

Methodologies for Monitoring Open Science

ELPUB 2019 - Workshop

June 2019

Notes for the session

https://frama.link/open_science

Agenda

Part 1: Presentations and discussions : A landscape of existing tools and methodologies

Presentations of existing work
Expectations for the worksop

Part 2: Group discussions

For you, what is the goal behind monitoring Open Science? And Open Access in particular ?
What are your expectations concerning the data sources used?
What are your expectations concerning the outcome of the monitoring, esp. for Open Access ? (level of detail, openness, impact ...)
Could you formulate any recommendation?
Is there any potential interest in data sharing, result sharing, or even collaboration?

Part 3: Group discussions

Challenges of monitoring Open Science, beyond Open Access

For you, beyond Open Access itself, what are the stakes involved? (policies, research data, open code ...)
What are the main challenges ?

French Open Science Monitor

June 2019

AGENDA

- Objectives of the French OSM
- Methodology and results
- Conclusion and further work

Objectives of the French Open Science Monitor

■ Steer Open Science in France

Part of the first commitment of the French National Plan for OS

■ Measure on a regular basis the Open Access trends in France

- from trustworthy and open data
- that can be splitted by fields and institutions
- that can be re-shared and enriched with external identifiers

A 3-step methodology

- 1 - Identify the publications with a french affiliation
Funnel approach: from a comprehensive list of publications (as much as possible) to a list of french publications (one author has a french affiliation)

- 2 - Enrich the meta-data of these publications
 - Scientific field
 - Authors identification
 - Affiliations identifications

- 3 - Figure out which publications are Open Access
 - Using the Unpaywall service (DOI based)

Methodology (1/3)

Identify the publications with a french affiliation

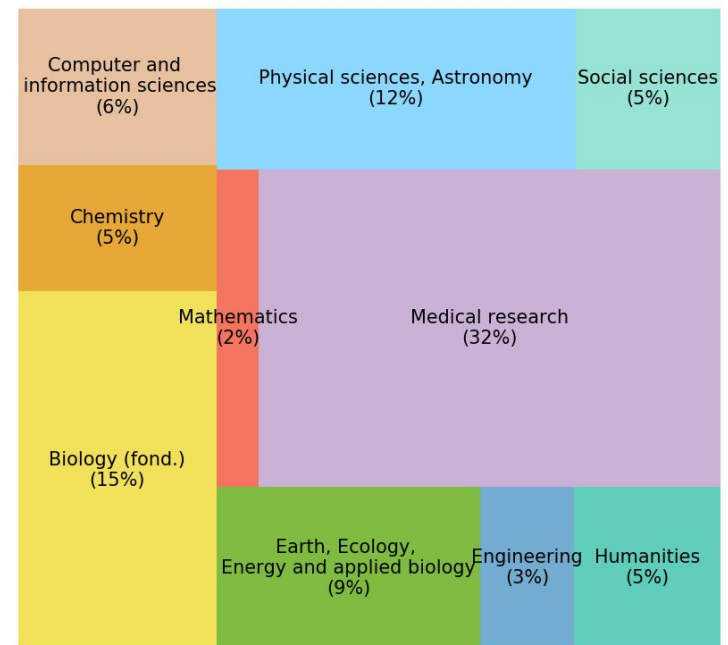
- We use the full Unpaywall database (based on Crossref)
 - which limits to publications with a DOI
 - that represents about 4.5 millions DOIs per year
- We detect publications with a french affiliations
 - First filter (rough) based on the author names (from a names database constructed from theses.fr, HAL, ORCID, Pascal&Francis ...)
 - Second filter based on affiliations
 - from HAL is metadata available
 - from scrapped webpages (DOI redirection) otherwise
- We estimate the accuracy of the detection: 4% false positive

Methodology (2/3)

Enrich the meta-data of these publications

- Metadata enrichment is a key aspect to build a granular monitor
- Scientific field
 - Machine learning technique to infer the scientific field from title and journal name
- Identification of the authors through IdRef (and ORCID)

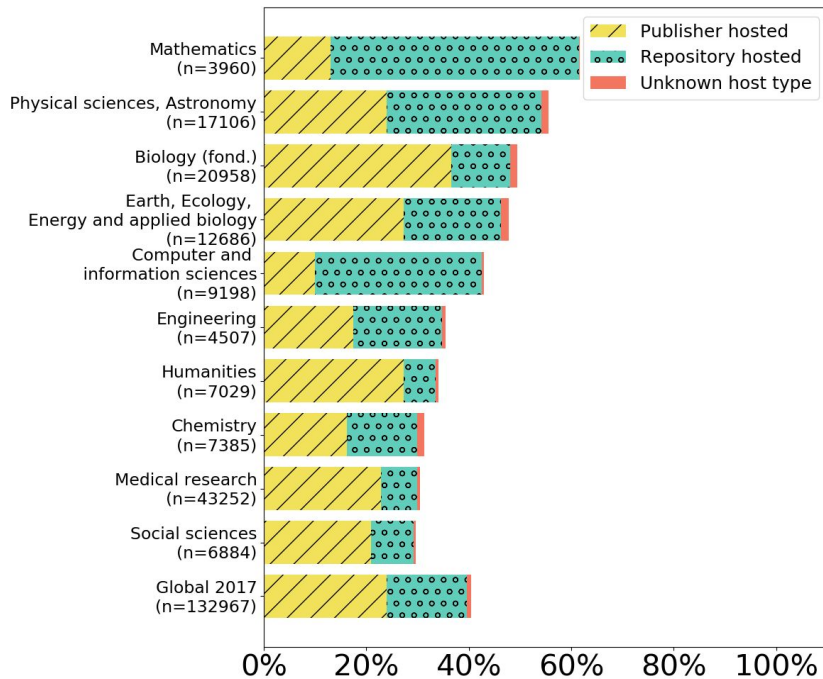
Scientific field share estimation for the publications detected (2017)



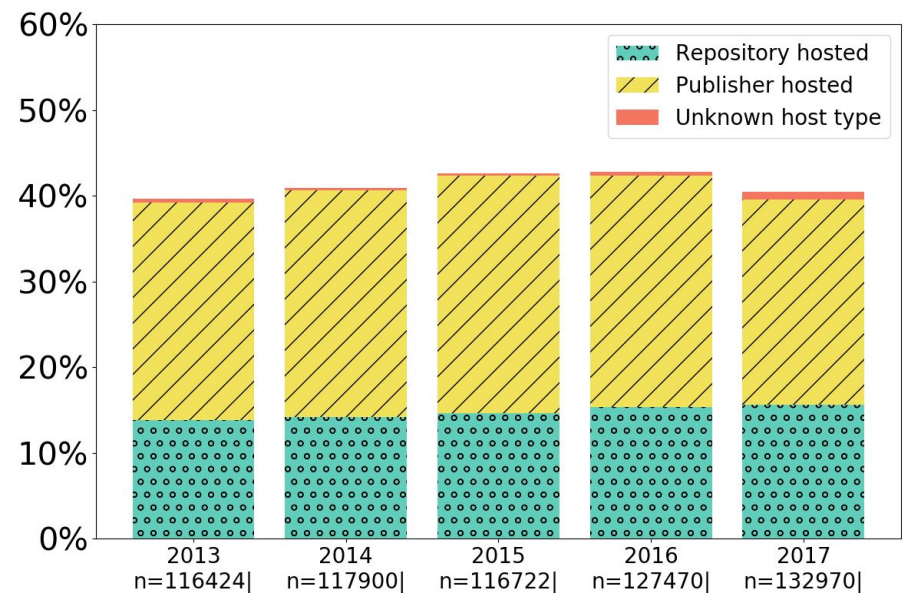
Methodology (3/3)

Open Access rate estimation with Unpaywall

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



Estimated OA rate evolution from 2013 to 2017



Conclusions

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