A new framework for the French Open Science Monitor (BSO)

Anne L'Hôte[1], Eric Jeangirard[1], Didier Torny[2], and Laetitia Bracco[3]

[1]French Ministry of Higher Education, Research and Innovation, Paris, France
[2]CNRS, France
[3]University of Lorraine, France

February 2022

Keywords: open science, open access, unpaywall, clinical trials, observational studies, scientometrics

1. Introduction

The French Open Science Monitor was launched in 2019 as part of the first French National Plan for Open Science (MESRI 2018). Its methodology was detailed in (Jeangirard 2019). It currently focuses on scholarly publications, for which at least one author has a French affiliation. It measures the rate of open access for these publications. It will eventually be extended to other dimensions of Open Science, whether they are transversal (management and opening of research data and softwares) or disciplinary.

To support the continuation of Open Science public policy with the second National Plan for Open Science (MESRI 2021), a new framework for the French Open Science Monitor has been produced. It introduces a monitor specific to the Health domain and also develops the features for the Open Access analysis.

The main goal of the French Open Science Monitor is to produce a dynamic vision of the openness level evolution and to analyse in detail how publications are opened, developing indicators specific to open repositories on one hand and indicators specific to the dissemination platforms on the other hand.

The objective of the French Open Science Monitor in Health is to report on some aspects of Open Science specific to medical research and health, in relation to the sharing of scientific knowledge that has become a paramount urgency during the COVID-19 pandemic. The aim is to have indicators that will make

it possible to take stock of the situation and monitor the public policies that will be implemented.

In addition to the open access to the publications, which is critical for all domains, the registration of clinical trials and observational studies, the publication of their results and the sharing of their data are specific dimensions in the Health domain, and more particularly of clinical research.

Clinical trials are research conducted on human subjects involving an intervention other than their usual care (delivery of a drug, treatment device, surgical procedure, etc.) for the purpose of developing biological, medical or public health knowledge.

Observational studies are "non-interventional" studies, also involving humans, but not involving any intervention other than the usual management of patients. They may focus on protocol compliance, adverse effects of a treatment after it has been put on the market, etc. This is the case, for example, with cohort studies, which consist of statistical monitoring of a panel of individuals over the long term in order to identify the occurrence of health events of interest and the related risk or protective factors.

This clinical research is subject to various biases, including publication biases, which are well identified by public health researchers. Amongst them, the most known is the tendency to publish only trials and studies whose results are conclusive and in line with the expectations of the researchers who carried them out (these are known as "positive" results). The consequence of this bias is that the syntheses or meta-analyses carried out on the basis of scientific publications with a view to guiding public health policies are in fact based on a partial and biased view of scientific knowledge.

Two main ways exist to correct this bias:

- systematic declaration of studies, before they are carried out, in dedicated registers;
- systematic publication of study results, even when they are "negative", for example through initiatives like Registered Reports.

Regulations have been implemented to improve transparency: in the United States, the declaration of clinical trials and their results is compulsory, and in Europe, the declaration of clinical drug trials will be compulsory as of 2022. In contrast, observational studies are not subject to any regulations regarding their reporting or publication.

2. Method

2.1 Publications

2.1.1 Perimeter definition

2.1.1.1 French Open Science Monitor The French Open Science Monitor is a tool that aims at steering the Open Science policy in France. As such, it produces statistics that are analyzed over time, and it has to focus on "French" productions. Also, as stated in (COSO 2018), we want to use only public or open datasources. Two constraints of perimeter thus appear naturally :

- **only publications with at least an author who has a French affiliation** are considered. The nationality of the authors does not come into play. Still, this raises the issue of access to affiliation information. Affiliation metadata are present in specific sources, like PubMed, but very rarely in the whole Crossref data. To fill in the gaps, we propose to crawl the affiliation information displayed publicly from the publications webpages. On top of that, identifying a country from an affiliation text is not that straightforward. If you are not convinced, think about an affiliation stating "Hôtel Dieu de France, Beirut, Lebanon": this does not refer to a French affiliation even though the word "France" is present. We use an automatic detection algorithm, based on Elasticsearch, described in (L'Hôte and Jeangirard 2021), to infer the countries from the affiliations field.
- **only the publications with a Crossref DOI** are considered. Duplicates have to be avoided, in order not to count twice (or more) a publication and thus add a bias to the statistics tha are produced. It is then key to use a Persistent Identifier. Also, we choose to use Unpaywall data for Open Access (OA) discovery. This service produces open data and offers the possibility to snapshot the whole database, which is an asset to analyse the OA dynamics. For now, Unpaywall focuses only on Crossref DOI, which leads us to adopt the same perimeter. We are aware that this is a bias against some disciplines, most notably Humanities and Social Sciences.

All genres of publications are considered (journal articles, proceedings, books ...) as long as the publication is associated to a Crossref DOI. Many types are being coded in the metadata, but for the sake of clarity, we group them in categories, namely journal articles, proceedings, preprints, book chapters, books, the rest being grouped in a category 'Others'. It is important to note that the 'preprint' type does not appear as such directly in the available metadata (it is generally declared as a journal article). Some preprint detection is based on the dissemination platform information. At the time this article is written, only the Cold Spring Harbor Laboratory (BioRxiv, MedRxiv) case is covered, but it can be extended as soon as another preprint dissemination platform would start using Crossref DOIs, as for example ArXiv has planned it.

2.1.1.2 French Open Science Monitor in Health The French Open Science Monitor also introduces a focus on the Health domain. Delimiting a clear perimeter for Health is not very easy. For now, we simply have chosen to consider in the scope **all PubMed publications, and only these**. The publications' data used in the French Open Science Monitor in Health is then a subset of the publications described above, adding the PubMed presence criterion. Note that "Health" is seen more as a domain than a discipline. In fact, publications from a lot of disciplines are taken into account in the French Open Science Monitor in Health. A domain-specific set of disciplines is used in the French Open Science Monitor in Health, as described below.

2.1.2 Open access dynamic

From the first edition of the French Open Science Monitor, it was clear that the open access rate was far from stable, so we should try to capture the opening dynamics (Jeangirard 2019). Indeed, the immediate open access exists but we cannot assume it represents the totality of the open access, considering the various publishers, funders and national embargo policies. Therefore, for a given set of publications, say the publications published during the year Y, it makes sense to measure the open access rate at different point in time, for example at some moment in year Y+1, Y+2 …

To do so, it becomes necessary to historicize the database containing the open access information. So, instead of maintaining a database that keeps track of the opening of each publication, which is the current Unpaywall data policy, we have to make regular snapshots of the whole Unpaywall database. Each snapshot is used as an observation date to measure the open access rate. It is important to note that this method natively embeds the potential open access discovery errors from the underlying Unpaywall database, that can be false negative (a publication is actually open at this point in time but it is not detected) or false positive (wrongly seen as open whereas it is closed). As a side note, it would also allow us to follow "temporary open" publications, resulting from new publishers policies adopted for Covid-19 related publications.

This method of analysis therefore reveals two temporal dimensions: publication dates and observation dates. Obviously, the observation date must be after the publication date. To avoid that the proliferation of possible analyzes blurs the message, we propose to look mainly at two elements :

- A main statistics that is the **1Y Open Access rate**: it represents the open access rate of the publications published during year Y and measured (observed from the snapshot of the OA discovery database) at one point in time during year Y+1 (generally in December if the data is available).
- Also, the **shape of open access curve** (open access rate function of the publication year). For a given observation date, the open access rate can be estimated broken down by publication year. This then produces a curve of the open access rate as a function of the publication year (at a

given point in time which is the observation date). This curve may have any shape, and in particular it is not always expected to be a monotonic increasing. Indeed, a monotonic increasing curve means that more recent publications are more and more open. That can (hopefully!) happen, but moving barriers and embargoes would genereally lead to another type of shape, that would be an inverted-V shape. The next figure illustrates different shapes of Open Access curves.

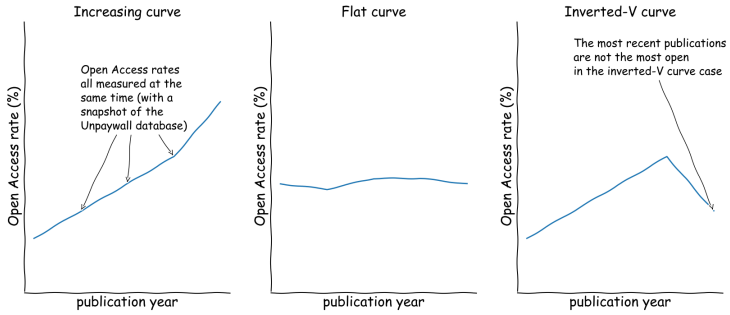

Figure 1: Different shapes of Open Access curves

From an observation date to another, the OA curve shape may change. This evolution of the shape gives an insight of the opening speed. Indeed, moving from an inverted-V shape, where the most recent papers are not the most open, to an increasing shape would be a proof of the opening acceleration. The next figures illustrates the evolution from an inverted-V shape, to flat and then to an increasing OA curve shape.

2.1.3 Open access types

As Unpaywall is the Open Access discovery tool we used, we initially based our results on the OA classifications described in (Piwowar et al. 2018). It breaks down the OA types in 5 categories: 'Gold', 'Hybrid', 'Bronze', 'Green', 'Closed'. These categories are also present in the Unpaywall database (and oaDOI API) in the field 'oa_status'. We first simply grouped the categories 'Gold', 'Hybrid' and 'Bronze' under a 'Publisher hosted' label. However, we now propose another classification that we think more appropriate for the French OA policy steering.

(Piwowar et al. 2018) defines 'Green' as 'Toll-access on the publisher page, but there is a free copy in an OA repository'. That implies that a publication that would be free to read on the publisher webpage and that would, at the same time, have a free copy on a repository would not be counted as 'Green'. That derives from the idea that the Version of Record (VoR), available on the publisher website, is the preferred OA version of the publication. As a consequence, the contribution of the repositories is mechanically reduced in favour of the pub-

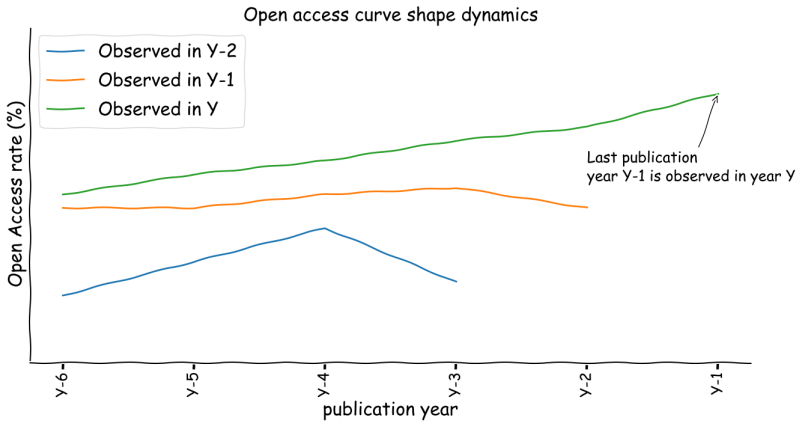

Figure 2: Open Access curve dynamics

lishers. This therefore blurs the picture of the extension of repositories impact. That led us to propose a first level of analysis, with 3 categories (excluding 'Closed'):

- **hosted only on an open repository**: Toll-access on the publisher page, but there is a free copy in an OA repository, corresponging exaclty to the 'Green' definition of (Piwowar et al. 2018), that we could rather label 'Green only'
- **hosted only by the publisher**: Free to read on the publisher webpage, but no free copy in any OA repository harvested by Unpaywall.
- **hosted on an open repository and by the publisher**: Free to read on the publisher webpage and there is a free copy in an OA repository.

Obviously, this does not impact the overall Open Access rate, but this balanced division, with no preference for the VoR, gives a different picture. The next figure shows the kind of impact choosing one or the other OA type break down.

Another graphical way to represent this balance is to use a bubble chart. Each bubble represents a cluster of publications (one bubble is the equivalent for each discipline, for each dissemination platform ...), its size depends on the number of publications in the cluster. The x-axis represents the share of OA publications hosted by the publisher, corresponding to the sum of publisher-only and publisher / open repository hosted publications. Conversely, the y-axis represents the share of OA publications hosted on a repository, corresponding to the sum of open repository-only and open repository / publisher hosted publications.

The source of data used to compute these OA types is still Unpaywall, but instead of the 'oa_status' field, we use the 'oa_locations' field. For a publication

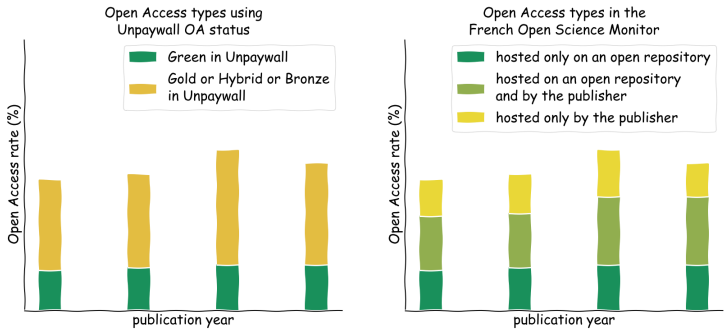

Figure 3: Open Access hosting types

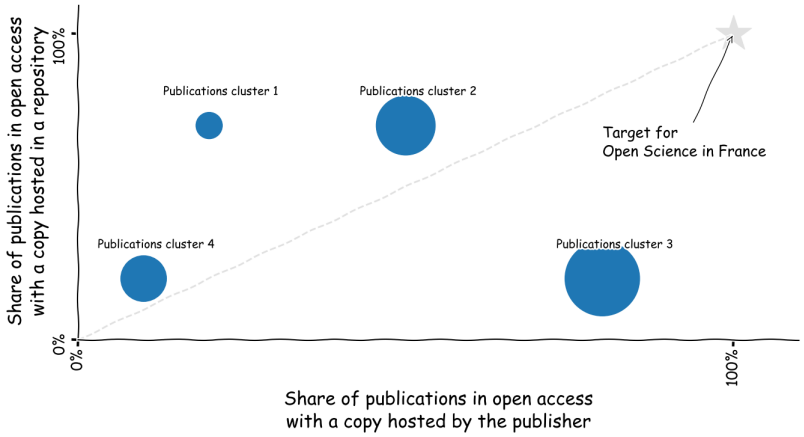

Figure 4: Share of publications in open access hosted on an open repository vs. by the publisher

which is in open access, it lists all the existing free copies that Unpaywall detected, at the time of the snapshot. Each location is described, in particular with an URL that gives a link to the free copy, and some metadata for the location is associated, in particular the 'host_type', that can take two possibles values, 'publisher' or 'repository'. It is important to note that, for now, preprint servers are considered as repositories.

2.1.4 Discipline and language impact

All disciplines and publication languages are covered, while no metadata exists to describe the discipline or the publication language. To enrich the metadata, we then rely on machine learning approches, that try to infer discipline and language from the available metadata.

For the language detection, only the title and abstract are used if available, with the lid.176.bin fasttext word embedding machine learning algorithm (Joulin et al. 2016).

Discipline detection also uses journal and keywords metadata if available. A general classifier is implemented for all domains, which classifies the publications into 10 macro disciplines: Mathematics, Chemistry, Physics & astronomy, Fondamental biology, Medical research, Computer sciences, Earth science ecology energy & applied biology, Humanities, Social sciences, Engineering. It is trained on data from the Pascal & Francis database and uses a Fasttext classifier. More details are discussed in the previous paper (Jeangirard 2019).

A domain-specific classifier is implemented for the Health domain. It classifies the publications into 17 disciplines, built from the Fields of Research taxonomy. The full methodology is detailed in (Jeangirard 2021).

The main purpose of these metadata enhancements is to be able to analyse the open access rate in function of languages and disciplines. We expect to observe differences not only in the global OA rate (which discipline is the most open ?), but also in the dynamics trends (which discipline show the strongest increase over time ?) or in the opening uses (relying on publisher hosted open access versus open repositories).

2.1.5 Publishers and dissemination platforms strategies

2.1.5.1 Identification of the dissemination platforms The data in the 'publisher' field of Crossref shows many inconsistencies. There are many journals, with a single ISSN, that belong to more than one publisher - whether they are different lexical forms or really different entities. Consequently, we have made a triple grouping in order to favour the coding of an economic entity diffusing the journal in question.

- Firstly, we considered the diversity of lexical forms of the same publisher, existing in developed form and in the form of acronyms, or without and with its economic status (LLC, Ltd.,…);

- Secondly, we have taken into account the capitalist evolution of the sector, which is marked by a growing concentration, with successive takeovers. The latter do not necessarily make the old group names disappear, most often used as a brand name;
- Thirdly, we have taken into account the separation between publisher and dissemination platform, with many scholarly societies remaining the owner and publisher, but delegating the dissemination of their publications to the publisher.

We historicized the last two groupings to account for the effective date of the link between these different entities. All coding is available in the open source code hosted at https://github.com/dataesr/bso-publications/tree/main/bso/server/main/publisher

2.1.5.2 Business models and open licenses As explained above, the 'oa_status' in Unpaywall data hides some part of the role of open repositories. It also hides Diamond open access, that is to say it mixes in the same 'Gold' category all publications published in an open-access journal that is indexed by the DOAJ, whether Article Process Charges (APC) were paid or not. That is why we introduce another level analysis, about the dissemination platform business model, with 3 categories :

- **Diamond**: journal-article published in an open-access journal indexed by the DOAJ, and without APC (according to the DOAJ data). This category may be under-estimated as some journal have no APC but are not in the DOAJ.
- **(Full APC) Gold**: publications published in an open-access journal (using the field 'journal_is_oa' = True from Unpaywall) and with APC.
- **Hybrid**: publications published in a journal that is not full open access (using the field 'journal_is_oa' = False from Unpaywall) and with APC.
- **Other**: all other cases, in particular publications with moving barriers, but also cases for which no information about APC has been collected. This catrgoy may be over-estimated as some journal have no APC but this information is not present in a structured database.

The objective of this level of analysis is to separate different business models (APC vs Diamond vs Hybrid), not to analyse the open licenses associated to the OA copies, so this categorization is quite different from the Gold / Hybrid / Bronze from (Piwowar et al. 2018).

For that matter, a third analysis level is used that distinguishes, for open access publications:

- **Creative commons** licenses (cc0, cc-by, cc-by-nc etc …)
- **Other licenses** (publisher specific, country-specific …)

- **No license**

To be clear, the no license category does not mean that the publications are closed, on the contrary they are in open access but no open license was detected, meaning the reuse conditions, beyond reading, are not defined.

Again, the informations from the field 'oa_locations' comes from Unpaywall, therefore the results depend on the quality of the Unpaywall database.

2.1.5.3 Article Processing Charges (APC) estimation Estimating APC for each journal article remains difficult as few open sources exist. We leverage on the openAPC database (at the publication level) and on the DOAJ data (at the ISSN level). We use the following heuristics to estimate the APC of a publication :

- If the DOI is not open access with a free copy on the publisher webpage, there is no APC estimation to make.
- Else, if the DOAJ specifies there are no APC for the ISSN, then it is a Diamond DOAJ OA, with no APC.
- Else, if the DOI is explicitly in the openAPC database, we simply use the APC from openAPC.
- Else, if the DOI is not in the openAPC database, but its ISSN or publisher is, with a sufficient number of observations, we use the mean of the APC observed for the same ISSN or the same publisher, during the same year if enough data is available, or over the whole openAPC database otherwise.
- Else, if the DOAJ specifies there are APC for the ISSN, we simply use the APC from DOAJ, after a conversion to Euros if needed (based on the exchange rate at the publication date).
- Otherwise, no estimation is made.

We are aware that this estimation is far from being perfect, but it still brings some insights. On top of that, even if we focus on French publications (publications with at an author with a French affiliation), the sum of the APC estimated is higher than the real amount of APC money spent by French institutions, as a large share of the publications are co-authored with scholars affiliated to foreign institutions. Informations on the corresponding author could be a proxy to focus on APC spent by France but for now, we do not have an open, reliable and massive source for this information.

2.1.6 The role of the open repositories

2.1.7 Other analysis axis

In the case of the Health domain, we use metadata coming from PubMed. These metadata are quite rich and enable extra analysis. In particular, some **funding**

metadata are present in PubMed, as well as the **affiliations for each author** (it is not always the case when using other sources and scrapped metadata).

PubMed gives information on grant declaration. To be clear, the absence of this metadata does not mean that there was no specific funding leading to the given publication. So the only thing we can do is to check whether a correlation exist between the open access rate and the presence of the grant metadata in PubMed.

As the affiliations information is given for each author, we can use (L'Hôte and Jeangirard 2021) to infer the country of affiliations of each author. We wish to analyze whether the country of affiliations of the corresponding author correlates to the open access rate or not. Unfortunately, the corresponding author metadata is not available, that is why we chose an approximation looking at the affiliation country of the **first and the last authors**. That will give an insight to know whether, for French publications, the OA rate is in general higher when one of the first or last authors has a French affiliation, or, conversely, if the OA rate is higher when the first and last author are affiliated abroad.

2.2 Clinical trials and observational studies

The French Open Science Monitor focuses, for now, only on publications. Current work is being conducted on monitoring also Research Data and Software Code. The French Open Science Monitor in Health, however, already introduces new research objects specific to the Health domains: the clinical trials and the observational studies.

In the US, reporting and publication of results is mandatory for all clinical trials. The reporting registry used is https://clinicaltrials.gov/. This site is also used by many international actors. It also welcomes the report of observational studies, though this reporting is not mandatory.

In the European Union, the reporting obligation will only extend to clinical drug, from 2022 on. The European registry https://www.clinicaltrialsregister.eu/ (EUCTR) therefore mainly includes clinical trials involving medicines, and less frequently observational studies, clinical trials involving surgical protocols, medical devices or psychotherapeutic protocols.

The issue of opening up or sharing data arises for clinical research in the same way as for other areas of scientific research. However, it has a particularly complex dimension, since it involves personal data, some of which directly concern the health of individuals. Nevertheless, it is possible to define the modalities for sharing this data.

Two dimensions will be developed:

- The openness of the results and publications when the study is completed.
- The declaration of clinical and observational studies in these public registries.

2.2.1 Perimeter

Two datasources are used for now to collect metadata about clinical trials and observational studies: clinicaltrials.org and EUCTR. clinicaltrials.org proposes an API while EUCTR does not; that is why the information is crawled from the website. Only the trials and studies that involves at least **one location in France** are analyzed.

Some trials or studies appear in both registries, the matching between the two databases being done based on the PIDs NCTId (from clinicaltrials.org) and eudraCT (from EUCTR), both registries keeping track of external PIDs. However, duplicates may still remain if no link was established between the existing PIDs in both registries.

To distinguish clinical trials on one side and observational studies on the other, we use the study type field, that can be either 'Interventional' (for clinical studies) or 'Observational' (for observational studies).

2.2.1 Main opening indicators

Mainly two types of indicators are analysed:

- The declaration of results and / or scholarly publications after a trial or study is completed. (Goldacre et al. 2018) showed that a large fraction of trials do not report their results. On top of the results declaration rate itself, we look into the results' date of registration, showing how much time has passed between the end of the trial and the actual date when the results are reported.

We propose both indicators mixing or separating results and scholarly publications. For the publications, it is important to note that only the metadata from the studies registries are used, without trying to link trials to DOIs using the publications metadata (whith PubMed for example). The open access status of these publications is also retrieved from the Unpaywall data.

- The delay to register the study: is the trial or study publicly registered before it actually starts, or is it done after ? And what is, in month, the actual delay to register ? Does it evolves over time ?

2.2.2 Lead sponsor impact

(Goldacre et al. 2018) gives evidence that the rate of results declaration is very impacted by the type of lead sponsor, commercial sponsors having a much higher declaration rate. We therefore propose to break down most of the analysis axis with the type of lead sponsor, being either academic or industrial. This categorization has been done manually based the lead sponsor name.

2.3 'Local' Open Science Monitors

The University of Lorraine was the first institution to propose a local version of the French Monitor. The code created on this occasion is freely accessible:(Bracco 2020).

This local version, published during spring 2020, was designed with reusability in mind. For this purpose, the code has been detailed step by step in Jupyter Notebooks and includes a readme file explaining all the required actions to obtain its own Barometer.

The availability of the code was combined with numerous training sessions as well as individual assistance provided by the University of Lorraine to each institution that requested it.

Following the publication of this code, many institutions were able to generate their own Open Science indicators. This enthusiasm for local implementation has underlined the need for institutions to have reliable and effective tools for monitoring Open Science.

The new version of the national Monitor allows, directly from the website, to generate graphs from a list of DOIs previously sent to the MESRI team. The University of Lorraine has been asked to test and implement this new version.

The constitution of the DOI corpus remains an essential step for the institutions. The code proposed by the University of Lorraine makes it possible to generate this list simply by crossing various databases such as the Web of Science, PubMed or HAL.

This simplified version will probably encourage other institutions to establish their own Monitor.

2.4 Data collection system and architecture

In this section, we will try to present the global workflow to collect, enrich and consolidate the data as described before with the technical and the storage challenges.

2.4.1 Data manipulation

Collect, Select, Enrich and Save We collect data from multiple sources (PubMed via Medline, Crossref, our own list of DOIS), and then try to guess the country according to the affiliations. And from the DOIs, we collect more details about that publication via Unpaywall. By details, we mean open access, DOAJ, APC … from multiple sources.

Each step consumes time and CPU. Assuming any step can fail at any time, we choose to develop each step as independent and idempotent.

From PubMed, we collect all the database via Medline and store it as JSONL files on Object Storage on OVH Cloud in a dedicated container. At that point,

we have all the notices of medical publications. We find there the affiliations of each publication. With the affiliation we tried to detect the countries of the institutions mentioned in the affiliations, in order to filter on French publications. The selected publications are stored as JSONL files in another dedicated container on Object Storage on OVH Cloud. Now focusing on the French publications, we use the extracted notices to match them against a MongoDatabase that we built on a dump of Unpaywall. We use the DOI to consolidate the data and then add many details.

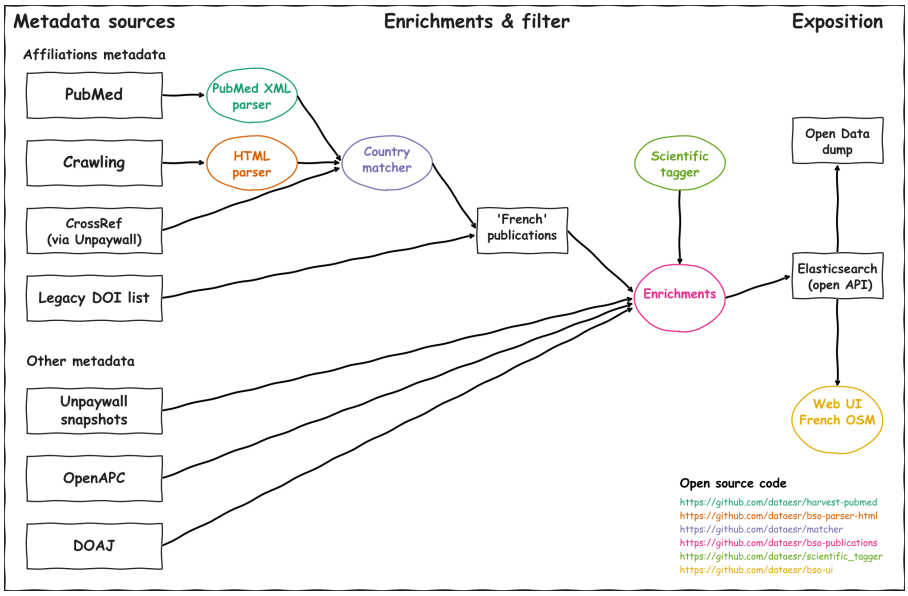

Figure 5: Global overview of the publications data flows

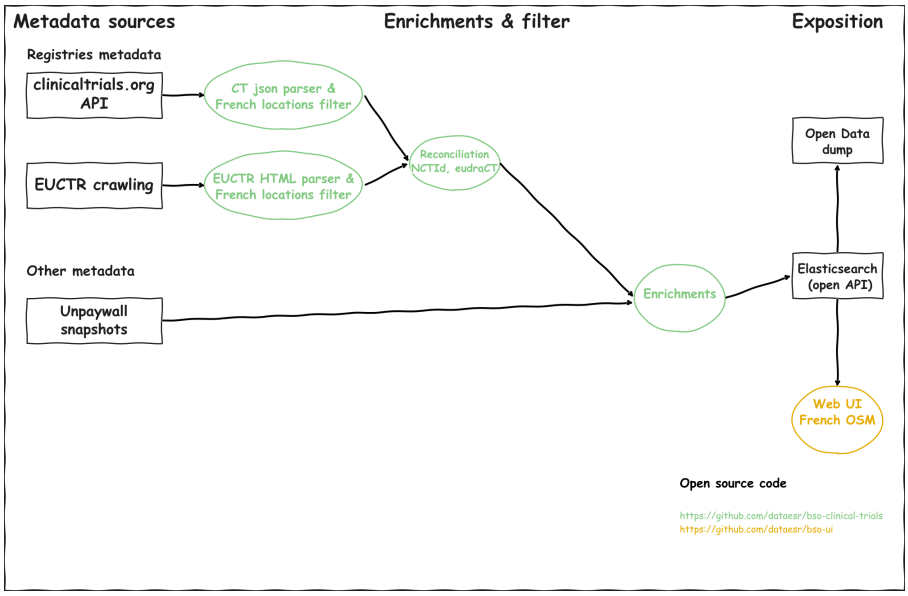

Figure 6: Global overview of the trials and studies data flows

3. Results

4. Discussion and conclusion

4.1 Findings

4.2 Limitations and future research

4.2.1 Limitations

DOI only

mixes preprint servers / open repo

richer metadata

based only on metadata from registries

4.2.2 Future work

research data, software code

Software and code availability

https://github.com/dataesr/bso-publications

https://github.com/dataesr/bso-clinical-trials

Data availability

portail MESRI

Acknowledgements

First, we want to thank Florian Naudet (https://orcid.org/0000-0003-3760-3801) from University of Rennes 1, Rennes, France, who helped us a lot to analyse the issues related to the clinical trials data, as well as Nicholas DeVito (https://orcid.org/0000-0001-8286-1995). We also want to thank the agency WeDoData (https://wedodata.fr/) that helped us designing the new web interface for the French Open Science Monitor.

References

Bracco, Laetitia. 2020. "Baromètre Lorrain de La Science Ouverte." Université de Lorraine. https://hal.univ-lorraine.fr/hal-03450104.

COSO, French Open Science Committee. 2018. "Feedback on EC Open Science Monitor Methodological Note." https://www.ouvrirlascience.fr/feedback-ec-science-monitor/.

Goldacre, Ben, Nicholas J DeVito, Carl Heneghan, Francis Irving, Seb Bacon, Jessica Fleminger, and Helen Curtis. 2018. "Compliance with Requirement to Report Results on the EU Clinical Trials Register: Cohort Study and Web Resource." *BMJ*, September, k3218. https://doi.org/10.1136/bmj.k3218.

Jeangirard, Eric. 2019. "Monitoring Open Access at a National Level: French Case Study." In *ELPUB 2019 23d International Conference on Electronic Publishing*. OpenEdition Press. https://doi.org/10.4000/proceedings.elpub.2019.20.

———. 2021. "Content-Based Subject Classification at Article Level in Biomedical Context." *arXiv:2104.14800 [Cs]*, May. http://arxiv.org/abs/2104.14800.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. "Bag of Tricks for Efficient Text Classification." *arXiv:1607.01759 [Cs]*, August. http://arxiv.org/abs/1607.01759.

L'Hôte, Anne, and Eric Jeangirard. 2021. "Using Elasticsearch for Entity Recognition in Affiliation Disambiguation." *arXiv:2110.01958 [Cs]*, October. http://arxiv.org/abs/2110.01958.

MESRI. 2018. "National Plan for Open Science." https://cache.media.enseignementsup-recherche.gouv.fr/file/Recherche/50/1/SO_A4_2018_EN_01_leger_982501.pdf.

———. 2021. "2nd National Plan for Open Science." https://cache.media.enseignementsup-recherche.gouv.fr/file/science_ouverte/20/9/MEN_brochure_PNSO_web_1415209.pdf.

Piwowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. 2018. "The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles." *PeerJ* 6 (February): e4375. https://doi.org/10.7717/peerj.4375.