

Diego Antolinos Basso

email@diegodiego.net

@diegantobass@fosstodon.org

Whence

- I'm a research software engineer
- I'm a computational linguist
- I'm an open source software enthusiast

Work

- 10 years at Sciences Po's médialab
- Currently at UniNE's AJM

Disclaimer

- This is work in progress
- This is repetition of common knowledge
- This is an invitation/provocation to debate

Take up the torch

What does being recruited on an ongoing research project (should) mean?

Panda

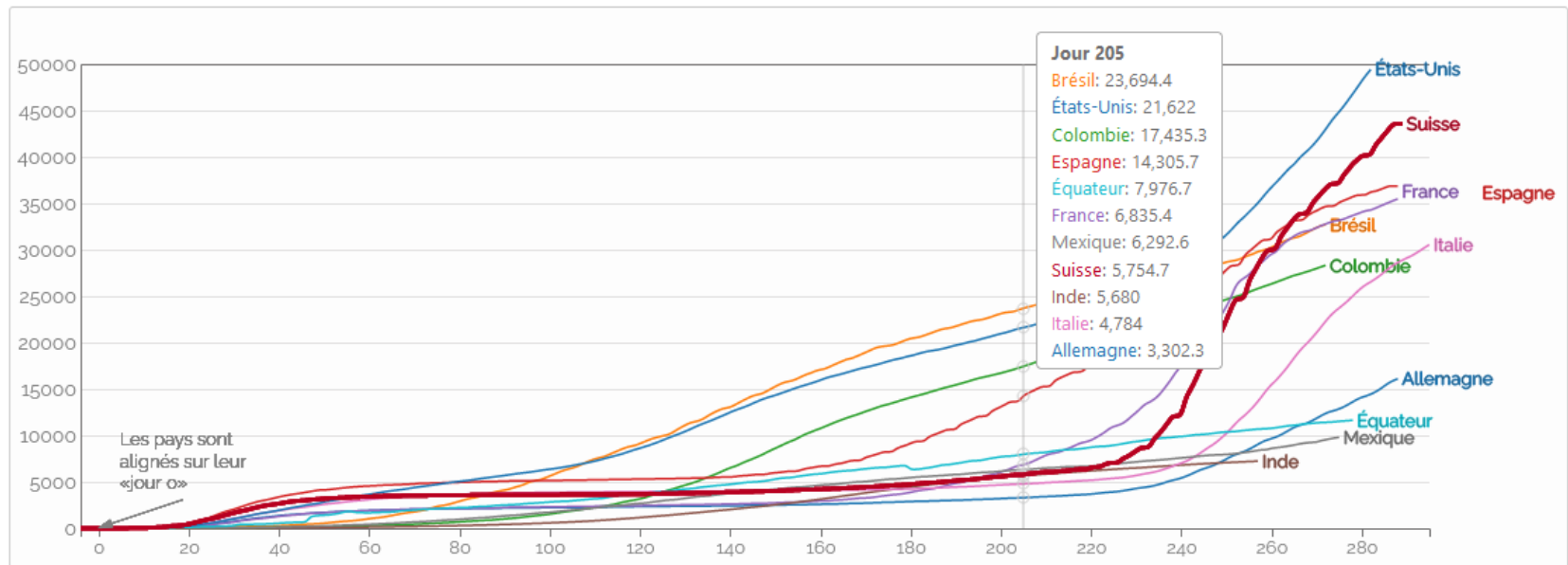
Pandemic Data: Production, diffusion et compréhension des données en temps de pandémie

- FNS NPR80 "Covid-19 in Society" 3 years project
- Team from UniNE's Faculté des sciences économiques
- "How did we produce and narrate data during the covid pandemic ?"

Panda is actually about the thousands of dataviz we got bombarded with during the pandemic

Taux de cas confirmés par million d'habitants

Les pays sont alignés sur le jour où les cas confirmés atteignent le taux de 1 cas par million d'habitants. **Pour la Suisse**, ce «jour 0» est le 29 février, avec 12 cas. Des différences dans les méthodes et le nombre de tests effectués rendent cependant les comparaisons entre pays difficiles.



Source: Centre européen de prévention et de contrôle des maladies (ECDC), données retraitées par Le Temps.

Note: le 25 mai, l'Espagne a réévalué ses bilans à la baisse (-1918 décès et -372 cas). Ces chiffres ont ensuite été fortement corrigés à la hausse.
















Today's talk

I was recruited in the project to fill in for a PhD student that left the project after a year

What does/should it allow/force you to do..?

**1 - Can documentation mean
"passing the torch"?**

I was handed a lot of python notebooks

 data_raw	backup data: fix all missing dates in dynamic dataviz
 data_source	Handover
 figures	Handover
 panda-iframe-analysis	recuperation info des dataviz en unes Heidi
 1_RAW_Get_Dataviz_Dom_2_METADATA_Get_Dataviz_Metadata.ipynb	recuperation des dates de publi manquantes
 1_RAW_Get_Datawrapper_Script.ipynb	Handover
 1_RAW_Merge_RTS_Corpus.ipynb	Handover
 2_METADATA_Get_Datawrapper_Metadata.ipynb	recuperation des dates de publi manquantes
 2a_Categorize_Datawrapper_Type.ipynb	Handover
 3_SOURCES_Aggregate_Source_First_Pass.ipynb	Handover
 3_SOURCES_Aggregate_Source_Second_Pass.ipynb	Handover
 _3_DATAVIZ_Stat.ipynb	Handover
 _3_SOURCES_Over_Time.ipynb	Handover
 _3_SOURCES_Stat.ipynb	Handover
 _3a_Categorize_Sources.ipynb	Handover

It reminded me how

- Commenting code is fundamental
- Documenting the data processing pipeline is key
- Keeping a chronological exhaustive journal helps
- Linking technical aspects to the global project gives meaning to an engineer's work

**2 - Sweeping dust under the
carpet**

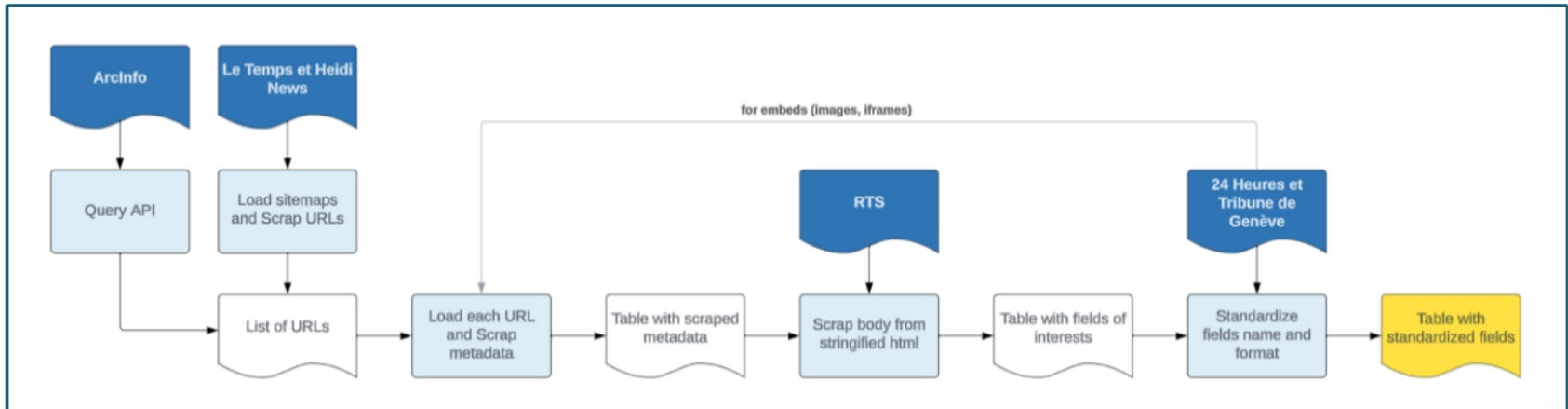


Figure 1: Flowchart of data collection and cleaning

Credit : Céline Dupuis, Panda 2024

(Data acquisition of the Panda project was a technical rollercoaster)

We should try not to hide complexity and "dirty parts", but be rigorous and document caveats

Because

- Tracing back undocumented steps costs more than keeping a clean log
- Re-running the code and getting different data harms confidence

Publication

- The "method" section of an article rarely suffices
- A methodology or data paper counts as a publication

Reminder

Not everyone finds this type of work tedious and unrewarding!

3 - Interdisciplinarity and competences

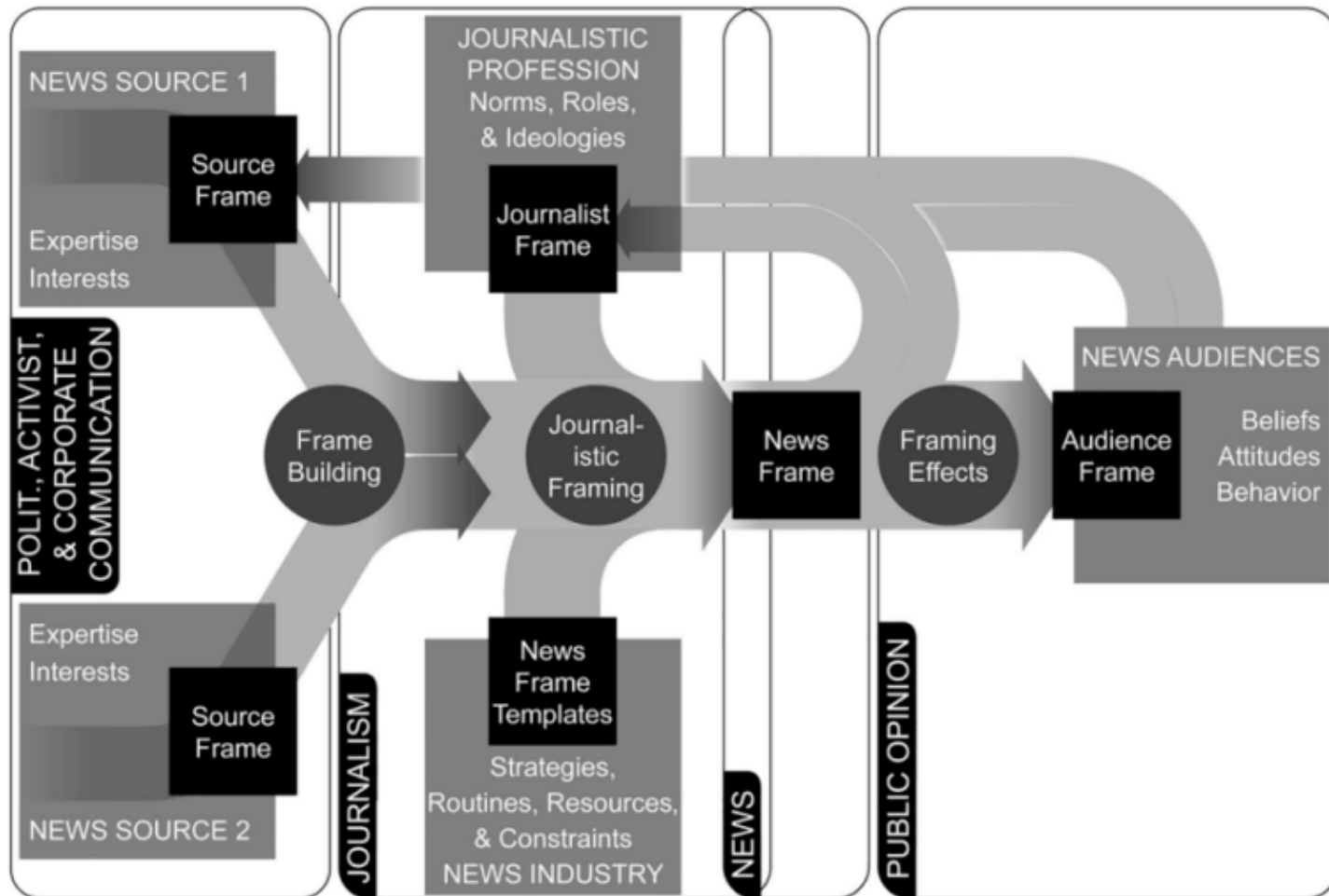


Figure 15.1 Model of the News Framing Process

Basic knowledge for some, new to many

Credit: Baden, C. (2019). Framing the news. In K. Wahl-Jorgensen & T. Hanitzsch (Eds.), The Handbook of Journalism Studies (pp. 229–245). Routledge

Sharing a bibliography

- Participating in a huge Zotero database is a good thing (and including technical/tools references)
- Having a subset of "must-read" papers is a good idea for researchers but also for engineers

Ultimately, how can we allow new-comers to shift perspectives by bringing in new ideas?

Final thoughts

I can talk today about all of this specifically because in Panda, a lot of care is given to all of these supposedly peripheral aspects

- **How can we make sure research projects take it (maintenance, documentation, care, etc.) into account?**
- **How can we get the point across to professors that are submitting projects and recruiting us into their projects?**