

Monitoring Open Access at a national level

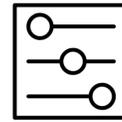
French case study

Eric Jeangirard – Ministry of Higher Education, Research and Innovation, Paris, France
Contact: eric.jeangirard [at] recherche.gouv.fr

Why a national open science monitor ?



Measure on a regular basis the Open Access **trends** in France from **trustworthy and open data** that can be decomposed into **fields** and **institutions**



Steer open science (part of the first commitment of the French National Plan)

A transparent methodology

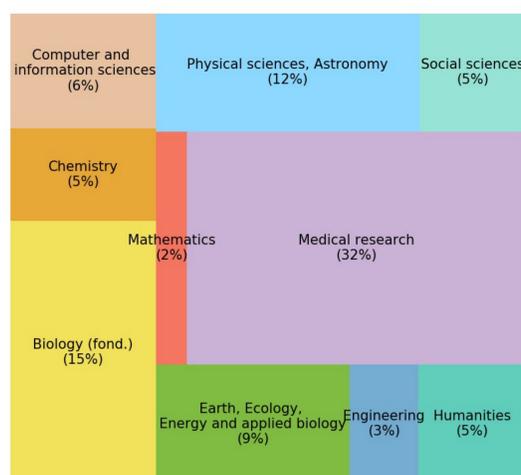
1 Identify the publications with a french affiliation

- For now, we restrict to publications with a DOI (which has an impact at the field level).
- We use data from Unpaywall / CrossRef and HAL (open repository) to list all the publications published during the year (about 4 million DOIs per year). Few meta-data are available: title, authors name, source details. **Affiliations data are missing in most of the cases** (when data comes CrossRef).
- If the affiliation is not available on HAL, we designed a tool that first checks if one of the authors is in our authors referential (built specifically from our use case) and if so, checks the affiliations on the publication's webpage (using the DOI redirect).
- The affiliations meta-data are key for building an OA monitoring at a national level. The accuracy of the method is estimated manually with a sample of 100 publications. **The false positive error rate is estimated at 4%.**

2 Enrich the meta-data of these publications

Meta-data enrichment is a key aspect to build a granular monitor, especially for scientific fields decomposition or affiliations decomposition (institutions, structures). Concerning scientific fields, a supervised machine learning model has been built from the labeled database Pascal-Francis (23 millions labeled scientific articles). Also, we identify the authors and the detected affiliations with external identifiers (IdRef for authors, RNSR, Sirene and Grid for affiliations).

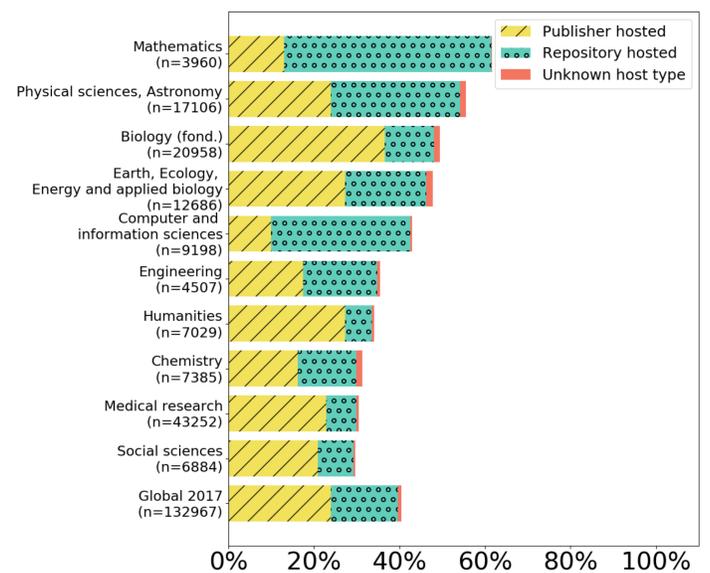
Scientific field share estimation for the publications detected (2017)



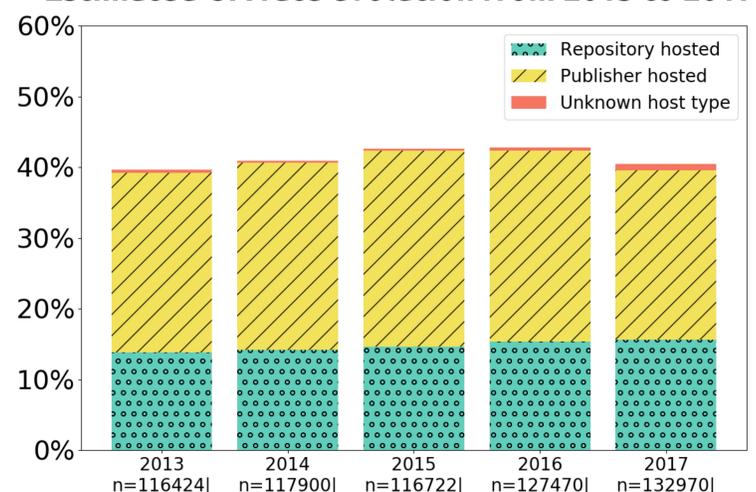
3 Figure out which publications are Open Access

We use the Unpaywall service to detect Open Access (based on DOI).

OA rate estimation per scientific field for the detected publications (2017)



Estimated OA rate evolution from 2013 to 2017



The trend is slightly up, except for the last year, but we gather evidence that shows this is a consequence of the moving nature of the OA status. Therefore these figures should be seen as a snapshot rather than definitive.

Next steps

Strengthen the methodology

- Improve the french affiliation detection.
- Use the Unpaywall feed to update the monitoring on a regular basis
- Analyse the OA status dynamics (when does a publication become OA ?)
- New sources inclusion (publications with no DOI)

Open Science, beyond Open Access

- Append open sources for citations
- Open Data and Open Source monitoring
- Funders and research journal policies