C'est 0,4 point de moins qu'en 2018. Le taux de chômage des femmes y est de 7,3 %, en %.</p> <p><a href="https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1">https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1</a></p> <p>Consulter le site pour voir la tablette</p> <p>tl_path : /data/insee/insee_00/4901601/tabclesse-figure1.ttl</p>	03/07/2020	2089.0000	<b>Auvergne-Rhône-Alpes : 7,9</b>
1	<p>Fichier Localisé Social et Fiscal (FILocalSF) - Année 2019 Distribution des revenus déclarés par unité de consommation et composition du revenu déclaré - Ménages ayant pour source principale de revenus déclarés des indemnités de chômage Mise en ligne le:</p> <p>17/01/2022 Géographie au: 01/01/2020</p> <p><a href="https://insee.insee.fr/fr/statistiques/4903991/indic-struct-distrib-revenu-2019-SUPRA.xls">https://insee.insee.fr/fr/statistiques/4903991/indic-struct-distrib-revenu-2019-SUPRA.xls</a></p> <p>Consulter la feuille n° 25 du fichier "FIL02019_DREC_RES"</p> <p>tl_path : /data/insee/insee_00/6036937/indic-struct-distrib-revenu-2019-SUPRA/multiple_FIL02019_DREC_RES/24/0.xls</p>	17/01/2022	1163.0000	<b>Auvergne-Rhône-Alpes : 79,5</b>

ceder-el-sociale/invalidtwitter\_extraction

## Recherche Analyse Evaluation

Afficher le menu

**26 avril 2022**

• Nouveau

+ Adrien Quatennens, le 26 avril 2022 à 09h54, [lien du post](#):

Un **ENT** salarié de 2022 produit **3 fois** plus de richesse qu'il y a 50 ans. On peut donc financer plus de retraites avec moins de travail à condition de mieux partager la richesse produite. Ce que Macron refuse de faire. Or l'espérance de vie en bonne santé stagne depuis 10 ans. BFMTV

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ONG Pfizer** pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs **ENT** enfants (**5%** jusqu'ici) ?! (Données ministère de la Santé)

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ONG Pfizer** pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 **CONT\_201** ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs enfants (**5%** jusqu'ici) ?! (Données ministère de la Santé)

**25 avril 2022**

+ François Asselineau, le 25 avril 2022 à 02h12, [lien du post](#):

**HDX** Marine Le Pen obtient + de **60%** des voix à **LOC Mayotte**, où **95%** de la **ENT** population est musulmane, Mais aussi Guadeloupe Martinique Guyane où la majorité des populations est d'origine africaine ou amérindienne. COMMENT LES MÉDIAS VONT-ILS CONTINUER À AFFIRMER QU'ELLE EST RACISTE ?

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

MAYOTTE : LE PEN OBTIENT PRÈS DE 60% DES VOIX DANS LE DÉPARTEMENT LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des **17** **ENT** communes et obtient plus de **70%** des voix dans 5 d'entre elles.

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

**LOC MAYOTTE** : LE PEN OBTIENT PRÈS DE **60%** **HDX** DES VOIX DANS LE **ENT** DÉPARTEMENT **MSDC** LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des 17 communes et obtient plus de 70% des voix dans 5 d'entre elles.

# Fact-checking and enriching debates

Coordinated ANR ContentCheck (2016-2020): **Le Monde**

- ❑ Fact-checking pipeline to check statistical statements against all INSEE publications (→ graph database)
- ❑ Project restarted in 2022, adding EuroStat, FB, etc.

10/01/2022 14:49

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil | Les Echos

Avec le Covid-19, la campagne électorale se jouera dans les médias, plus que dans les meetings, et nous sommes prêts. En 2021, le paysage de l'information a continué à être bouleversé : l'économie de l'attention et de l'émotion domine avec, pour effets, le buzz, la polarisation et une viralité exponentielle des infos et des infox.

Pour nous, l'enjeu est d'être dans la nuance, dans le vrai plutôt que dans l'annonce. Et surtout de lutter contre la désinformation. Pour cela, nous engageons un partenariat avec Polytechnique et l'Inria pour développer l'intelligence artificielle dans nos rédactions dès cette campagne électorale : la technologie nous permet de réagir vite en traçant la circulation de l'information en ligne et en croisant des bases de données pour la vérifier.

INTERVIEW

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil



La PDG de Radio France se félicite que l'application Radio France ait dépassé le million d'utilisateurs chaque mois. Sibyle Veil explique aussi pourquoi son groupe veut lancer un marché international du podcast. Elle veut aussi garder la main libre sur la succession de Laurence Bloch à la tête de France Inter.

**radiofrance**



Recherche Analyse Evaluation

Afficher le menu

26 avril 2022

• Nouveau

+ Adrien Quatennens, le 26 avril 2022 à 09h54, [lien du post](#):

Un **SMART** de 2022 produit **3 fois** plus de richesse qu'il y a 50 ans. On peut donc financer plus de retraites avec moins de travail à condition de mieux partager la richesse produite. Ce que Macron refuse de faire. Or l'espérance de vie en bonne santé stagne depuis 10 ans. BFMTV

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **Pfizer** pédiatriques (soit + de **2** doses par enfant de 5 à 11 ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs **enfants** (**5%** jusqu'ici) ? (Données ministère de la Santé)

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **Pfizer** pédiatriques (soit + de **2** **SMART** doses par enfant de 5 à 11 **CONTINU** ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs enfants (**5%** jusqu'ici) ? (Données ministère de la Santé)

25 avril 2022

+ François Asselineau, le 25 avril 2022 à 02h12, [lien du post](#):

**HDX** Marine Le Pen obtient + de **60%** des voix à **Mayotte**, où **95%** de la **population** est musulmane, Mais aussi Guadeloupe Martinique Guyane où la majorité des populations est d'origine africaine ou amérindienne. COMMENT LES MÉDIAS VONT-ILS CONTINUER À AFFIRMER QU'ELLE EST RACISTE ?

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

MAYOTTE : LE PEN OBTIENT PRÈS DE 60% DES VOIX DANS LE DÉPARTEMENT LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1<sup>re</sup> à Mayotte avec 59,1% des voix. Elle remporte 12 des **17** **ENT** communes et obtient plus de **70%** des voix dans 5 d'entre elles.

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

**LAC** **MAYOTTE** : LE PEN OBTIENT PRÈS DE **60%** **MES** DES VOIX DANS LE **DEPARTEMENT** **MES** LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1<sup>re</sup> à Mayotte avec 59,1% des voix. Elle remporte 12 des 17 communes et obtient plus de 70% des voix dans 5 d'entre elles.

# Fact-checking and enriching debates

Coordinated ANR ContentCheck (2016-2020): **Le Monde**

- Fact-checking pipeline to check statistical statements against all INSEE publications (→ graph database)
- Project restarted in 2022, adding EuroStat, FB, etc.

10/01/2022 14:49

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil | Les Echos

Avec le Covid-19, la campagne électorale se jouera dans les médias, plus que dans les meetings, et nous sommes prêts. En 2021, le paysage de l'information a continué à être bouleversé : l'économie de l'attention et de l'émotion domine avec, pour effets, le buzz, la polarisation et une viralité exponentielle des infos et des infox.

Pour nous, l'enjeu est d'être dans la nuance, dans le vrai plutôt que dans l'annonce. Et surtout de lutter contre la désinformation. Pour cela, nous engageons un partenariat avec Polytechnique et l'Inria pour développer l'intelligence artificielle dans nos rédactions dès cette campagne électorale : la technologie nous permet de réagir vite en traçant la circulation de l'information en ligne et en croisant des bases de données pour la vérifier.

INTERVIEW

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil



La PDG de Radio France se félicite que l'application Radio France ait dépassé le million d'utilisateurs chaque mois. Sibyle Veil explique aussi pourquoi son groupe veut lancer un marché international du podcast. Elle veut aussi garder la main libre sur la succession de Laurence Bloch à la tête de France Inter.

**radiofrance**





# Fact-checking with RadioFrance: Heterogeneous data integration: Leveraging Information Extraction, NLP

# Fact-checking and enriching debates

Coordinated ANR ContentCheck (2016-2020): **Le Monde**

- ❑ Fact-checking pipeline to check statistical statements against all INSEE publications (→ graph database)
- ❑ Project restarted in 2022, adding EuroStat, FB, etc.

10/01/2022 14:49

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil | Les Echos

Avec le Covid-19, la campagne électorale se jouera dans les médias, plus que dans les meetings, et nous sommes prêts. En 2021, le paysage de l'information a continué à être bouleversé : l'économie de l'attention et de l'émotion domine avec, pour effets, le buzz, la polarisation et une viralité exponentielle des infos et des infox.

Pour nous, l'enjeu est d'être dans la nuance, dans le vrai plutôt que dans l'annonce. Et surtout de lutter contre la désinformation. Pour cela, nous engageons un partenariat avec Polytechnique et l'Inria pour développer l'intelligence artificielle dans nos rédactions dès cette campagne électorale : la technologie nous permet de réagir vite en traçant la circulation de l'information en ligne et en croisant des bases de données pour la vérifier.

INTERVIEW

« Radio France a dépassé Apple dans les podcasts », annonce sa présidente, Sibyle Veil



La PDG de Radio France se félicite que l'application Radio France ait dépassé le million d'utilisateurs chaque mois. Sibyle Veil explique aussi pourquoi son groupe veut lancer un marché international du podcast. Elle veut aussi garder la main libre sur la succession de Laurence Bloch à la tête de France Inter.

**radiofrance**



Recherche Analyse Evaluation

Afficher le menu

26 avril 2022

• Nouveau

+ Adrien Quatennens, le 26 avril 2022 à 09h54, [lien du post](#):

Un **ENT** salarié de 2022 produit **3 fois** plus de richesse qu'il y a 50 ans. On peut donc financer plus de retraites avec moins de travail à condition de mieux partager la richesse produite. Ce que Macron refuse de faire. Or l'espérance de vie en bonne santé stagne depuis 10 ans. BFMTV

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ONG** Pfizer pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs **ENT** enfants (**3%** jusqu'ici) ?! (Données ministère de la Santé)

+ Florian Philippot, le 26 avril 2022 à 09h04, [lien du post](#):

Au 24 avril, pourquoi a-t-on fait livrer 12030000 (!) doses de **ONG** Pfizer pédiatriques (soit + de **2** **ENT** doses par enfant de 5 à 11 **CONT\_3M** ans) et pourquoi poursuit-on les livraisons régulières alors que les parents ne vaccinent pas leurs enfants (**3%** jusqu'ici) ?! (Données ministère de la Santé)

25 avril 2022

+ François Asselineau, le 25 avril 2022 à 02h12, [lien du post](#):

**HDX** Marine Le Pen obtient + de **60%** des voix à **LOC** Mayotte où **95%** de la **ENT** population est musulmane, Mais aussi Guadeloupe Martinique Guyane où la majorité des populations est d'origine africaine ou amérindienne. COMMENT LES MÉDIAS VONT-ILS CONTINUER À AFFIRMER QU'ELLE EST RACISTE ?

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

MAYOTTE : LE PEN OBTIENT PRÈS DE 60% DES VOIX DANS LE DÉPARTEMENT LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des **17** **ENT** communes et obtient plus de **70%** des voix dans 5 d'entre elles.

+ François Asselineau, le 25 avril 2022 à 02h02, [lien du post](#):

**LOC** MAYOTTE : LE PEN OBTIENT PRÈS DE **60%** **HDX** DES VOIX DANS LE **ENT** DÉPARTEMENT **MSC** LE PLUS MUSULMAN DE FRANCE Marine Le Pen est arrivée largement 1re à Mayotte avec 59,1% des voix. Elle remporte 12 des 17 communes et obtient plus de 70% des voix dans 5 d'entre elles.

## Fact-checking for RadioFrance

radiofrance.fr/inria-research

Recherche Analyse Evaluation

Recherche chomage auvergne rhône alpes/2019

Rang	Lien	Date de publication	Score	Cellule de donnée
0	<p>Figure 1 - Taux de chômage en 2019 (en %) - Lecture : en 2019, le taux de chômage en Auvergne-Rhône-Alpes est de 7,3 %. C'est 0,4 point de moins qu'en 2018. Le taux de chômage des femmes y est de 7,3 %, en %.</p> <p><a href="https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1">https://insee.insee.fr/fr/statistiques/4901601/tabclesse-figure1</a></p> <p>Consulter le site pour voir le tableau</p> <p>tl_path : /data/insee/insee_06/4901601/tabclesse-figure1.xls</p>	03/07/2020	2089.0000	<b>Auvergne-Rhône-Alpes 7,3</b>
1	<p>Fichier Localisé Social et Fiscal (FIL0201F) - Année 2019 Distribution des revenus déclarés par unité de consommation et composition du revenu déclaré - Ménages ayant pour source principale de revenus déclarés des indemnités de chômage Mise en ligne le:</p> <p>17/01/2022 Géographie au: 01/01/2020</p> <p><a href="https://insee.insee.fr/fr/statistiques/4833902/indic-struct-distrib-revenu-2019-SUPRA.xls">https://insee.insee.fr/fr/statistiques/4833902/indic-struct-distrib-revenu-2019-SUPRA.xls</a></p> <p>Consulter la feuille n° 25 du fichier "FIL02019_DEC_RES"</p> <p>tl_path : /data/insee/insee_06/4901601/indic-struct-distrib-revenu-2019-SUPRA/multiple_FIL02019_DEC_RES/24.xls</p>	17/01/2022	1163.0000	<b>Auvergne-Rhône-Alpes 7,5</b>

41 Ioana Manolescu, Inria and Institut Polytechnique de Paris La soirée de l'IA, April 2022

**Inria**



# Credits

ConnectionLens and SourcesSay

- ❑ A. Anadiotis, O. Balalau, N. Barret, T. Bouganim, C. Chanial, F. Chimienti, C. Conceição, R. Dziri, J. Feitz, H. Galhardas, M.-Y. Haddad, J. Leblay, T. Merabti, M. Mohanty, E. Pietriga, P. Upadhyay, Y. Youssef; S. Laurent, S. Horel (Le Monde)
- ❑ <https://sourcessay.inria.fr>

Statistical fact-checking:



- ❑ O. Balalau, T. Cao, S. Ebel, T. Galizzi, Q. Massonnat, X. Tannier; M. Ferrer (Le Monde), A. Krempf (RadioFrance)



## Fact-checking for RadioFrance

Document title: chomage auvergne rhône alpes/2019

Rang	Lien	Date de publication	Score	Cellule de donnée
0	<p>Figure 1 - Taux de chômage en 2019 (en %) - Lecture : en 2019, le taux de chômage en Auvergne-Rhône-Alpes est de 7,3 %. C'est 0,4 point de moins qu'en 2018. Le taux de chômage des femmes y est de 7,3 %, en %.</p> <p><a href="https://insee.insee.fr/fi/statistiques/4501601/tabclesse-figures">https://insee.insee.fr/fi/statistiques/4501601/tabclesse-figures</a></p> <p>Consulter le site pour voir le tableau</p> <p>tl_path : /data/insee/insee_06/4501601/tabclesse-figure1.xls</p>	03/07/2020	2089.0000	<b>Auvergne-Rhône-Alpes 7,3</b>
1	<p>Fichier Localisé Social et Fiscal (FILocalIS) - Année 2019 Distribution des revenus déclarés par unité de consommation et composition du revenu déclaré - Ménages ayant pour source principale de revenus déclarés des indemnités de chômage Mise en ligne le:</p> <p>17/01/2022 Géographie au: 01/01/2020</p> <p><a href="https://insee.insee.fr/fi/statistiques/4501601/indic-struct-distrib-revenu-2019-SUPRA.xls">https://insee.insee.fr/fi/statistiques/4501601/indic-struct-distrib-revenu-2019-SUPRA.xls</a></p> <p>Consulter la feuille n° 25 du fichier "FIL02019.DEC.RES"</p> <p>tl_path : /data/insee/insee_06/4501601/indic-struct-distrib-revenu-2019-SUPRA/multiple_FIL02019_DEC_RES/24.xls</p>	17/01/2022	1163.0000	<b>Auvergne-Rhône-Alpes 7,5</b>

41 Ioana Manolescu, Inria and Institut Polytechnique de Paris La soirée de l'IA, April 2022

**Inria**



# Credits

ConnectionLens and SourcesSay

- ❑ A. Anadiotis, O. Balalau, N. Barret, T. Bouganim, C. Chanial, F. Chimienti, C. Conceição, R. Dziri, J. Feitz, H. Galhardas, M.-Y. Haddad, J. Leblay, T. Merabti, M. Mohanty, E. Pietriga, P. Upadhyay, Y. Youssef; S. Laurent, S. Horel (Le Monde)
- ❑ <https://sourcessay.inria.fr>

Statistical fact-checking:

- ❑ O. Balalau, T. Cao, S. Ebel, T. Galizzi, Q. Massonnat, X. Tannier; M. Ferrer (Le Monde), A. Krempf (RadioFrance)

**Thank you**

**Questions?**

**ConnectionLens:** <https://team.inria.fr/cedar/connectionlens/>