

# Open Science for Social Science and Humanities: Open Access availability and distribution across disciplines and Countries in OpenCitations Meta

Seyedali Ghasempouri<sup>1</sup>[0000–0003–3446–1115], Maddalena Ghiotto<sup>1</sup>[0009–0009–1309–6340], and  
Sebastiano Giacomini<sup>1</sup>[0009–0007–7813–0939]

Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy

## Abstract.

**Purpose:** This study aims to investigate the representation and distribution of Social Science and Humanities (SSH) journals within the OpenCitations Meta database, with a particular emphasis on their Open Access (OA) status, as well as their spread across different disciplines and countries. The underlying premise is that open infrastructures play a pivotal role in promoting transparency, reproducibility, and trust in scientific research.

**Approach:** The research methodology involved the use of secondary data sources, namely the OpenCitations Meta database, the ERIH-PLUS bibliographic index, and the DOAJ index. A custom research software was developed in Python to facilitate the processing and analysis of the data.

**Findings:** The results reveal that 78.1% of SSH journals listed in the European Reference Index for the Humanities (ERIH-PLUS) are included in the OpenCitations Meta database. The discipline of Psychology has the highest number of publications. The United States and the United Kingdom are the leading contributors in terms of the number of publications. However, the study also uncovers that only 38% of the SSH journals in the OpenCitations Meta database are OA.

**Originality:** This research adds to the existing body of knowledge by providing insights into the representation of SSH in open bibliographic databases and the role of open access in this domain.

**Limitations:** While this study provides valuable insights into the representation of SSH journals in the OpenCitations Meta database, it is limited to the data available on DOAJ. Furthermore, the study does not delve into the specific reasons behind the observed disparities in disciplinary representation or the low percentage of OA journals. It also does not explore the comparison between countries in terms of the kinds of articles.

**Keywords:** Open Science · Social Science · Humanities · Open Access · OpenCitations Meta · ERIH-PLUS · DOAJ · Bibliographic databases

## 1 Introduction

Social Science and Humanities (SSH) are known to be among the most difficult domains of application for bibliometric studies, a field concerned with the application of quantitative methods to measure scientific publications and their impact in the field of scholarly research. Major research in the domain includes the understanding of scientific citations and the investigation of publication pattern, as well as the use of such measurements in policy and management contexts. Bibliometrics application in the SSH seems to be hindered by their underrepresentation, as bibliometric studies rely in large part in the databases' representativeness of the scientific activity studied.

Throughout the years, the comparative study of databases' disciplines has focused on Web of Science and Scopus, two authoritative but commercial sources, demonstrating in both older and more recent studies[1] that "the use of either WoS or Scopus for research evaluation may introduce biases that favor Natural Sciences and Engineering as well as Biomedical Research to the detriment of Social

Sciences and Arts & Humanities”[2]. Nonetheless, the coverage of SSH Journals seems to be lower compared to two other disciplines even when involving other sources[3].

The problem seems to arise from the different publication patterns in SSH publications with respect to other domains. For example, ”significant parts of the scholarly production in the SSH are published in national journals, book chapters, and monographs. As a result of this diversity, the challenge of setting criteria for the selection of source items is seen as much greater than for the sciences”[4].

Only in recent years, mainly thanks to Open Science initiatives and the shift towards Open Access publishing, new infrastructures have emerged or opened their data, contributing to the advancement of reliable and transparent scientific studies. A significant progress for SSH has been the creation of the European Reference Index for the Humanities (ERIH PLUS)[5], ”a deliberate attempt to go beyond the commercial indexing services such as Web of Science and Scopus by covering more comprehensively all peer-reviewed scholarly journals in the SSH that are publishing at a minimum national level”[6].

Another turning point for open bibliometric studies has been the Initiative for Open Citations (I4OC)[7], which promotes the unrestricted availability of scholarly citation data as a crucial requirement for the bibliometrics and scientometrics domain. Thanks to this initiative, many publishers have decided to open their deposited references at Crossref[8], an association that provides persistent identifiers assigned to academic publications and publishes metadata related to these publications. In their recently published comparative study, Borrego et al. assessed the ERIH-PLUS coverage of Crossref and identified it as the most promising resource for bibliographic discovery in the Arts and Humanities[9]. They also pointed out the need for improvement of its metadata completeness.

One of the founders of I4OC is OpenCitations[10], a not-for-profit infrastructure organization for open scholarship dedicated to the publication of open bibliographic and citation data using Semantic Web technologies. OpenCitations aligns fully with open science guidelines and complies with FAIR principles[11]. Currently, OpenCitations retrieves and organizes citation data from various sources, including all Crossref open references. However, as of the present day, no study on the coverage of this new resource for SSH journals has been conducted.

Our research aims to address the lack of representation of OpenCitations in the panorama of comparative studies on bibliographic indexes, specifically regarding the coverage of SSH journals. At the same time, our research aims to conduct further evaluation by observing the distribution of journals across disciplines and countries, as well as assessing their Open Access status. To do so, we will consider OpenCitations Meta[12] the database that stores all bibliographic metadata for all publications involved in the OpenCitations indexes and compare it against ERIH-PLUS as it is the most comprehensive SSH index at the present day, with 11129 Journals listed.

Our research questions are the following:

- **RQ1:** What is the coverage of publications in Social Science and Humanities (SSH) journals (according to ERIH-PLUS) included in OpenCitations Meta?
- **RQ2:** What are the disciplines that have more publications?
- **RQ3:** What are countries providing the largest number of publications and journals?
- **RQ4:** How many of the SSH journals are available in Open Access according to the data in DOAJ?

The study has its foundation in the belief that open infrastructures are fundamental to ensure transparency, reproducibility, and trust in scientific research, having an essential value for the bibliometrics domain. The present study can contribute to fostering a culture of open science and calling for more incentives and adequate policies where needed.

Moreover, in agreement with Tennant et al.[13], this study considers OA as of the challenges faced by the “open science” transformation since it has been proven to lead to significant academic and societal advancements, enabling unprecedented data access for scientific studies to nonacademic scientists, and to machines for data mining, for instance.

Nevertheless, open Access publishing is still damaged by questionable or inaccessible practices, such as Article Processing Charges and Predatory Journals. Severin et al.[14] research on discipline-specific OA publishing practices found the OA uptake in SSH to be lower than in most other fields. Among the reasons behind this trend, one can find lower levels of general OA awareness among scholars, the notion

of prestige related to the choice of venue publications and misunderstandings related to licensing and plagiarism. Therefore, the need for the promotion of Open Access good practices and the dissemination of reliable OA Journals is of extreme importance today.

In the context of our research, to ensure the reliability of the data about OA status, we will refer to the Directory of Open Access Journals DOAJ[15], since it is a trusted source with 19,366 indexed journals at the present day.

The rest of the paper is structured as follows. In Section 2, we present the Materials and Methods used in our research. In Section 3, we discuss the Data used in our study. In Section 4, we outline our Methodology. In Section 5, we present our Results, in Section 6, we provide Discussions including limitations, in Section 7 the Conclusions and further research are provided.

## 2 Materials and Methods

The adoption of open science good practices to transparently manage every research outcome, from managing data and developing a methodology to the publication of results, is a central aspect of the present research:

The Data Management Plan (DMP)[16] that was published at the beginning of our work ensured transparency about the handling of the data produced. It describes in detail the choices made to comply with FAIR data principles, making our data findable, accessible, interoperable, and reusable. The DMP includes two datasets: 1) a research software[17], and 2) a data catalog with the output files of the software where our results are stored[18].

The first version of the DMP has undergone peer review, and a second version has been published with necessary modifications. Finally, a third version with minor revisions reflecting subsequent changes made during our work process has been published at the end of the research.

The same practice has been adopted to develop the study's methodology, which is a core research outcome essential for enabling reproducibility. Our methodology will be described in detail below and is also available as a workflow on protocols.io[19]. As mentioned, the published workflow has also been modified during the research, incorporating peer reviews and in-process changes that we deemed necessary. This has resulted in a second and a third, final version of the workflow.

## 3 Used Data

The research used secondary data retrieved from the following sources: the OpenCitations Meta database, the ERIH-PLUS bibliographic index, and the DOAJ index.

The OpenCitations Meta (OC Meta) dataset was downloaded as a dump on 24/02/2023[20]. We opted for the dump download because accessing the data through a SPARQL endpoint or REST API would have taken too much time due to the large quantity of data we needed. The dataset represents the largest source in this research, consisting of a zipped folder (8GB) containing over 22,000 CSV files (36GB unzipped), covering more than 90 million bibliographic entities.

The ERIH PLUS (EP) dataset "Approved Journals" contains metadata about all SSH Journals included in the Index and is publicly available for download in CSV format as of 27/04/2023 (2MB). Since they continuously update their data and older version could not be found, the dataset have been published on zenodo[21]. The metadata for each Journal entry includes a unique ERIH PLUS identifier, Print-ISSN and Online ISSN, Original and International title, Country of publication, ERIH PLUS disciplines classification, and the OECD classification. For the purpose of this study's disciplinary classification, the ERIH PLUS classification was adopted.

The DOAJ dataset was also downloaded in its publicly available dump in CSV format on 28/05/2023 (22.9MB). Since they continuously update their data and older version could not be found, the dataset have been published on zenodo[22]. According to the guidelines[23], the DOAJ includes only OA journals whose copyright holders grant usage rights to others using an open license. Additionally, the full text of all content must be available for free and open access without delay. The main metadata

needed from the DOAJ dataset was the Journal ISSN and EISSN, as well as the Country of Publisher, to integrate with ERIH PLUS countries.

## 4 Methodology

To answer research questions, a purpose-tailored research software in Python programming language was created. The main library used to explore, manipulate and combine available data into meaningful new datasets is Pandas[24]. Other libraries used include glob, os, tqdm[25], concurrent.futures[26], argparse, csv. As a first step, a dataset named *SSH\_Publications\_in\_OC\_Meta\_and\_Open\_Access\_status* is created.

as the result of the parsing and merging of the three datasets. This dataset is intended to indicate how many SSH journals listed in EP are included in the OC Meta database, along with the number of publications (OC Meta) and Open Access (DOAJ) status for each journal. It is created as the output of the method *process\_files()* called on an object of the class *PlayaristsProcessor*, initialized with the path to the three considered datasets, a batch size and the number of CPU workers to use for parallel processing. These last two variable are required given the dimensions of the unzipped OpenCitations Meta dump, which is divided in batches and processed in parallel for enabling an efficient parsing with a reasonable execution time. In detail, the concurrent.futures[26] module is used to provide a high-level interface for asynchronously executing callable. In this step, a *ProcessPoolExecutor* object is created and a function that performs the filtering of OpenCitations Meta dump according to the journals present in EP is mapped to the object. The ancillary function that performs this step is *process\_meta\_csv()*. First, it creates a dictionary out of the EP data with unique Print ISSN and Online ISSN as keys and EP Journal ID as value, then it retrieves the issn identifiers in OC Meta venue column with a string matching technique and it proceeds to verify whether these identifiers are present in the key values of the EP dictionary. The positive matches are stored in a new *DataFrame* containing the *uniqueOC\_omid* identifier, a list of the issn of each journal, the unique EP identifier and the number of Publications present in OC Meta for each journal. Finally, the Open Access status is added by means of the *process\_doaj\_file()* ancillary function. This function takes in input the DOAJ dataset as a *DataFrame* and the newly created *DataFrame* with the SSH journals in OC Meta; it verifies whether the identifiers in the issn column of this last *DataFrame* are present in the DOAJ values of Journal ISSN and Journal EISSN columns and extend the result *DataFrame* by adding the Open Access column. This column's value is set as "True" if a positive match is found and as "Unknown" if not; since there might be Open Access Journals not enlisted in DOAJ.

This first *DataFrame* provides the data needed to answer questions 1 and 4 of this research about the OC Meta coverage of EP journals and their OA status and is exported in csv format with the name *OCMeta\_DOAJ\_ErihPlus\_merged.csv*. As a second step, to answer research questions 2 and 3, two other datasets are created: SSH Publications by Discipline and SSH Publications and Journals by Country. Both of these datasets have three columns, the first, respectively, for disciplines and country classification, the second holding journal counts value and the third holding publication counts value. The count of journals per discipline is not comprehended in our research question, nonetheless, since the process was similar to the one adopted to retrieve the count of journal per country, it was considered worth adding as an additional result. It is worth mentioning that EP journals specify more than one discipline. Given the lack of access to more granular data about single publications disciplinary classification, our approach was to consider all publications under to the same journal as belonging to all disciplines specified for that journal. To retrieve information about disciplines a dictionary is created by the method *create\_disciplines\_dict()* of the class *DisciplinesProcessor*. This method merges *SSH\_Publications\_in\_OC\_Meta\_and\_Open\_Access\_statuswithEP* dataset to include disciplines information, then it creates the dictionary by setting each unique discipline in the EP dataset as key and a list of EP unique identifiers of the journals belonging to each discipline as value. The process unfolds approximately in the same way to retrieve information about countries, but in this case it's the *create\_countries\_dict()* of the class *CountriesProcessor* to be called. This method differs from

the one specified above because it needs to parse the DOAJ dataset to retrieve Country of publisher values whenever countries information are missing in ERIH PLUS. To do so, it relies on the ancillary function *retrieve\_doj\_countries()* that adds missing countries to the dictionary. This function filters *SSH\_Publications\_in\_OC\_Meta\_and\_Open\_Access\_status* to create a sub DataFrame with only rows of identifiers that lack country information, then merges it twice with DOAJ, once for Journal ISSN column and once for Journal EISSN, using the “how=left” parameter. Then, it adds the countries in DOAJ to the missing values of the Country column. While performing this step, the function also saves a list of the issn that still remained without country information, for transparency and data completeness. Lastly, it iterates over the updated DataFrame and either adds new journals identifiers to the respective country key in the dictionary, if it already exists, or creates a new country key and related journal identifiers as value. Finally, to compute the counts of publications and journals for both disciplines and countries, the method *counts()* of the *CountsProcessor* class is run. This method takes in input either the country or the discipline dictionary created in the previous step and the label to set as the first column value in the final dataset. 29 It iterates over the dictionary keys to filter *SSH\_Publications\_in\_OC\_Meta\_and\_Open\_Access\_status* according to journals in the list specified as value, then it stores the length of the filtered DataFrame as the count of the journals. Lastly, it sums all the values in the column *Publications\_in\_venue* to calculate the count of publications. The final DataFrame is exported as csv file according to the export path defined as *CountsProcessor* attribute.

As far as software testing is concerned, given the complexity and size of the datasets, synthetic data have been created to test the software functioning. Creating diverse data occurrences and possible errors or odd instances to evaluate how the software would react in such situations. Here are the designed requirements: both OCMeta and DOAJ contain publications or articles not included in ERIH PLUS; OCMeta dumps are divided into multiple files: the program must correctly concatenate data from different files; OCMeta files contain different types of entries, not just publications (for example, venues). Also, multiple publications are associated with the same venue. OCMeta publications have a variable number of venue identifiers, which are not associated with a precise number of columns. We designed three test cases: 1) to test the merging of OC Meta and ERIH-PLUS, 2) to test the merging of the *OCMeta\_DOAJ\_ErihPlus\_merged*, resulting from the above step with DOAJ dataset, 3) to test the correct assignment of countries and disciplines to different journals. Considering the construction of our software pipeline we can assume that if the software’s success in these tests can be a solid proof of its functioning to fulfil its requirements, minor errors could still be present in the handling of data but they shouldn’t be compromising for the overall quality of our research. As a final step in the workflow, suitable visualization to make the results’ analysis more accessible were produced. These were generated using *matplotlib*[27] and *seaborn*[28].

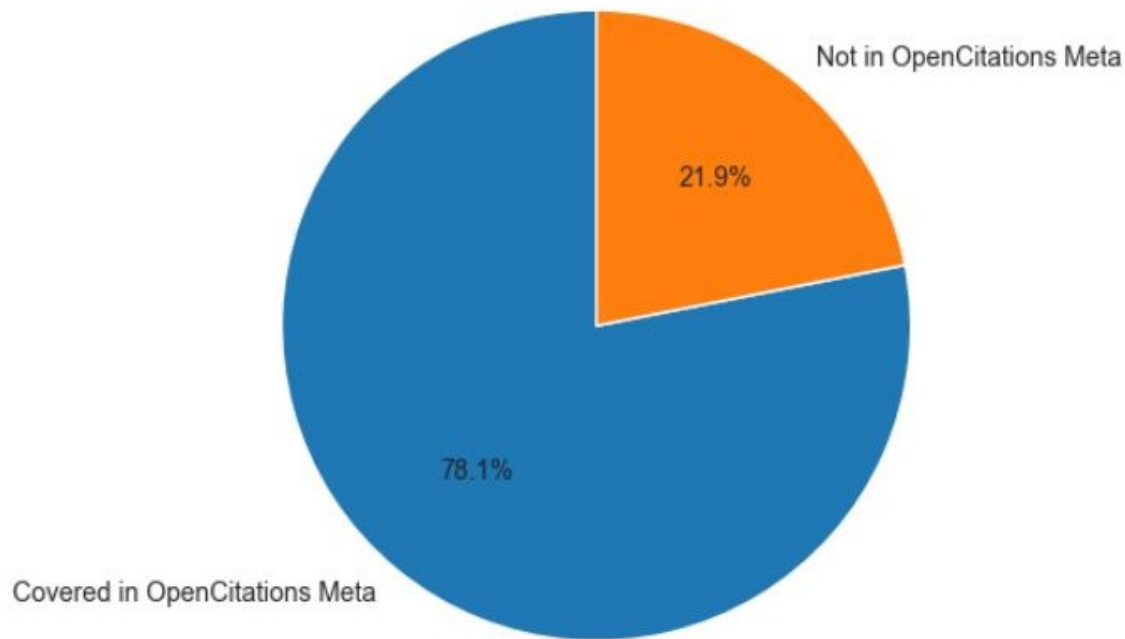
## 5 Results

Here we present the most significant results of the research, along with some other supporting results.

### 5.1 What is the coverage of publications in Social Science and Humanities (SSH) journals (according to ERIH-PLUS) included in OpenCitations Meta?

OpenCitations Meta(Fig.1) cover the 78.1% of ERIH PLUS index, for a total of 8691 journals, and 5.496.449 publications.

Comprehensively, the SSH Journals in OpenCitations Meta represent the 17% of the whole database. This result shows that, on the one hand, OC Meta covers a significant proportion of the ERIH PLUS journals. According to the most recent study (2023), this percentage is almost comparable to Cross-refs’ coverage[9] 80% and significantly higher than Scopus’ coverage 49%, affirming OpenCitations’ promising position as a bibliographic database for SSH. Nevertheless, it is worth noting that a higher coverage of journals does not necessarily imply a larger coverage of publications, since the quantity of articles published by a journal may differ significantly, as well as the completeness on the coverage



**Fig. 1.** Pie Chart. ERIH Plus Journals in OC Meta Coverage

of publications for each journal may still vary among different databases. As it has been found in our research, the counts of publications per journal in OpenCitations Meta SSH journals ranges from 1 to 257538 (Plos ONE Online Journal, issn:1932-6203).

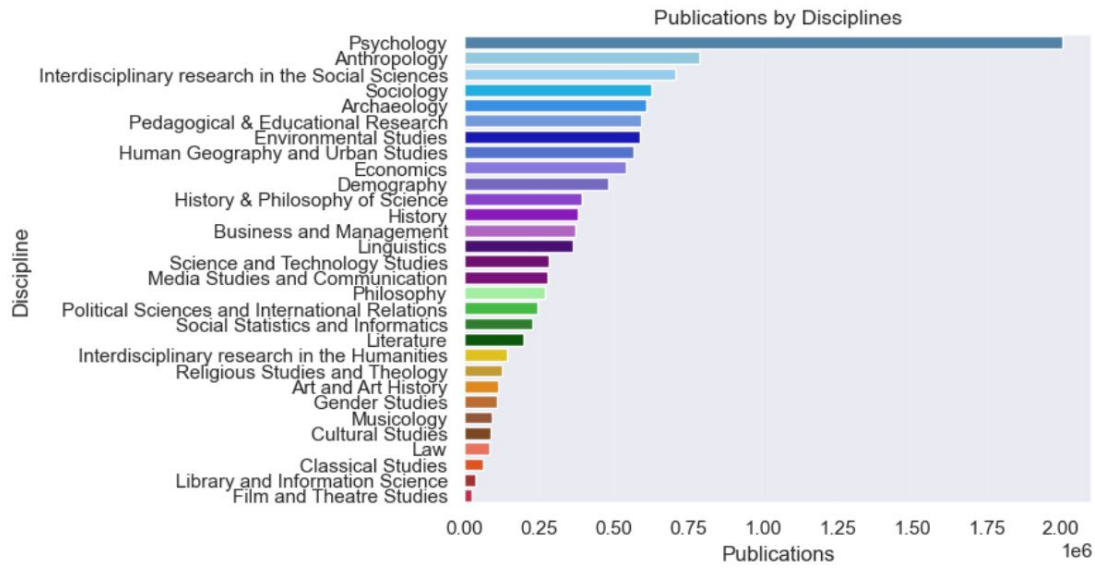
## 5.2 What are the disciplines that have more publications?

Disciplines have been ranked according to their publications number in descending order (Fig.2). A wide gap between Psychology, the discipline with the highest number of publications and the other disciplines can be observed. Although ERIH PLUS doesn't specify which disciplines belong to the Social Sciences and which belong to the Humanities, referring to a reasoned classification provided by Spinaci et al.[29] it is possible to identify a trend that shows Social Science as covering more publications than Humanities. In fact, among the top 10 disciplines, only Archaeology belongs to the Humanities domain, whereas all the fields placing lower in the rank belong to the Humanities.

## 5.3 What are countries providing the largest number of publications and journals?

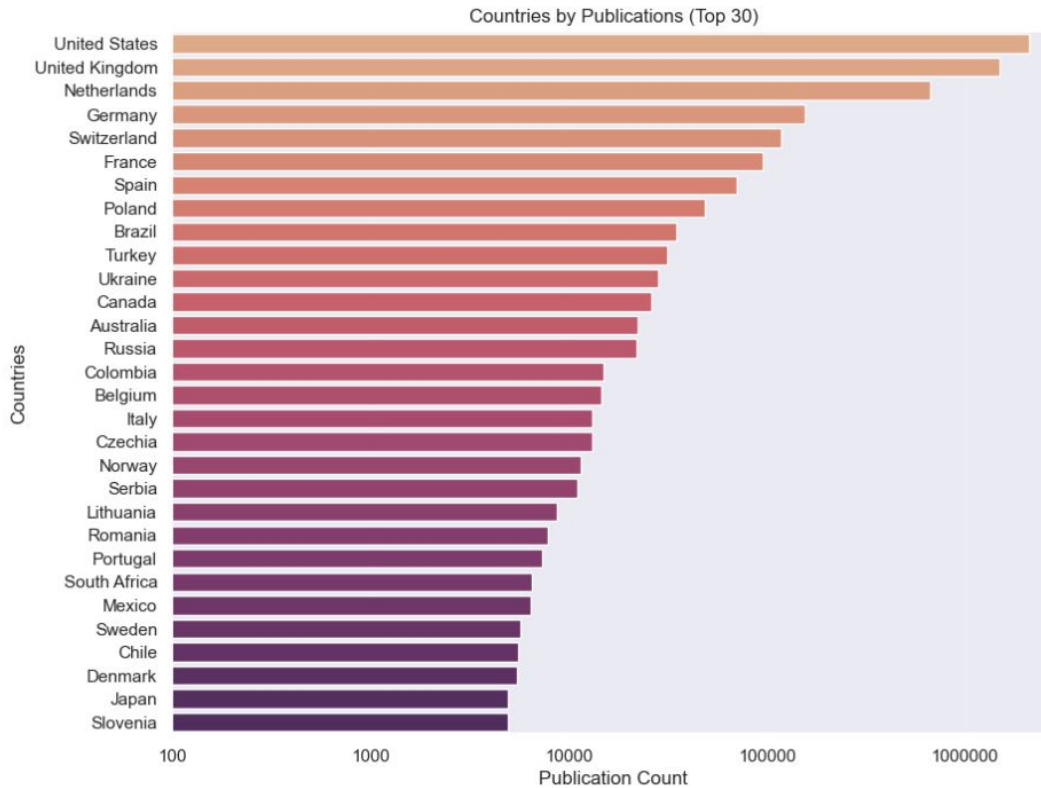
The results showed the coverage of 92 countries in for SSH publications and journals in OpenCitations Meta. The following bar plot (Fig.3) is based on a logarithmic scale to squeeze the gap between the first Country of the rank (US) and the other values. This means that, although the US and UK seem to have a similar number of contributions, the difference is remarkable: 2,096,630 publications from the US against 1,501,629 (i.e.: -28.31). Nevertheless, if we consider the geographical distribution of the data, the value for a small European country can't be possibly compared with the value for wide territorial extensions. such as the USA or South Africa. For this reason, while doing the same research on Scopus and Crossref, Borrego et al.[9] have aggregated European countries into East, North, South and West Europe, highlighting the predominance of European countries' publication in the database.

This is not surprising, given that, as ERIH PLUS states: "the main target group of the index are researchers and research within a European framework. To the extend which the index holds journals from other parts of the world, it is because they are assumed to add value to the ERIH PLUS main target group and scope"[5].

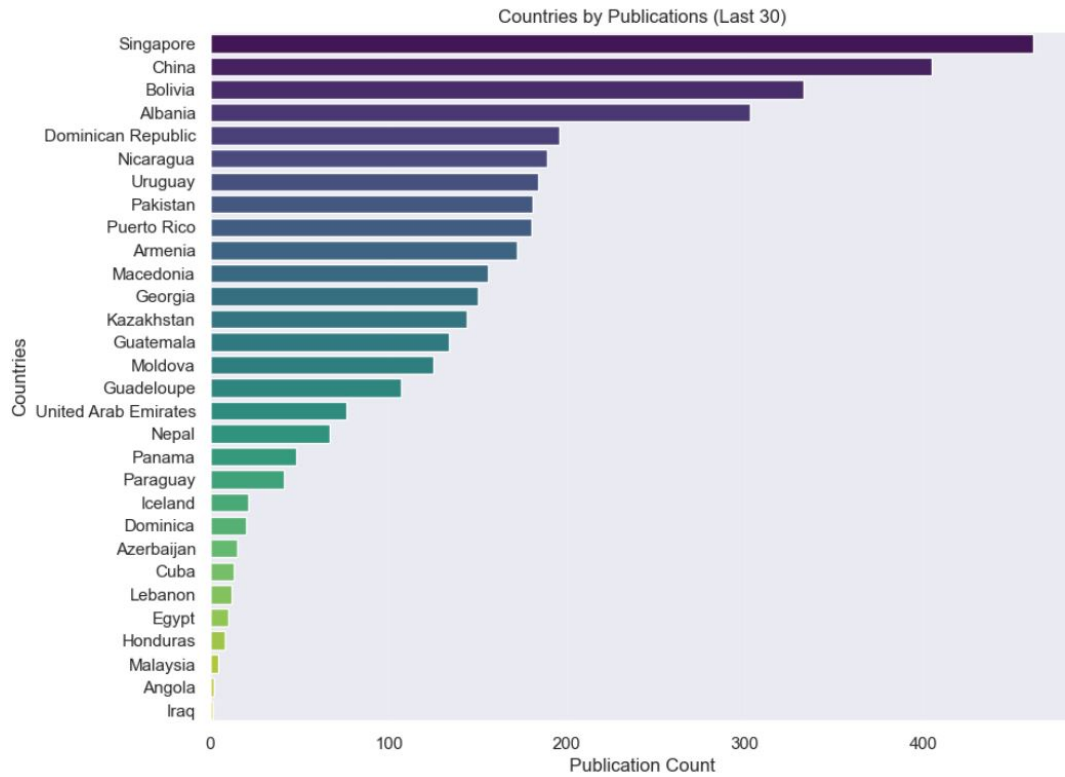


**Fig. 2.** Bar plot. Disciplines ranked according to publications number

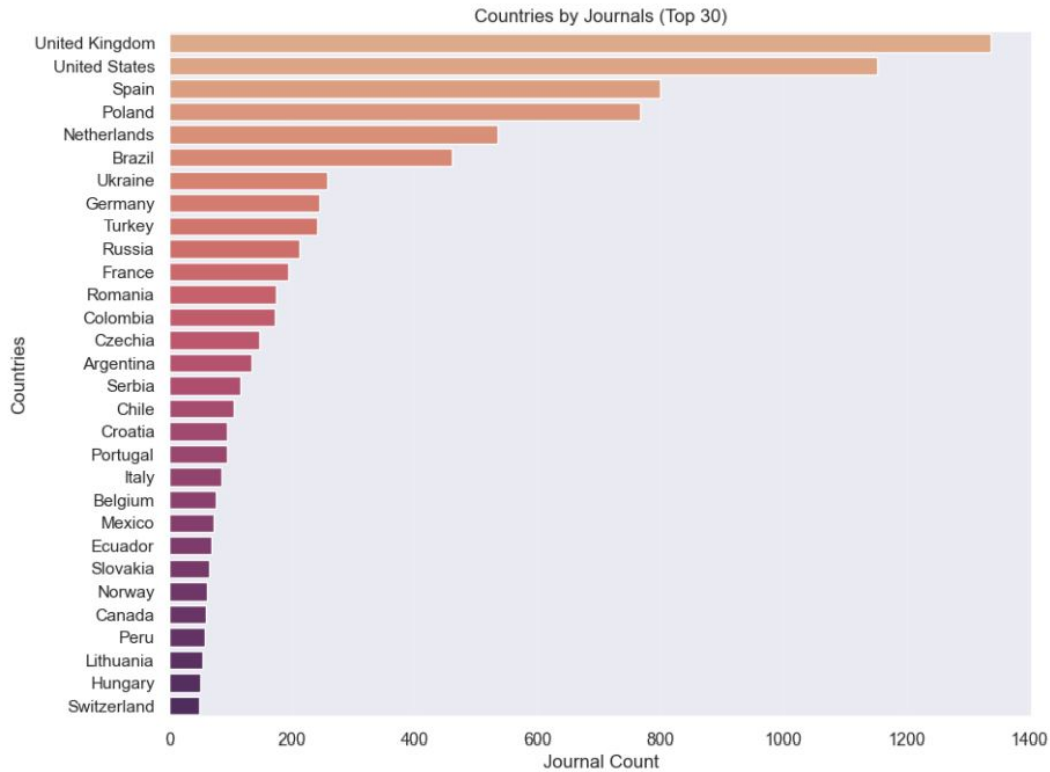
**Publications:** Accordingly, in Fig.4 it is possible to see what are the countries that have the least number of publications and almost all of these are non-European countries.



**Fig. 3.** Bar plot. The gap between countries.

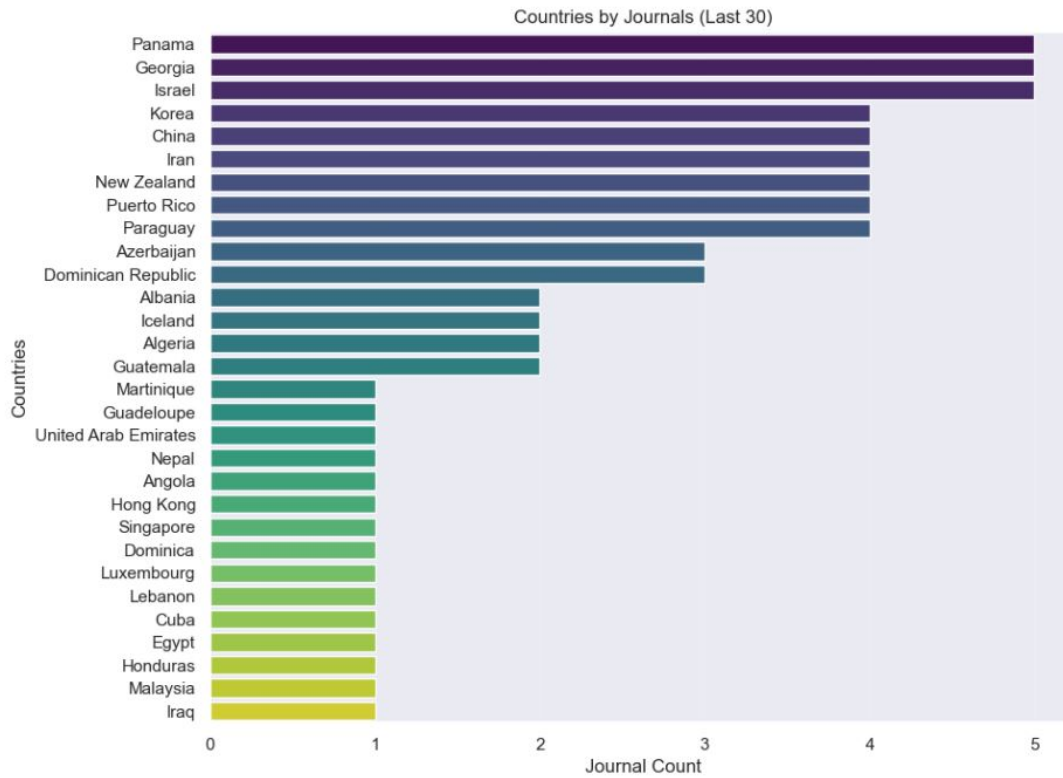


**Fig. 4.** Bar plot. The least number of publications are almost non-European countries.



**Fig. 5.** Bar plot. Distribution of journals over countries

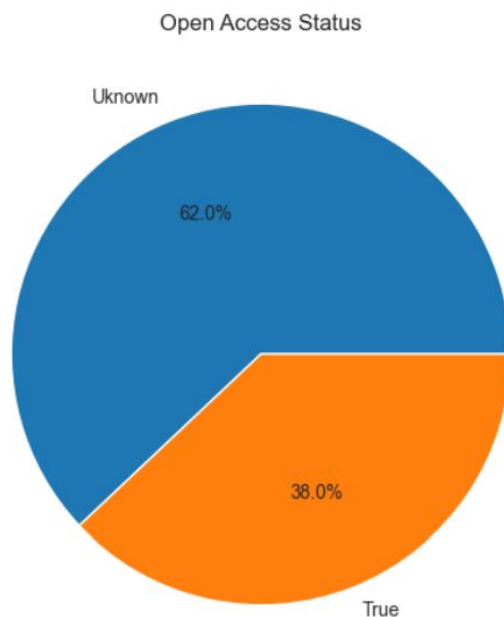




**Fig. 6.** Bar plot. Distribution of journals over countries

**Journals:** Highlighting the distribution of journals over countries is especially relevant in the present study, considering the factor mentioned above of the variation of publications per journal. In Fig.5, it is particularly remarkable to observe the different position held by Australia and Switzerland. The first is the 13th country for number of publications with 24225 publications but doesn't even place in the top 30 if considering journals, as it publishes only 38 Journals; the second is the 5 th for number of publications with 127230 articles, but is at the 30th position considering journals, having only 49. A similar comparison can be profoundly interesting for evaluating different countries publication practices. As we have seen, scientific research in the SSH is carried out on a national level more than in other domain, for this reason these data could be very useful for further comparative research on national publication practices in the SSH.

#### 5.4 How many of the SSH journals are available in Open Access according to the data in DOAJ?



**Fig. 7.** Pie chart.

Only 38% of the SSH journals (Fig. 7) covered by OpenCitations Meta, according to DOAJ criteria, are OA for a total of 3299 journals and 743559 publications. This percentage aligns with many of the studies compared by Severin et al. in 2020[14] that found SSH is among the slowest domains in the uptake of OA practices.

## 6 Discussion

To summarize the present research result, a satisfactory coverage of SSH journals in OpenCitations has been observed. Although there is space for improvement, the resource proves to be a good base for bibliometric studies on SSH at a European level, especially in comparison with other bibliographic indexes and considering the consistent and comprehensive provision of metadata and citations data that OpenCitations holds. Nonetheless, it is difficult to generalize these results on a global scale.

This is mainly due to the Eurocentric approach of the ERIH PLUS index considered for this study, to do so it would be advisable and interesting to reproduce the same study with more balanced indexes, even though at the present day qualitative, public resources of this kind seem to be lacking. Furthermore, there was a limited presence of OA journals in the field. This result comes to support the movement for the promotion of open accessibility of SSH resources.

### 6.1 Limitations

Finally, on the grounds established by our first research questions, our results on countries and disciplines distribution can be considered a solid representation of these distributions, offering interesting results for bibliometrics studies. These results could be enriched by investigating, for instance, what is the distribution of disciplines in each country. They also provide grounds for further investigation into the factors contributing to disparity across disciplinary representation to help in understanding and addressing them, as well as long-term evaluation of the changes in the SSH domain.

Overall, the highlighted strengths and limitation of the present study prove the importance as well as the need of having accessible open data for bibliometric studies, in order to be able to perform trustworthy, balanced and replicable studies on scholarly research. So far, the open science movement was able to advance the adoption of good practices for transparent and replicable scientific research, nonetheless the space for improvements remains fairly vast and it is of extreme relevance today to call for more awareness to be able to promote a more open research culture.

## 7 Conclusions

This study demonstrates that the representation and distribution of SSH journals within the OpenCitations Meta database. The findings reveal that a significant proportion of SSH journals listed in the ERIH-PLUS are included in OpenCitations Meta, with Psychology emerging as the discipline with the highest number of publications. However, the study also uncovers that a considerable percentage of the SSH journals in the OpenCitations Meta database are not open access.

To further enhance the usefulness of this study and provide a comprehensive overview of the issue it would be relevant to deepen the knowledge of OA availability across different SSH disciplines and countries. Future research could delve deeper into the impact of OA on various aspects of citation patterns. This could involve studying whether OA articles receive more citations than those behind paywalls, analysing the citation lifecycle of OA versus non-OA articles to ascertain if OA prolongs the relevancy and citation of a paper, and examining citation networks to understand if OA journals tend to cluster together or bridge gaps between different research communities.

Such inquiries could provide valuable insights into the broader implications of the OA movement on the visibility, influence, and interconnectedness of SSH research, resulting extremely useful to convince different actors involved in scholarly publication in the SSH of the validity of OA publishing. A similar thorough analysis could significantly help to drive political decisions on resources allocation to efficiently support the shift towards accessible scholarly research.

## References

1. Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. In *Proceedings of the science and technology indicators conference* (pp. 598-604).
2. Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, 213-228.
3. Visser, M., Van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20-41.
4. Sivertsen, G., & Larsen, B. (2012). Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: An empirical analysis of the potential. *Scientometrics*, 91(2), 567-575.
5. Norwegian Centre for Research Data. (27, April 23). *ERIH PLUS*. Retrieved Month Day, Year, from <https://kanalregister.hkdir.no/publiseringsskanaler/erihplus/>
6. Lavik, G. A. V., & Sivertsen, G. (2017). ERIH PLUS—Making the SSH visible, searchable and available. *Procedia Computer Science*, 106, 61-65.
7. Initiative for Open Citations. (24, Feb 23). <https://i4oc.org/>
8. Crossref. (28, May 23). <https://www.crossref.org/>
9. Borrego, Á., Ardanuy, J., & Arguimbau, L. (2023). Crossref as a bibliographic discovery tool in the arts and humanities. *Quantitative Science Studies*, 4(1), 91-104.
10. Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428-444.
11. GO FAIR. (28, May 23). *FAIR Principles*. <https://www.go-fair.org/fair-principles/>
12. OpenCitations. (28, May 23). *Meta*. <https://opencitations.net/meta>
13. Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H.

- J. (2016). The academic, economic and societal impacts of Open Access: An evidence-based review (5:632). F1000Research.
14. Severin, A., Egger, M., Eve, M. P., & Hürlimann, D. (2020). Discipline-specific open access publishing practices and barriers to change: An evidence-based review (7:1925). F1000Research.
15. Directory of Open Access Journals. (28, May 23). *Home*. <https://doaj.org/>
16. Ghiotto, M., Ghasempouri, S., & Giacomini, S. (2023). *DataManagementPlan - Open Science for Social Science and Humanities: Open Access availability and distribution across disciplines and Countries in OpenCitations Meta (Version 4)*. Zenodo. <https://doi.org/10.5281/zenodo.8174644>
17. Ghasempouri, S., MaddaGh., & Giacomini, S. (2023). *Open-sci/2022-2023-playarists-code: Open Science for Social Science and Humanities: Open Access availability and distribution across disciplines and Countries in OpenCitations Meta - RESEARCH SOFTWARE (v1.0.0)*. Zenodo. <https://doi.org/10.5281/zenodo.7979757>
18. Ghasempouri, S., Ghiotto, M., & Giacomini, S. (2023). *Open Science for Social Science and Humanities: Open Access availability and distribution across disciplines and Countries in OpenCitations Meta - RESULTS DATASET (Version 0) [Data set]*. Zenodo. <https://doi.org/10.5281/zenodo.7979592>
19. Ghiotto, M., Giacomini, S., & Ghasempouri, S. (2023). *Research on ERIH PLUS approved SSH journals present in OpenCitations Meta database V.4*. <https://dx.doi.org/10.17504/protocols.io.5jyl8jo1rg2w/v4>
20. OpenCitations (2022). OpenCitations Meta CSV dataset of all bibliographic metadata. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.21747461.v3>
21. Ghasempouri, S., Ghiotto, M., & Giacomini, S. (2023). *ERIH PLUS (EP) dataset "Approved Journals" 27-04-2023*. <https://doi.org/10.5281/zenodo.8163337>
22. Ghasempouri, S., Ghiotto, M., & Giacomini, S. (2023). *The DOAJ dataset downloaded 28-05-2023 (journalcsv\_\_doaj-20230528\_0035\_utf8) [Data set]*. Zenodo. <https://doi.org/10.5281/zenodo.8163248>
23. Directory of Open Access Journals. (n.d.). Guide for applying for DOAJ. Retrieved Month Day, Year, from <https://doaj.org/apply/guide/>
24. PyPI. (28, May 23). Pandas. <https://pypi.org/project/pandas/>
25. PyPI. (28, May 23). Tqdm. <https://pypi.org/project/tqdm/>
26. Python.org. (28, May 23). concurrent.futures - Launching parallel tasks. <https://docs.python.org/3/library/concurrent.futures.html>
27. PyPI. (28, May 23). Matplotlib. <https://pypi.org/project/matplotlib/>
28. PyPI. (28, May 23). Seaborn. <https://pypi.org/project>
29. Spinaci, G., Colavizza, G., & Peroni, S. (2022). A map of Digital Humanities research across bibliographic data sources. *Digital Scholarship in the Humanities*, 37(4), 1254-1268.