

Une perspective informatique sur le *fact-checking* journalistique

Ioana Manolescu

Inria Saclay—Île-de-France et Ecole polytechnique



Motivation

- Chercheur en informatique à Inria depuis 2002
 - Bases de données: une **industrie** de plus de 80 MD\$/an, omniprésentes dans toute société moderne
 - **Recherche**: Bases de données distribuées, données du Web

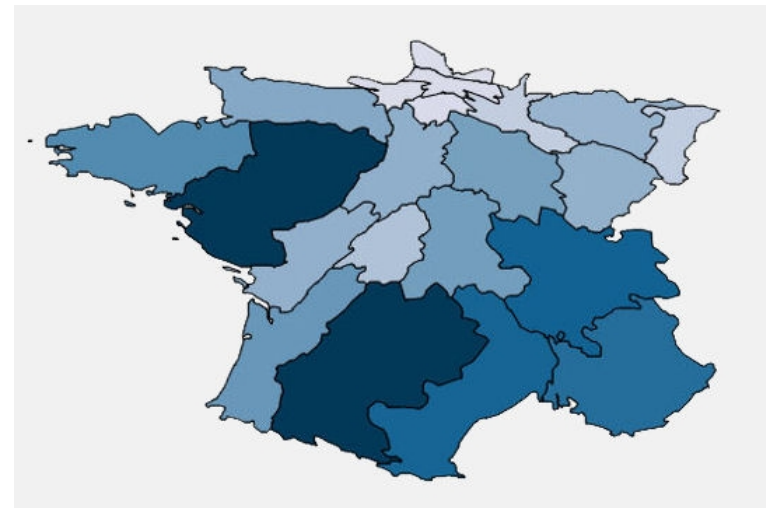
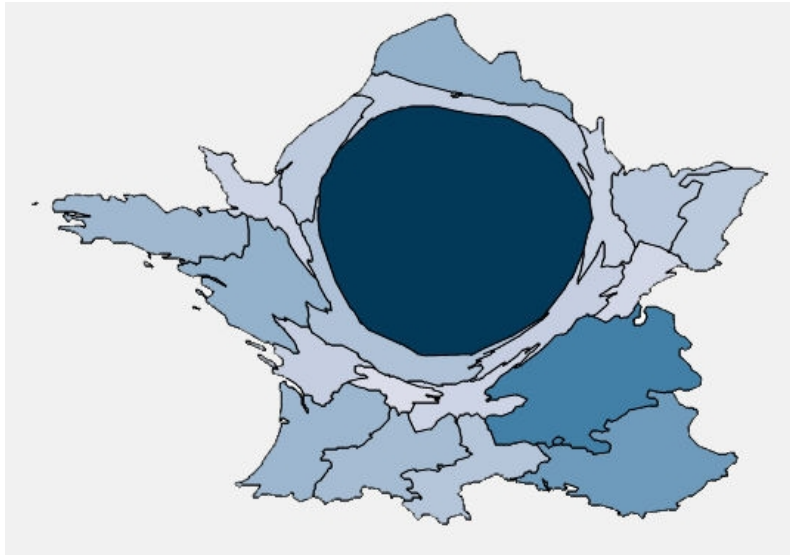


Motivation

- Chercheur en informatique à Inria depuis 2002
 - Bases de données: une **industrie** de plus de 80 MD\$/an, omniprésentes dans toute société moderne
 - **Recherche**: Bases de données distribuées, données du Web
- Données ouvertes (*open data*, p.ex. données statistiques nationales, EU etc.)

Motivation

- Chercheur en informatique à Inria depuis 2002
 - Bases de données distribuées, données du Web
- Données ouvertes (*open data*, p.ex. données statistiques nationales, EU etc.)
 - P. ex. données Etalab, infographies Le Journal du Net, 2010

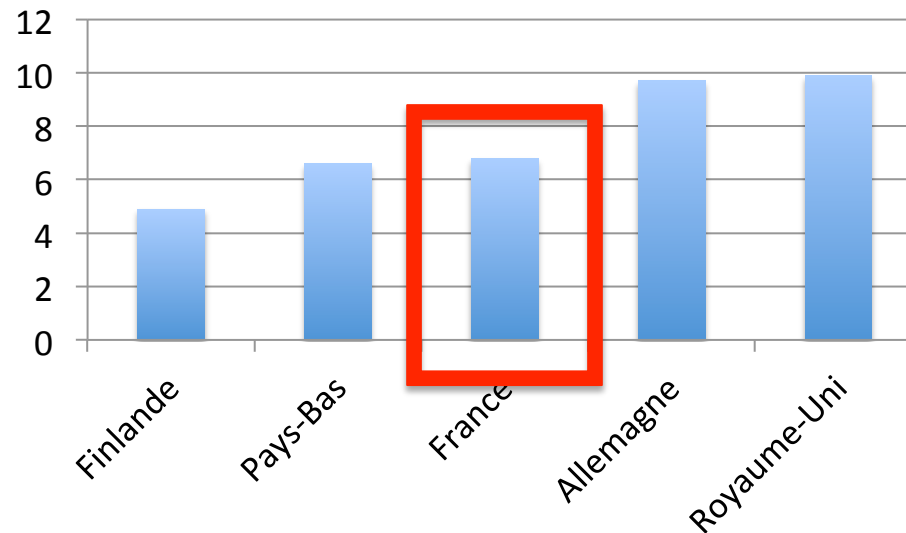


Motivation

- *« On met un pognon de dingue dans les minima sociaux [...] Les gens qui naissent pauvres, ils restent pauvres ; ceux qui tombent pauvres, ils restent pauvres. »*

Motivation

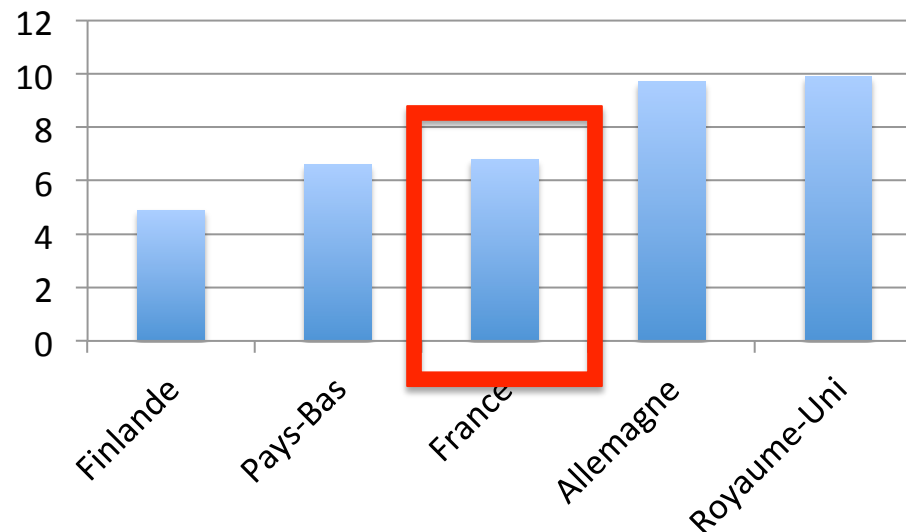
- « *On met un pognon de dingue dans les minima sociaux [...]*
Les gens qui naissent pauvres, ils restent pauvres ; ceux qui tombent pauvres, ils restent pauvres. »
- INSEE: % de la population disposant de la moitié du revenu médian



https://www.lemonde.fr/societe/article/2018/11/05/en-france-vivre-avec-moins-de-800-euros-par-mois_5378868_3224.html

Motivation

- « *On met un pognon de dingue dans les minima sociaux [...]*
Les gens qui naissent pauvres, ils restent pauvres ; ceux qui tombent pauvres, ils restent pauvres. »
- INSEE: % de la population disposant de la moitié du revenu médian



Aussi: % de la population ayant connu la pauvreté sur les 12 derniers mois

https://www.lemonde.fr/societe/article/2018/11/05/en-france-vivre-avec-moins-de-800-euros-par-mois_5378868_3224.html

Projet ANR ContentCheck



"Content Management Techniques for
Fact-Checking: Models, Algorithms, and Tools"
2016-2019



Lyon 1



LES DÉCODEURS

VENONS-EN AUX FAITS

<http://contentcheck.inria.fr>

Projet ANR ContentCheck



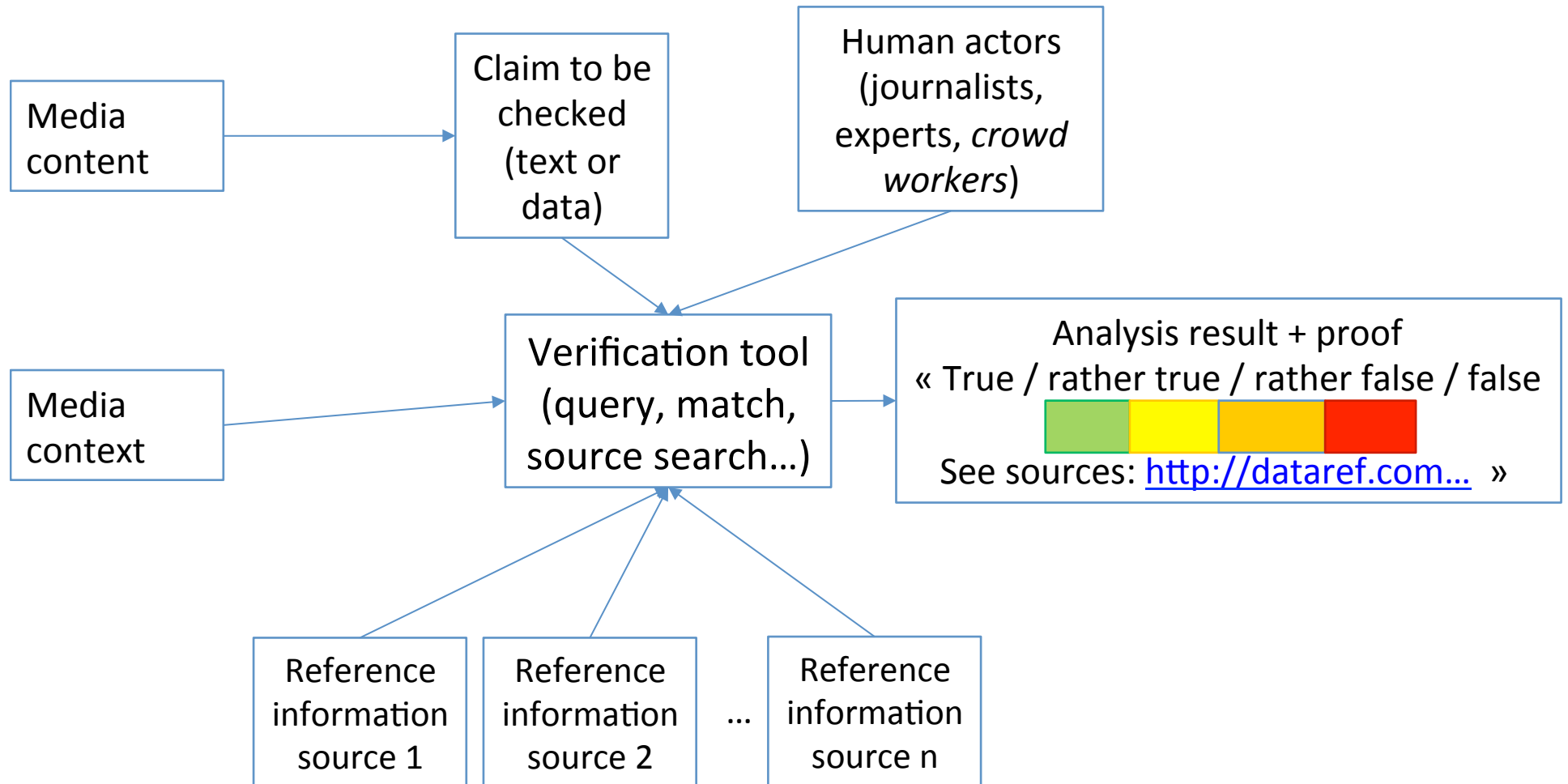
Contexte

- *Fact-checking* journalistique de plus en plus connu/visible dans le paysage médiatique
- Percées de l'automatisation par l'informatique

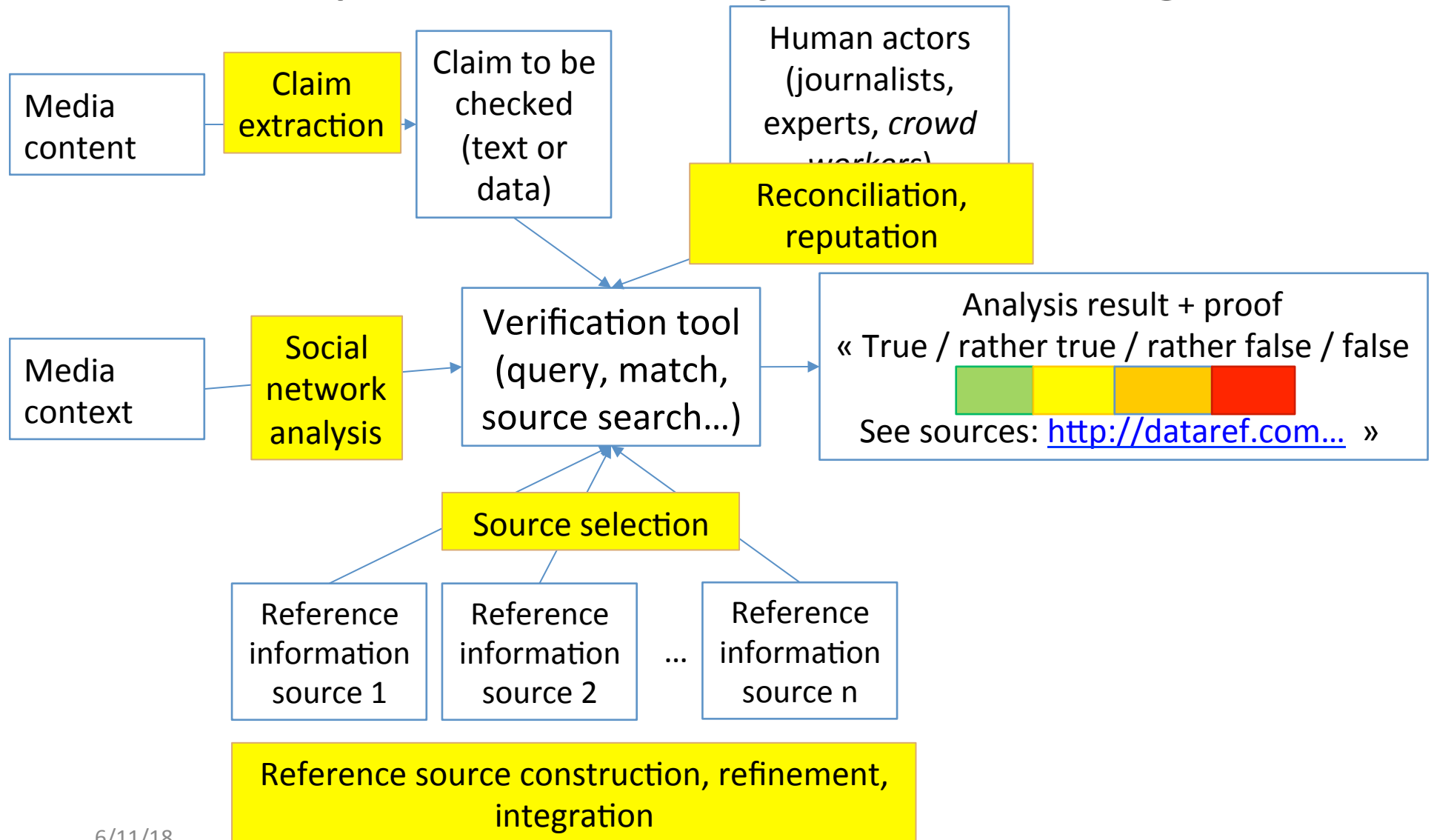
Objectifs

1. **Compréhension scientifique** du processus de *fact-checking* vu comme un problème informatique de gestion de contenus
2. **Nouveaux algorithmes** pour des tâches de *fact-checking*: sémantique; texte et classification; fouille de données
3. **Prototypes** et **validation**
4. **Corpus de référence**

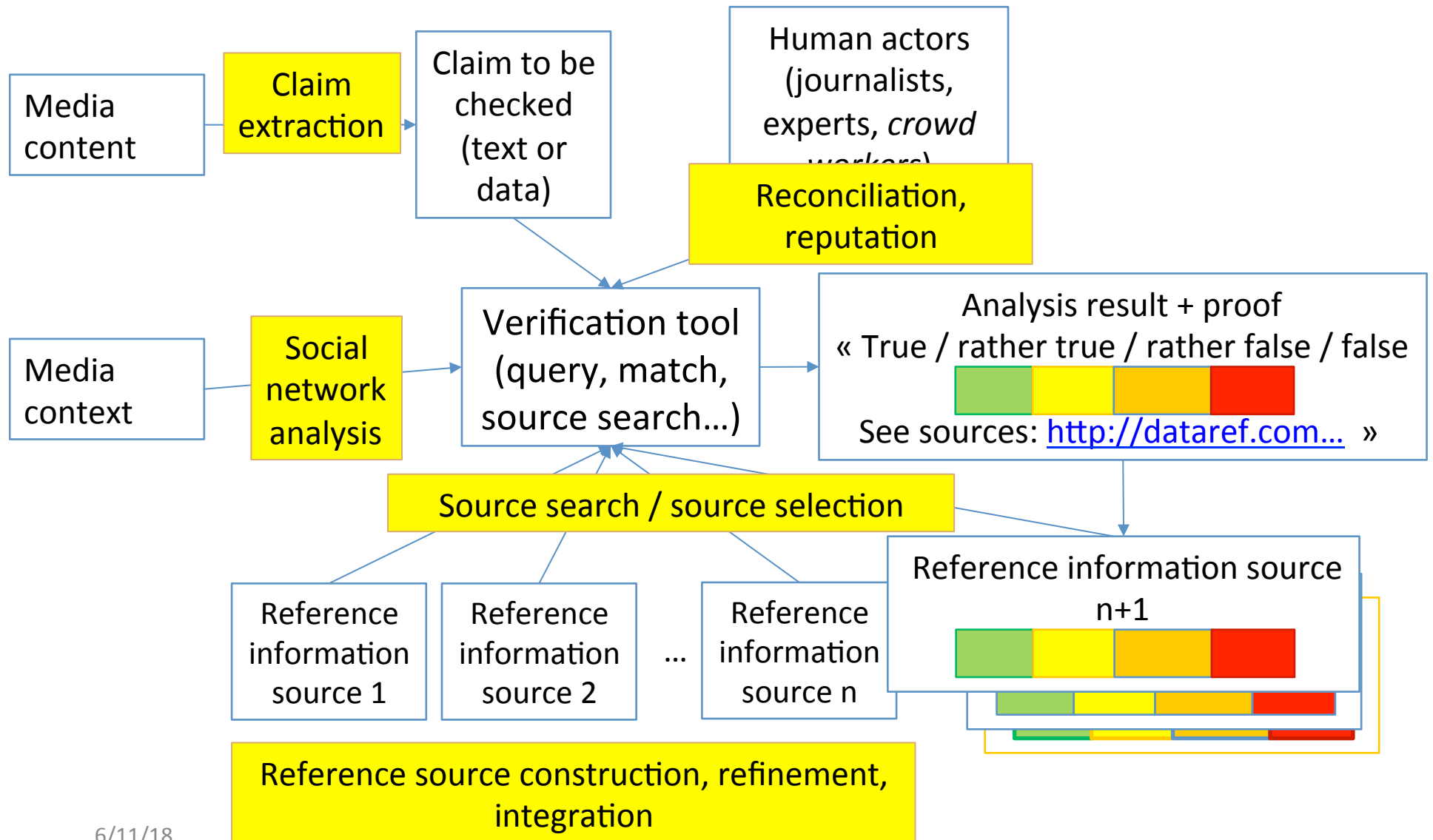
Une perspective informatique sur le processus de *fact-checking*



Une perspective informatique sur le processus de *fact-checking*



Une perspective informatique sur le processus de *fact-checking*



Faciliter l'accès aux sources de référence

T.-D. Cao, I. Manolescu, X. Tannier. "Extracting Linked Data from statistic spreadsheets" [SBD 2017, WebDB 2018]

“Créations d’entreprises en France en 2015” conduit à:

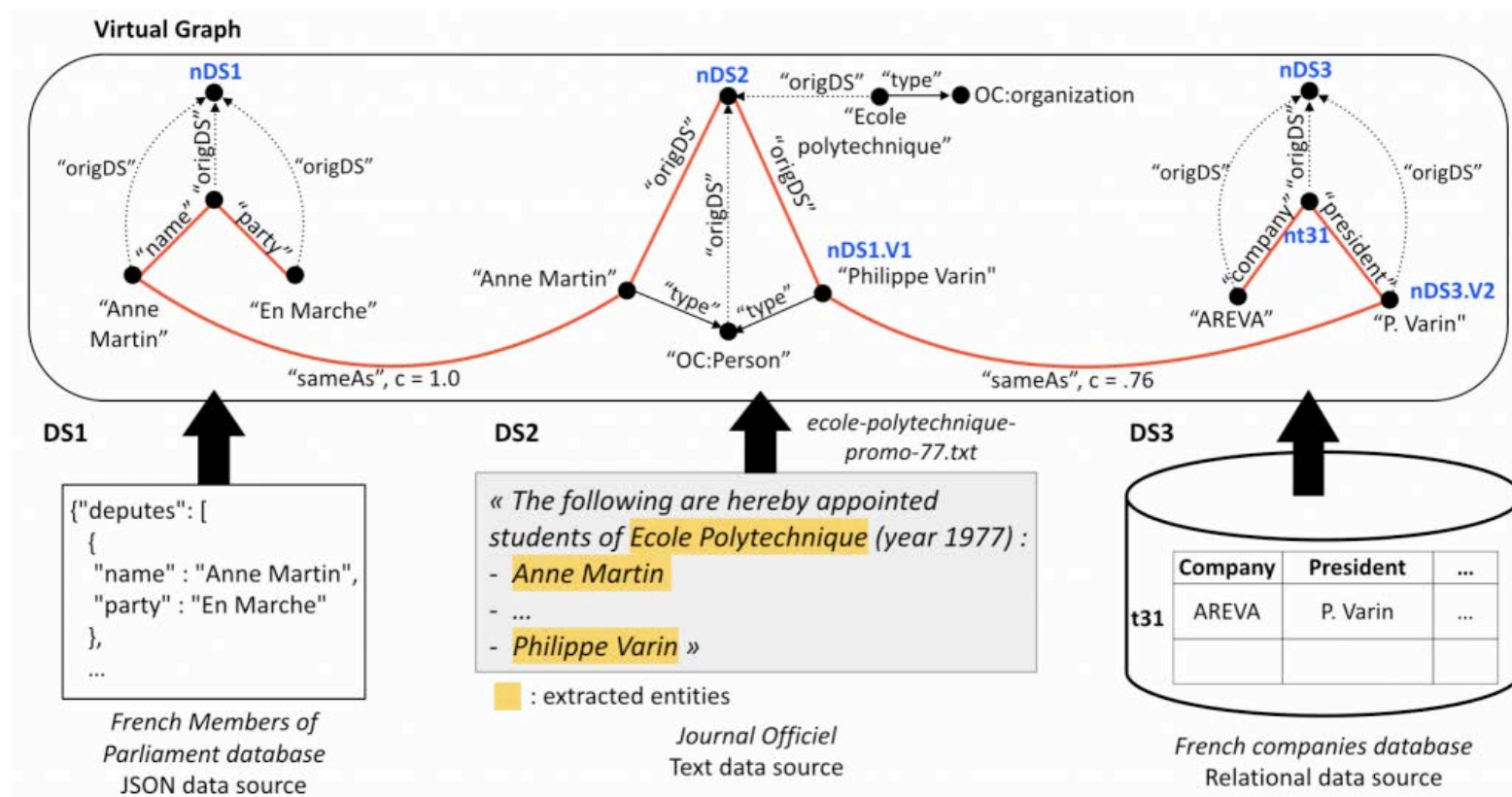
Créations d'entreprises dans quelques pays de l'Union européenne en 2015

en %

Pays	Taux de création
Allemagne	7,1
Belgique	6,2
Espagne	9,5
France (1)	9,5
Italie	7,5
Pays-Bas	10,1
Portugal	15,7
République tchèque	8,2
Royaume-Uni	14,3

Recherche par mots clés dans des sources de données hétérogènes

C. Chanial, R. Dziri, H. Galhardas, J. Leblay, M. Le Nguyen and I. Manolescu: "ConnectionLens: Finding Connections Across Heterogeneous Data Sources" [VLDB2018]



Recherche automatique dans les sources

F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu, S. Zampetakis: "Fact-checking and analyzing the Web" [SIGMOD 2013]

- Plug-in
- Formule des recherches automatiques dans une base de contenus à partir de la page visitée
- « Second screen »

The screenshot shows a web browser window with a news article on the left and a 'Second screen' overlay on the right. The article is titled 'Bill Clinton apporte son soutien à Barack Obama' and discusses Clinton's support for Obama during the 2008 election. The 'Second screen' overlay contains several panels:

- Concepts**: A list of entities extracted from the article, including Bill Clinton, Barack Obama, Des Moines, Iowa, Hillary Clinton, Michelle Obama, Bruce Springsteen, and Pennsylvanie, each with a dbpedia URI.
- Facts & figures**: A panel titled 'Curriculum' showing a portrait of Bill Clinton and his biographical details: Born William Jefferson, August 19, 1946, Hope, Arkansas, U.S., Political party Democratic Party, Spouse(s) Hillary Rodham.
- Quotes**: A panel with two quotes: 'Launching the Africa Regional Media Hub in Johannesburg' and 'President Obama speaking LIVE for the last time before the election', both with links to the original sources.
- Sources**: A panel showing related stories, including 'Ohio, Florida, ... être compliqué' and 'Des indicateurs ... d'Obama', with links to lemonde.fr.

MERCI / QUESTIONS?

Question auditeur: Est-il vrai qu'à Moscou,
on donne des Mercedes aux citoyens?

Radio Yerevan: Oui, mais il s'agit de Ladas plutôt que
de Mercedes,
et on les leur vole plutôt que de les leur donner...