

Content management techniques and tools for fact-checking

Francois Goasdoue^{a,b}, Ioana Manolescu^{b,c,d}, Xavier Tannier^{d,e}

^a U. Rennes 1 ^b Inria Saclay ^c LIX (CNRS and Ecole Polytechnique)

^d U. Paris Saclay ^e LIMSI (CNRS and U. Paris Sud)

PARIS DB DAY

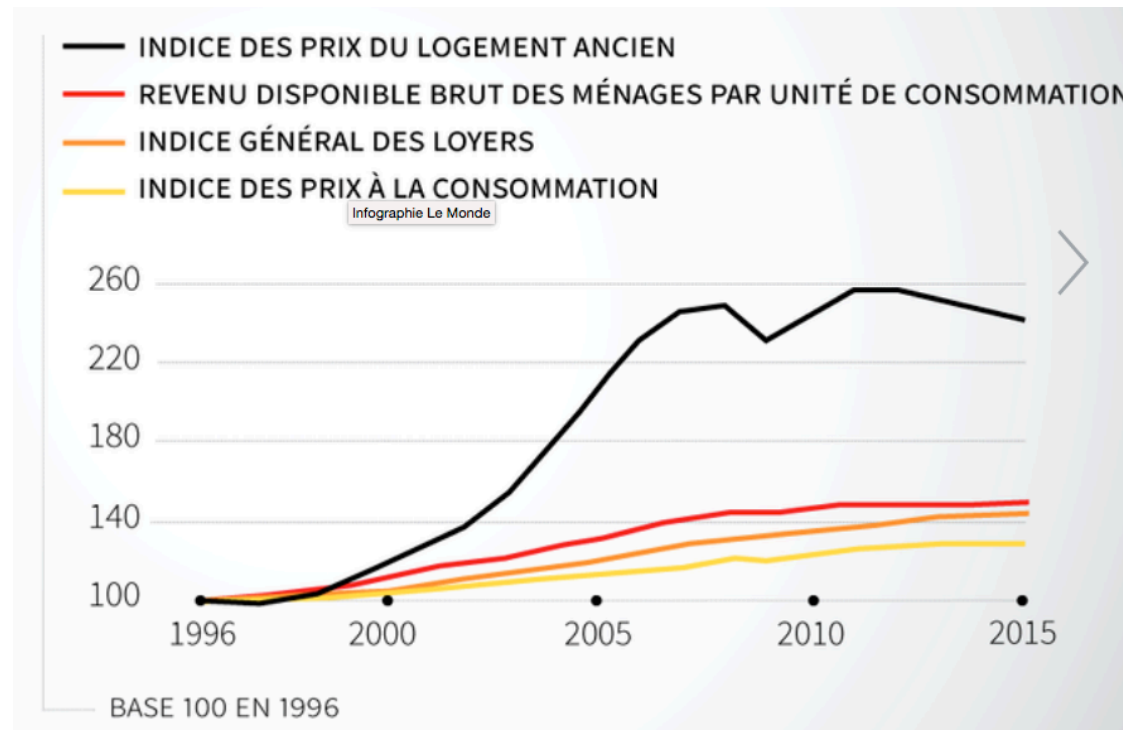
09/05/2017

Plan

1. Data journalism and fact-checking
2. ... are content management problems
3. Some works I have been involved to in this area

Data journalism

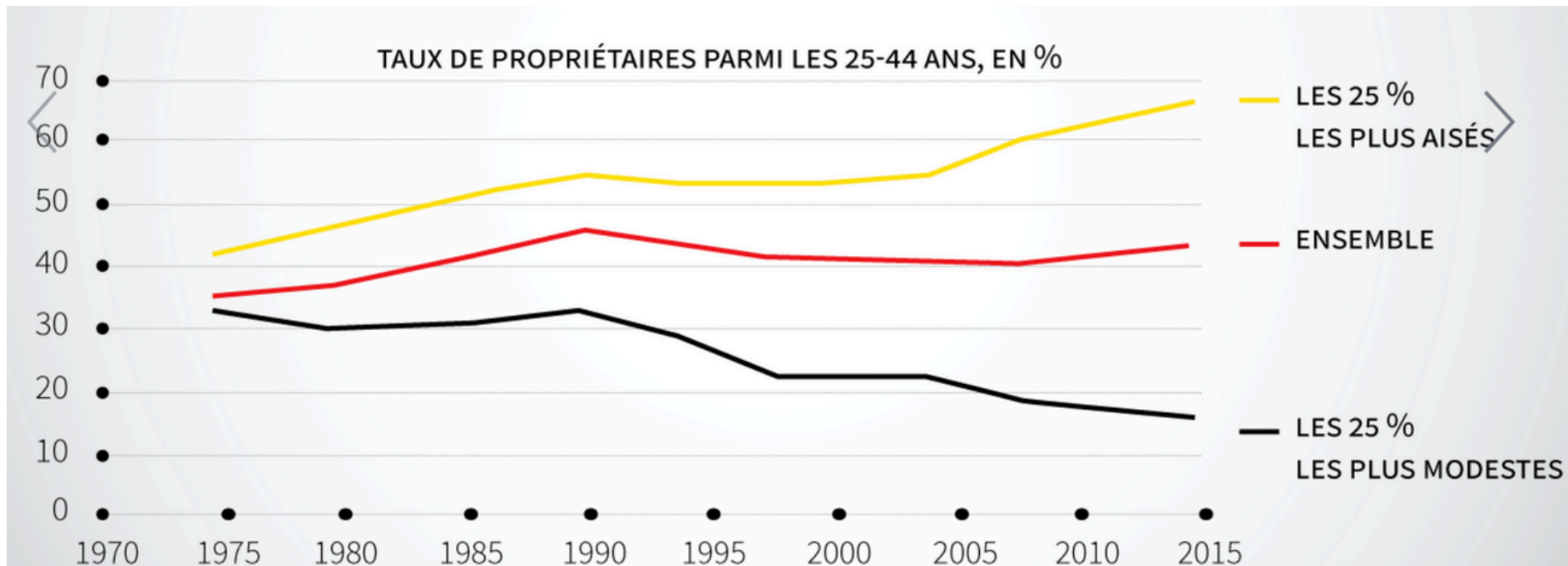
Investigative journalism based on **complex and/or large data**



http://abonnes.lemonde.fr/les-decodeurs/portfolio/2017/04/18/les-fractures-francaises-1-5-le-logement-les-raisons-de-la-crise_5112859_4355770.html

Data journalism

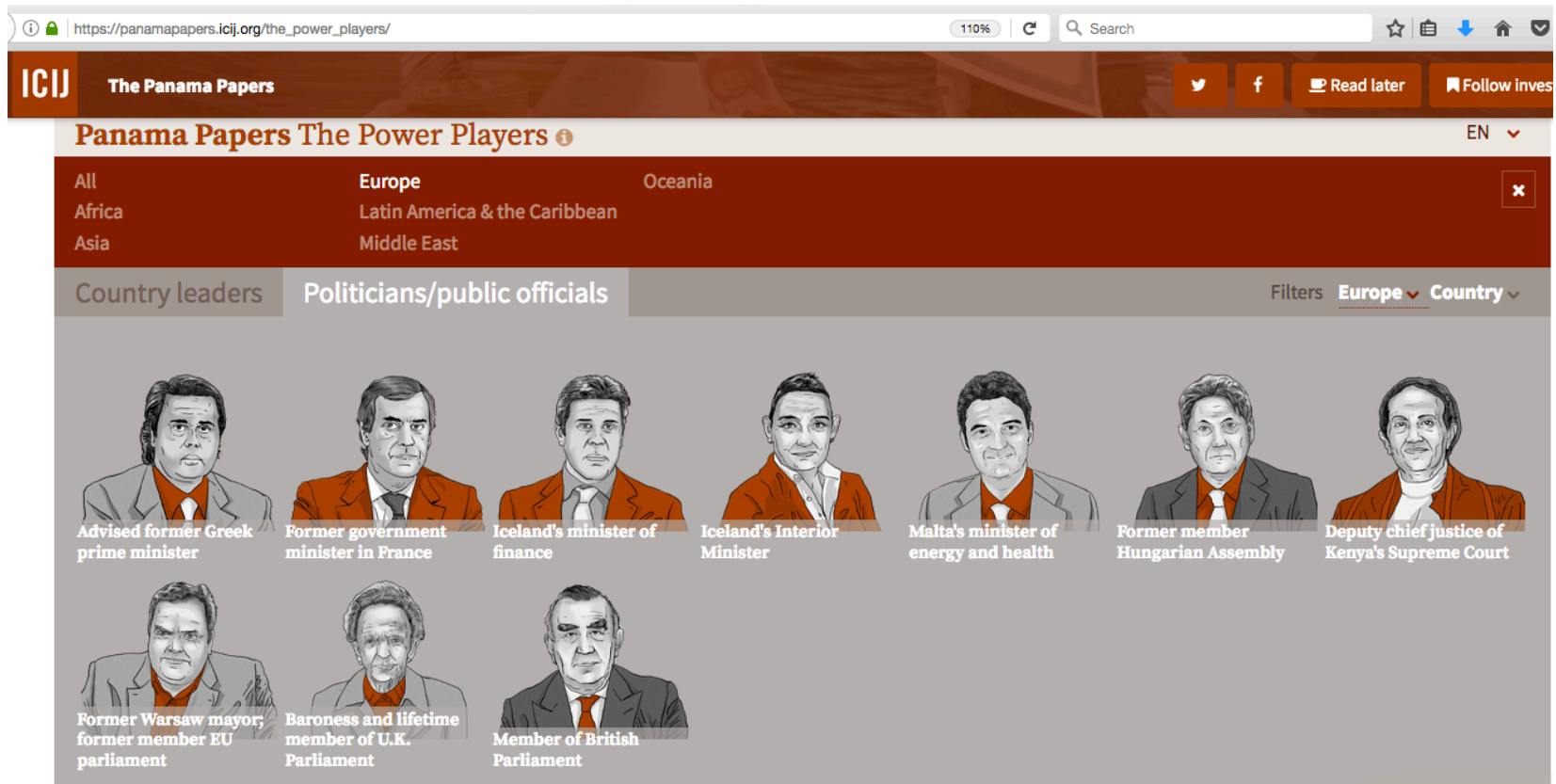
Investigative journalism based on **complex and/or large data**



http://abonnes.lemonde.fr/les-decodeurs/portfolio/2017/04/18/les-fractures-francaises-1-5-le-logement-les-raisons-de-la-crise_5112859_4355770.html

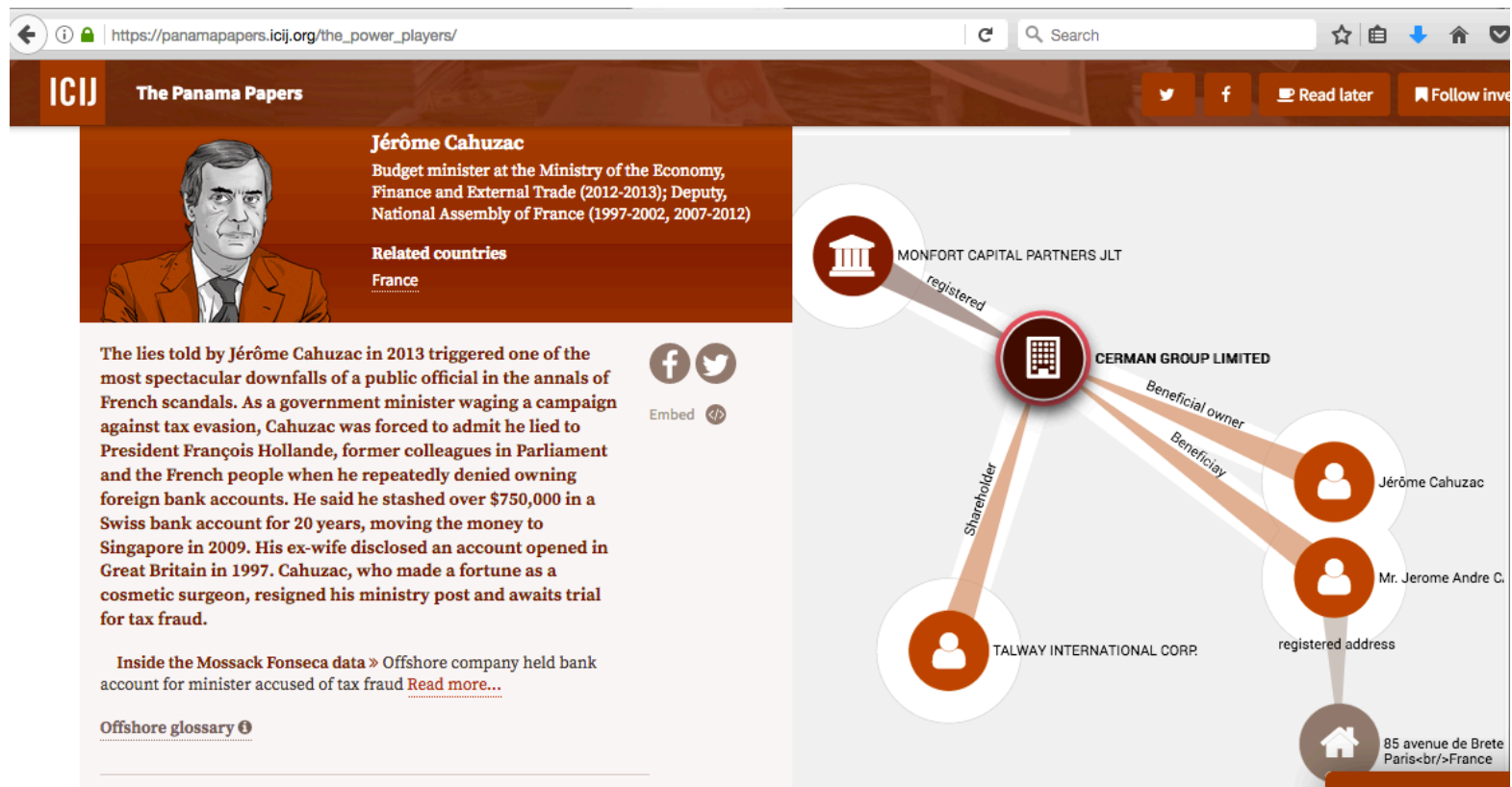
Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)



Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)



Fact-checking (since 1930 approx.)

Fact-checking: verification of facts mentioned in media content

- ❑ To protect media reputation and avoid legal action
- ❑ Verification supposes the existence of a reference dataset

“The day I became a fact-checker at The New Yorker, I received **one set of red pencils** [...] for underlining **passages on page proofs of articles that might contain checkable facts.** [...] confirmed **with the help of reference books** from the magazine’s library, including Merriam-Webster’s Geographical Dictionary, the New Grove Dictionary of Music and Musicians and Burke’s Peerage and Gentry.”



<http://www.nytimes.com/2010/08/22/magazine/22FOB-medium-t.html>

Fact-checking (2012 – ongoing)

 A Project of The Annenberg Public Policy Center

HOMEARTICLESASK A QUESTIONVIRAL SPIRALARCHIVESABOUT USSEARCHMORE

THE WIRE

Did Democrats Once Support Border Wall?

By Robert Farley Posted on April 26, 2017

Like 835

Tweet

Pinterest

Share

11

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

Like 953

Tweet

Pinterest

Share

98

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.

Share The Facts



Mick Mulvaney
Director, Office of Management and Budget

"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

SHAREREAD MORE



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 "tendentiously 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, and 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.

MISLEADING

FACTCHECK.ORG





Washington Post's TruthTeller (2013)

Discontinued

Video → Automated transcript → manual matching into FactCheck.org

Truth Teller beta




More Video 

Auto Transcript ☒ Autoscroll

security investigation I think now more than ever we need
transparency from people who seek to have the trust of
the folks virginia to assume the office of governor and it's
pretty rich to have the guy who ran out the lincoln
bedroom sold seats on air force one was an uninvited
token spirit or any teamsters election law and money
laundering case be talking about aren't six now?

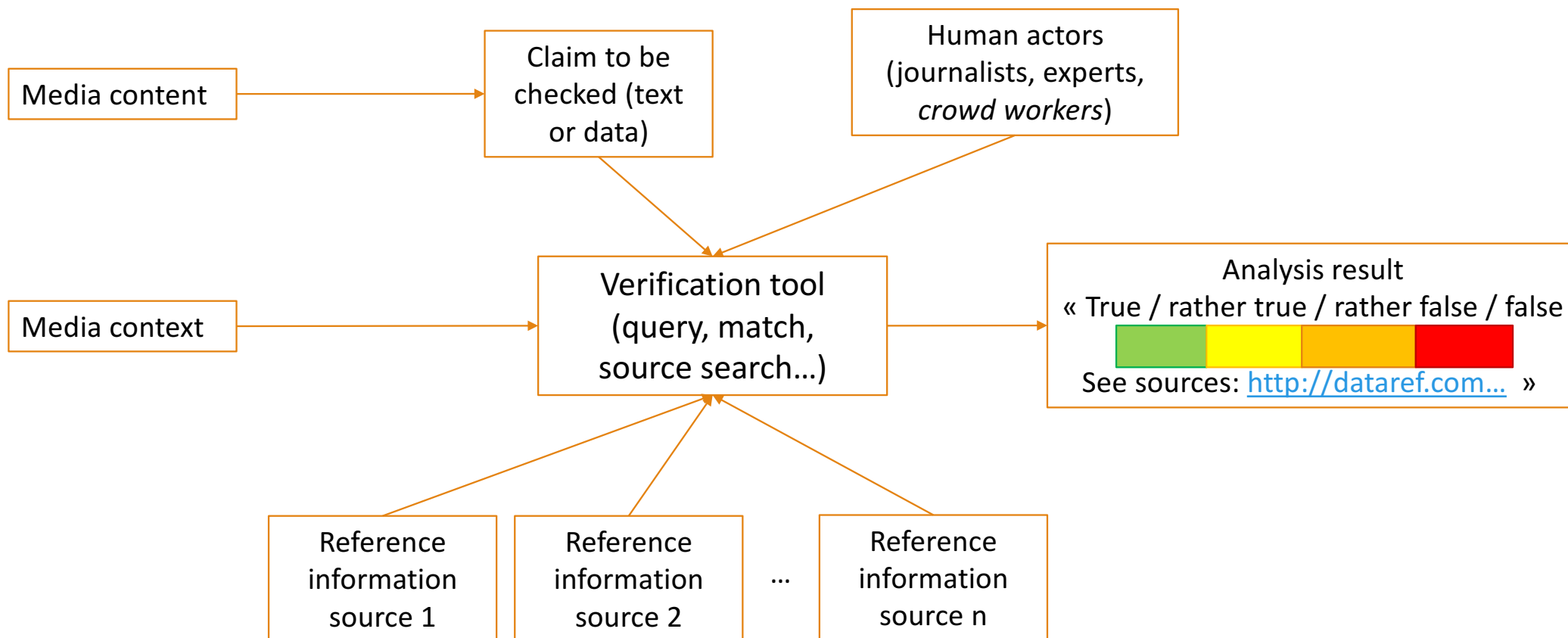
Fact Check

“it's pretty rich to have the guy who ran out
the lincoln bedroom sold seats on air force
one was an uninvited token”
01:09 - 01:24

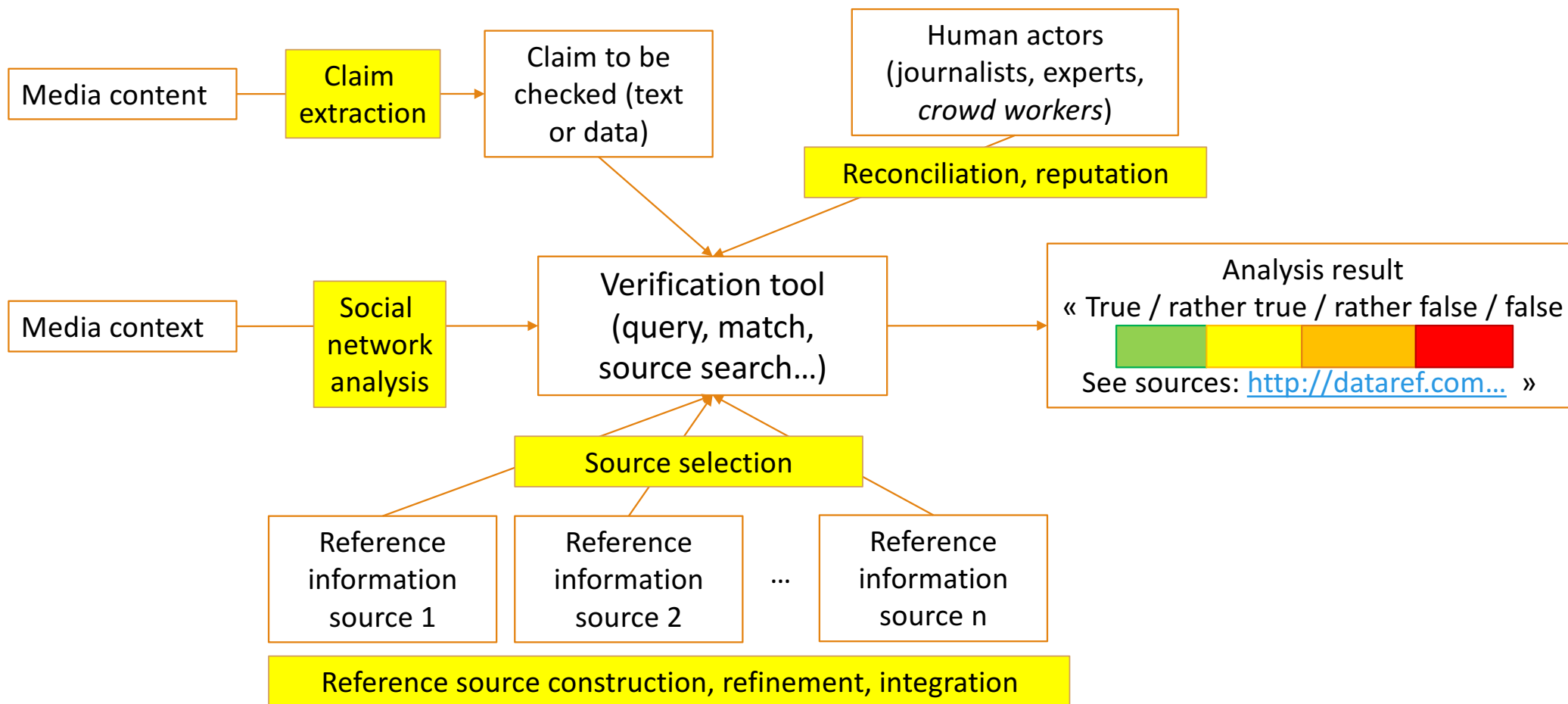

False

[Read a full fact check about this claim ...](#)

Fact-checking is a content management problem



Fact-checking is a content management problem



```
graph TD
    MC1[Media content] --> CE[Claim extraction]
    CE --> CTD[Claim to be checked<br/>(text or data)]
    HA[Human actors<br/>(journalists, experts,<br/>crowd workers)] --> RR[Reconciliation, reputation]
    CTD --> RR
    RR --> VT[Verification tool<br/>(query, match,<br/>source search...)]
    MC2[Media context] --> SNA[Social network analysis]
    SNA --> VT
    VT --> AR[Analysis result<br/>« True / rather true / rather false / false »<br/>See sources: http://dataref.com... ]
    AR --> SSS[Source search / source selection]
    SSS --> R1[Reference information source 1]
    SSS --> R2[Reference information source 2]
    SSS --> Dots[...]
    SSS --> Rn[Reference information source n]
    SSS --> Rn1[Reference information source n+1]
    R1 --> RS[Reference source construction, refinement, integration]
    R2 --> RS
    Dots --> RS
    Rn --> RS
    Rn1 --> RS
```

The diagram illustrates the DataRef framework for media verification. It starts with two inputs: 'Media content' and 'Media context'. 'Media content' flows through 'Claim extraction' to 'Claim to be checked (text or data)'. 'Media context' flows through 'Social network analysis' to the 'Verification tool (query, match, source search...)'. 'Claim to be checked' also flows into the 'Verification tool'. The 'Verification tool' outputs an 'Analysis result' (e.g., 'True / rather true / rather false / false') and a list of sources (e.g., 'http://dataref.com...'). The 'Analysis result' is visualized with a four-color bar (green, yellow, orange, red). The 'Verification tool' also flows into 'Source search / source selection', which identifies 'Reference information source 1', 'Reference information source 2', ..., 'Reference information source n', and 'Reference information source n+1'. These sources are visualized with similar four-color bars. Finally, all reference information sources flow into 'Reference source construction, refinement, integration'.

A point of caution: it's not just checking

Most aspects of modern reality are complex

From a journalistic perspective, explaining may be as important and useful as checking

Also: future is hard(er) to check



ContentCheck
ANR project
(2016-2019)

Toward automated fact-checking

FactMinder demo [SIGMOD 2013] (F. Goasdoué, K. Karanasos, Y. Katsis, J. Leblay, I. Manolescu and S. Zampetakis)

Browser plug-in

Bringing up rich context for a Web page, from document and knowledge bases

Source search

Supported by template queries over XML documents and RDF graphs

« Second screen »

The screenshot displays the FactMinder browser plug-in interface. The main content area shows a news article titled "Bill Clinton apporte son soutien à Barack Obama" from Le Monde. The article text is partially visible, mentioning Clinton's support for Obama's campaign. To the right of the article are four sidebars:

- Concepts:** A list of entities extracted from the document, including Bill Clinton (dbpedia:Person), Barack Obama (dbpedia:Person), Des Moines, Iowa (dbpedia:Place), Hillary Clinton (dbpedia:Person), Michelle Obama (dbpedia:Person), Bruce Springsteen (dbpedia:Person), and Pennsylvanie (dbpedia:Place).
- Facts & figures:** A section titled "Curriculum" containing biographical information about Bill Clinton, such as his birth date (August 19, 1946) and political party (Democratic Party). It also includes a portrait of Clinton.
- Related stories:** A list of related news items, including "Ohio, Florida, ... être compliqué" and "Des indicateurs ... d'Obama".
- Sources:** A list of sources, including "Romney... agressifs".

Content management tools for data journalism

Tatooine demo [VLDB2016] (R. Bonaque, T.-D. Cao, B. Cautis, F. Goasdoué, J. Letelier, I. Manolescu, O. Mendoza, S. Ribeiro, X. Tannier, M. Thomazo)

So many data sources, so little time!

~ data lake

« Can we **group tweets of politicians by political current** and analyze their **most frequent topics**? »

« Can we **classify articles** by their main **concept** and do a **tag cloud**? »

Can we **industrialize** answering such requests?



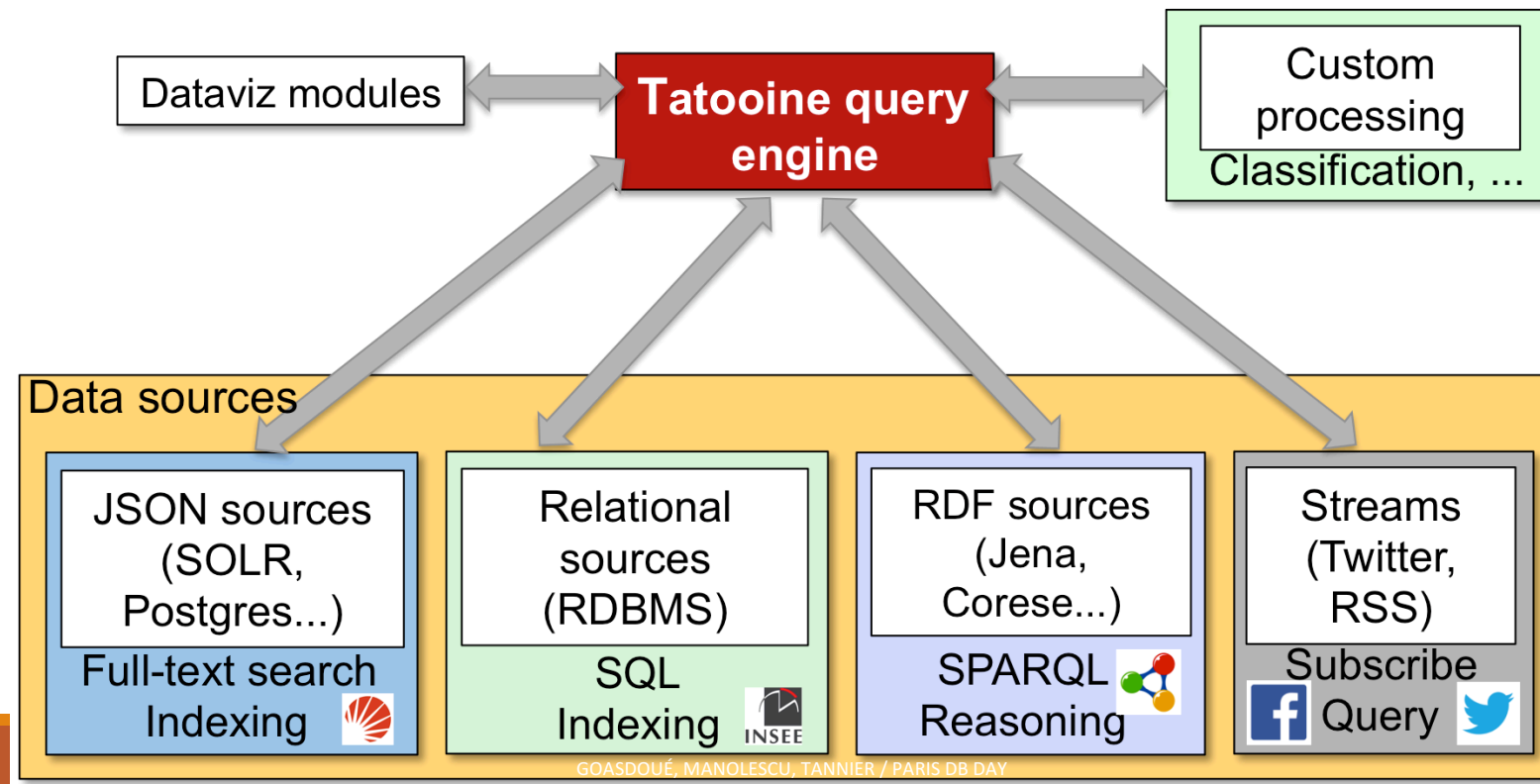
Content management for data journalism

Tatooine demo (R. Bonaque, T.-D. Cao, B. Cautis, F. Goasdoué, J. Letelier, I. Manolescu, O. Mendoza, S. Ribeiro, X. Tannier, M. Thomazo)

So many data sources, so little time!

Data journalism; reference source enrichment

Demo



Modeling facts, statements and lies

Ongoing work with Ludivine Duroyon (U. Rennes 1) and
Francois Goasdoué

We need to represent information such as:



1. Penelope worked for the National Assembly (NA) **from 2002 to 2012**
2. Her first work contract for the NA ran from 2002 to 2005 = **in 2002 the employee database showed that** she was going to work there from 2002 to 2005
3. **On Jan 15, 2017, François says** Penelope worked for the the NA from **2002 to 2007**.
On Jan 21, 2017, he corrects himself to state that Penelope worked for the NA **from 2002 to 2008**.
5. **On Jan 22, 2017, Le Canard Enchaîné wrote** that **François had stated on Jan 21** that Penelope worked for the NA **from 2002 to 2008**.
6. Charles and Marie both worked for the NA **at some point**, but **Charles' contract was after Marie's** (they never overlapped)

Modeling facts, statements and lies

Ongoing work with Ludivine Duroyon (U. Rennes 1) and
Francois Goasdoué



Approach:

1. **Facts** = RDF
2. **Timed facts** = bitemporal RDF (transaction time, validity time) [Gutierrez, Hurtado 2005]
3. **Beliefs (viewpoints)**: A states that B states that... [Gatterbauer, Suciu 2009]
4. **Timed beliefs**: at time T1, A states that at time T2, B ...

Defined:

- ☐ Saturation (set of all consequences of a set of statements)
- ☐ Query answering

TBD: query answering algorithm

Poster.

Linked Open Data extraction from INSEE statistic spreadsheets

[illegible]

Merci / questions?
