# Uncovering the Citation Landscape: Exploring OpenCitations COCI, OpenCitations Meta, and ERIH-PLUS in Social Sciences and Humanities Journals

**Olga Pagnotta**[1] [0009-0004-5901-5730] , **Lorenzo Paolini**[2] [ 0009-0003-3803-4011]
**Marta Soricetti**[3] [0009-0008-1466-7742] , **Sara Vellone**[4] [ 0009-0008-4037-6607]

1 Digital Humanities and Digital Knowledge, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy olga.pagnotta@studio.unibo.it

2 Digital Humanities and Digital Knowledge, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy lorenzo.paolini6@studio.unibo.it

3 Digital Humanities and Digital Knowledge, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy marta.soricetti@studio.unibo.it

4 Digital Humanities and Digital Knowledge, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy sara.vellone@studio.unibo.it

## Abstract

The main purpose of this research is to answer to three different questions and find out:

1. the number of citations which refer to publications in Social Sciences and Humanities journals included in ERIH-PLUS, by looking at citations data contained in OpenCitations COCI and OpenCitations Meta;
2. the most citing and the most cited SSH discipline, according to the above mentioned datasets;
3. the citations coming from and going to publications contained in OpenCitations Meta which are not included in SSH journals.

We want to draw a line that connects these three different datasets, aiming at offering an overall view of the citations landscape of each of them. For this purpose, we approach the problem from a computational point of view. We extract only the relevant data by operating a first preprocessing of COCI, ERIH-PLUS and META's datasets. Then we build a python software able to analyze CSVs data, querying them to retrieve information needed and to present the results in a clear and understandable way. The findings show that the majority of citations come from and go to psychology publications, and a deep gap exists between the number of citations included in SSH journals and the number of citations that are not included in SSH journals. The research conducted by us has the purpose to add information to existing resources with the aim of facilitating their use and allowing the researchers to have a clearer view of the data contained in each dataset. In addition, the research has the purpose to gather information that may be

useful for understanding which is the most influential discipline in the SSH field and to provide a solid starting point for further studies regarding this subject.

**Keywords**: COCI, Meta, ERIH-PLUS, OpenCitation, SSH, Journals

---

# INTRODUCTION

Citations have been used in the past years as a criterion for evaluating the importance of a journal and, consequently, the importance of the articles in which they are contained and of their authors, too.
Claudio Castellano and Filippo Radicchi investigated the viability of the use of relative indicators for comparing article impact in different scientific disciplines[1], discovering the urge of validating the hypothesis of universality for all scientific disciplines and not only for a subset of them, due to the fact that the distribution of the number of citations received by an article is strongly depending on the scientific discipline[2].
A different path has instead been taken by Dassa et al.[3] regarding the idea of creating a comparative table of contents of the databases that list the journals in the SSH, which shows the broader coverage of the ERIH list for the human sciences.
The main purpose of this research is trying to compare and to relate different indexes to understand the number of citations that are included in SSH journals, and thus which are the disciplines that - according to these citations - cites and are cited the most.
The relevance of this research stands in the possibility to reuse the findings for further studies related to the disciplines predominant in the citation's field, and therefore to understand if there is any useful information on the importance of the disciplines themselves.

# MATERIALS AND METHODS

The starting points of our research are three different datasets: OpenCitations COCI, OpenCitations Meta and ERIH-PLUS.
COCI[4] is the OpenCitations Index of Crossref open DOI-to-DOI citations, which contains the details of all the citations that are specified by the DOI-identified works present in Crossref[5].
OpenCitations Meta[6] is a database that stores and delivers bibliographic metadata for all publications involved in the OpenCitations indexes.

---

[1] Castellano, C., Radicchi, F. On the fairness of using relative indicators for comparing citation performance in different disciplines. Arch. Immunol. Ther. Exp. 57, 85–90 (2009). https://doi.org/10.1007/s00005-009-0014-0
[2] Ibidem
[3] Michèle Dassa, Christine Kosmopoulos et Denise Pumain, « JournalBase - A Comparative International Study of Scientific Journal Databases in the Social Sciences and the Humanities (SSH) », Cybergeo: European Journal of Geography [En ligne], Science et Toile, document 484, mis en ligne le 08 janvier 2010, consulté le 27 mai 2023. URL : http://journals.openedition.org/cybergeo/22862 ; DOI : https://doi.org/10.4000/cybergeo.22862
[4] Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics, 121(2), 1213–1228. https://doi.org/10.1007/s11192-019-03217-6
[5] https://www.crossref.org/
[6] http://opencitations.net/meta

ERIH-PLUS[7] is an academic journal index for the SSH society in Europe. It includes the original ERIH lists, which initially covered only the humanities disciplines, while now it has been extended to also the social science ones[8].
To make a connection between these dataset, we have analyzed which type of information they have in common and which information were relevant for our research.
COCI's columns named *citing* and *cited* have a correspondence with the Meta's column *id,* and they all represent the DOIs[9]. The Meta's column *venue* matches with ERIH-PLUS' *Print ISSN* and *Online ISSN*. The *ERIH PLUS disciplines*' column is also to be taken into account for our purposes.

**Pre-processing classes and methods**

We decided to reuse some methods of the OpenCitations Preprocess Software[10] for Meta and ERIH-PLUS, adapted to our needs, to read, filter and clean the data and to store them in a new output file.
The class *PreProcessing*[11] is the first class to be mentioned, since it works as our superclass: inside *PreProcessing*, the method *get_all_files* is defined. For our research, we have modified it according to our needs. This class allows the user to perform the first reading of the input folder, by passing the path and the extension of the file, and return all the files contained.
In addition to that, two classes have been created: *MetaPreProcessing* and *CociPreProcessing*, 'children' of the first class, both containing a method called *splitted_to_file*[12], which has been adjusted according to the specific classes' needs and it is used in another method, newly created, named *split_input*[13].

META is our focus for answering the research questions, but we have performed some filtering also on this dataset to be able to merge it with the others.
In the class *MetaPreProcessing* we manage the processing of the META dump.
For the columns "id" and "venue" of the original files we have decided to keep as identifiers of publications and venues only, respectively, the DOIs and the ISSNs, removing thus all the other identifiers specified for each entity in META
The method *splitted_to file* takes in input an integer number that represents the lines' count, a list of lines, the column's name needed and the path to store the output file. The list taken in

---

[7] https://kanalregister.hkdir.no/publiseringskanaler/erihplus/
[8] https://en.wikipedia.org/wiki/ERIH_PLUS
[9] *Digital Object Identifiers*
[10] https://archive.softwareheritage.org/swh:1:dir:9c619d8cf358d6db044107069234972fd751f325;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/
[11] https://archive.softwareheritage.org/swh:1:cnt:2faf157225885e5420cdd740bee5311649c1b1a1;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/base.py;lines=27
[12] https://archive.softwareheritage.org/swh:1:cnt:e1cec205850fa2e58fc639f8ae3ce5981535ede0;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/pubmed.py;lines=67
[13] https://archive.softwareheritage.org/swh:1:cnt:e1cec205850fa2e58fc639f8ae3ce5981535ede0;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/pubmed.py;lines=88

input is produced in the *split_input* method. In the original dataset, the *id* column contains more than one identifier in the same cell. So the method *split_input* checks if there are more than one, splits the identifiers, checks if the identifiers are or not DOIs and keeps only those that are DOIs, removing all the others. The same process is operated on the *venue* column, by removing all the identifiers that are not ISSN. Once this action is performed, a new line is appended to a list and the method *splitted_to_file* is invoked.
The *splitted_to_file* method will return the output files with all the relevant information gathered after the first process of Meta.
An additional method is included in this class: *create_list_dois*. The purpose of this method is to create a list of DOIs that is needed to check if each COCI's DOI is also included in Meta. Thus, the *MetaPreProcessing* **must** be performed before *CociPreProcessing*.

In the class *CociPreProcessing* we manage the preprocessing of the COCI dump.
After the preprocessing, we will keep only the citations that are entirely contained in META. This means that the citations which have either the citing or the cited entity (or both) not contained in META are excluded from COCI_preprocessed. The method checks this using the files produced by *MetaPreProcessing* containing all the DOIs of META (that are passed as input of the class). The output files will be thus formed by two columns, "citing" and "cited". The method *split_input* is in charge of the preprocessing of COCI.
The method *splitted_to_file* has the same structure of the one described for *MetaPreProcessing* and uses as input the list produced in the *split_input* method. This last method takes as input a boolean parameter, *list_dois_excluded_from_meta*, that is used to control the creation of additional output files containing information about the DOIs not found in Meta (*excluded_dois_from_meta)*. First, *split_input* creates an empty list (*lines_coci_pre*) that will store the result of the process of COCI. If the value of the input parameter is *True* a new list, *lines_dois_excluded,* is created and it will be used for the production of *excluded_dois_from_meta* dataset. A set of Meta's id is created using the *CSVManager*[14] class starting from the input list of all the DOIs included in META. The iteration of the input zipped file is operated by entering directly in each of the zipped sub-folders containing the csv files. Each csv is opened, read and a dictionary is created taking the DOI that cites and the DOI that is cited. Four booleans variables will be valued with True or False according to the inclusion or not of the citing (or cited) DOIs in Meta.
If both citing and cited DOIs are in Meta, the two DOIs are respectively inserted into a dictionary with "citing" or "cited" as key, and a new line is appended to *lines_coci_pre* list, that will be later used for the creation of the output files. If instead either citing or cited (or both) DOIs are excluded from META, a new dictionary is created containing, in addition to "citing" and "cited", other two keys, specifying through a boolean value whether the DOI is citing (or cited) in Meta. As a last step, the new line is appended to the *lines_dois_excluded* list.
Finally, if the list (valid both for *lines_coci_pre* and *lines_dois_excluded*) exists, the *splitted_to_file* method is invoked to create the output of this first process.

The class *ErihPreProcessing* is responsible for the preprocessing of the ERIH-PLUS dataset. It creates a new CSV file with two columns "venue_id" and "ERIH_disciplines". "venue_id" is the union of the original columns "Online ISSN" and "Print ISSN" of ERIH-PLUS.

[14] https://archive.softwareheritage.org/swh:1:cnt:8362d20804ab87cf6862ffe37f2624e26634eff2;origin=https://github.com/opencitations/oc_meta;visit=swh:1:snp:dcbd1d5baf07b98504292230c53215f1e080e782;anchor=swh:1:rev:8510cdd12e3693202231fef27b0311ca7b88414d;path=/oc_meta/lib/csvmanager.py

This class is different from *CociPreProcessing* and *MetaPreProcessing* mainly for the reason that the ERIH-PLUS dataset is smaller than COCI and META.
In fact, it contains only two methods: *preprocess_ERIH_plus* and *write_csv*.
The first method creates a new empty list, then it opens the csv file and creates two dictionaries in which are stored the two ISSN (corresponding to *Print ISSN* and *Online ISSN* columns) with the "*venue_id*" as key. To these dictionaries, a new key ("ERIH_disciplines") is added with the value obtained from the column *ERIH PLUS Disciplines*. The dictionary is appended to the list. By using the second method, *write_csv*, the list is thus written and stored in a new csv output file, containing only the venues' ids and the disciplines associated with them.

**ErihMeta Class**

After cleaning all the dataset and keeping only the information relevant for this research, ErihMeta class was created. This class merges the results of the preprocess conducted on ERIH-PLUS and META on the "venue" column.
The main method is *erih_meta*, which identifies all the ISSN included in the "venue" column of Meta and adds, by calling the method *find_erih_venue,* the disciplines associated with that list of ISSN. A new file storing all the columns of Meta plus the column containing the ERIH-PLUS disciplines is generated.

**Counter Class**

The Counter class is the class responsible for answering the research questions. Some of the methods have been already explained in the *PreProcessing* class- in particular *get_all_files* and *splitted_to_file*.
This class is able to execute two different methodologies, one that entails the production of output files ("**Methodology1**"), reusable for other researches on the topic, and the other one that gives directly the answers to the questions ("**Methodology2**").
The constructor of the class requires three parameters:

- coci_preprocessed_path: Path to the directory that contains preprocessed COCI data
- erih_meta_path: Path to the directory containing ERIH_META data.
- num_cpus: number of cpu available for the execution of the program, by default it is set as the entire number of cpu available in the machine. This is also useful to define the number of threads to use for the execution of the program, which is defined as num_cpu * 4.

The method *create_additional_files* takes in input a boolean parameter ("*with_disciplines*"): if set to *True,* it creates a subset, *erih_meta_with_disciplines,* and it is filled with the id and the discipline contained in the column "*erih_disciplines*". If the value is *False*, the method searches for all the DOIs which are not associated with a discipline and fills a new file that corresponds to the subset called *erih_meta_without_disciplines*.
To make a connection between ERIH-PLUS, Meta and COCI, the *create_disciplines_map* allows to iterate over the preprocessed COCI files and to use the class CSVManager for searching in *erih_meta_with_disciplines* the DOIs included in COCI and the discipline associated to them. The output files are generated with four columns: "id", "citing", "cited" and

“disciplines”. According to the role that the DOIs has in the COCI’s citation, “citing” and “cited” are filled with True or False.
The files obtained with the previous method are used by *create_count_dictionaries* to generate two dictionaries: the keys are the SSH disciplines and the values are the total count of the occurrence of each discipline, either as a citing or cited entity. The most citing discipline and the most cited discipline with the related occurrences are thus obtained.
A method called *create_dataset_for_count* has been defined to answer in particular to the first and the third research questions. The output datasets are built by using *COCI_preprocessed* and with the subsets of *erih_meta*, *erih_meta_with_discipline* and *erih_meta_without_disciplines* managed with CSVManager. The files have four columns (“citing”, “is_citing_SSH”, “cited”, “is_cited_SSH”): the second and the fourth column contain a boolean value, *True* if the DOI is a SSH publication and *False* otherwise.
A simple count method is represented by *count_lines*, that counts the lines of each output file.

All the methods above mentioned are included and used in the main method of this class: *execute_count*, which is the method that the final user has to call to answer to the research questions proposed in this paper. It takes in input six parameters:

- the path of the output folder where all the produced files will be stored (*output_dir*);
- a boolean parameter (*create_subfiles*) that controls the production of additional files: if it is set to *True*, *create_additional_files* and *create_dataset_for_count* will be called and the output of those methods will be saved in the specified output folder (**Methodology1**); if it is set to *False*, the answers will be provided without producing any additional file (**Methodology2)**;
- three boolean parameters that allows the user to decide the answer to produce (*answer_to_q1*, *answer_to_q2*, *answer_to_q3*);
- an integer parameter (*interval*) which controls the number of lines that will be added to each file.

Thus, to answer the first question, if both *create_subfiles* and *answer_to_q1* are set as *True*, the method creates a dataset with the columns “id” and “erih_disciplines" containing the DOIs with the SSH disciplines associated. Then it calls *create_dataset_for_counts* with the parameter *is_SSH* set as True, which returns a dataset in which all the DOIs are associated with a discipline. The method *count_lines* is used to count all the lines of the files produced and returns the number of the citation that, according to COCI, involve -either as citing or cited entities- publications in SSH journals (according to ERIH-PLUS) included also in Meta.
To answer the second question, if both *create_subfiles* and *answer_to_q2* are set as *True*, the method *create_discipline_map* is called to create the files that will be used by *create_count_dictionaries*. This method will count the disciplines and will return the most citing and the most cited one.
To answer the third question, if both *create_subfiles* and *answer_to_q3* are set as *True*, the method calls *create_additional_files* with the input parameter set to *False* to create *erih_meta_without_disciplines.* The dataset with all the DOIs that are not associated with disciplines will be used in the *create_datasets_for_counts* method, with the parameter *is_SSH* set as *False*. It will return the number of citations that, according to COCI, start from and go to publications in Meta and are not included in SSH journals.

If *create_subfiles* is set to *False*, the method *iterate_erih_meta* creates two lists (*ssh_papers* and *not_ssh_papers*), a dictionary containing the DOIs associated with a discipline and a set in which all the ERIH-PLUS disciplines are contained. It reads all the CSV files resulting from the merge between ERIH-PLUS and Meta.
Two dataframes will be created using a mask, which fills the first (*ssh_df*) with the DOIs associated with SSH disciplines and the second (*not_ssh_df*) with the DOIs not associated with SSH disciplines. The method gets the unique values of the "id" column and appends it respectively to the lists previously created. After decoupling DOIs from the two lists, two sets are created. Thus, the method returns *ssh_set*, *not_ssh_set*, *unique_id_discipline_map* and *ssh_disciplines*.

The method *count_citations_in_file* takes as input the tuple resulting from the method just described and the path of the file to read - preprocessed COCI - from which it considers only the "citing" and "cited" columns.
To answer the first and third question, it compares the value of the "citing" and "cited" columns with both the first and the second set. If the DOI is in the *ssh_set*, the citation count with the key 'ssh' (the dictionary *citation_counts*) is incremented. Otherwise, the *not_ssh* count is incremented.
To answer the second question, the method checks if the citing or cited DOIs with an SSH discipline associated are included in the *ssh_set*, thus, for each discipline encountered, the counter is incremented and the discipline with the higher value is returned.
Finally, the function returns a tuple of three elements which represents the results: the discipline counter, the count of the citation in SSH journals and the count of citations not in SSH journals.

**Requirements & Problems**

The last paragraph of this section will be used to describe some requirements and problems that one may encounter trying to reuse the methodology provided.
First, the datasets taken in consideration are from 30 GB to 285 GB, so it is necessary to have a machine or an external disk that is capable of storing such a large amount of data. Then, it has to be pointed out that the process of this data requires a powerful processor, because just two of our four machines were able to elaborate and run them in a reasonable amount of time and without any problem.
In particular, we used the MacBookPro M1 (14'') with 16GB of RAM and 8core CPU with SSDs for computing all the processes and storing the data.
For what concerns the versions and libraries required for the python software, they can be found in our software[15].

# RESULTS

In this section we want to highlight the results to the answers obtained thanks to the process previously described, but also to provide a better description and possible usage of the files generated if chosen to:

[15] Olga Pagnotta, Sara Vellone, Marta Soricetti, & Lorenzo Paolini. (2023). Uncovering the Citation Landscape: Exploring OpenCitations COCI, OpenCitations Meta, and ERIH-PLUS in Social Sciences and Humanities Journals - SOFTWARE (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7978742

- The preprocess of COCI produce 13967 csv files
  - 673 files are also produced containing the COCI's DOIs excluded from Meta
- The preprocess of Meta produce 8438 csv files
  - 7623 files contain the DOIs of all the publication stored in Meta
- The preprocess of ERIH-PLUS produce just one file
- The merge between ERIH-PLUS and Meta produce 7622 csv files
  - 550 csv files contains the DOIs with a discipline associated
  - 7073 csv files contains the DOIs without a discipline associated
- 22030 csv files represent the datasets in which we have the COCI's DOIs (both citing and cited) associated with the disciplines
- 67380 files results from the merge between the pre-processed COCI, ERIH-PLUS and Meta

**Answers to our questions**

The first question was about the number of citations referring to publication in SSH journals included in ERIH-PLUS, by looking at citations data contained in COCI and Meta. The number attested by this research is of 220.295.011 citations.

The result of the second question shows that the most cited and the most citing discipline in the field of Social Science and Humanities is psychology, having 54.512.160 citing DOIs and 83.291.583 cited DOIs.

The result of the third question highlights the high number of citations that are not included in SSH journals, which is 1.176.384.557.

# DISCUSSION AND CONCLUSIONS

The results obtained are really interesting, in particular considering that for a long time the social importance of a research was assessed using economic indicators[16], which are starting to be considered as inefficient. In fact, the quality of the impact of a social science can be discussed only contextualizing it, because it depends on "the person, the problem, the time"[17], thus a rigid quantitative measure of the societal impact should be avoided. In previous studies, as the one conducted by Benedict et al[18]., some new measures for evaluating the impact of the social sciences and humanities were proposed and discussed, such as "Career profiles and target agreements" and "Extrapolation of best practices"[19], that focus on creating the right conditions for a research to be impactful.
Our results regarding the first and the third question show that the number of citations in the SSH field is significantly lower than the number of citations that are not in the SSH field. We can guess that the reason stands in the mechanisms that value the "scientific disciplines" more than

[16] Fecher, Benedikt, Sokolovska, Nataliia, Kuper, Freia, & Fenton, Alex. (2021). Impact of social sciences - (How) Can it be measured?. https://doi.org/10.5281/zenodo.5704639
[17] Ivi, p.2
[18] Ibidem
[19] Ibidem

the SSH ones, due to the importance that the word of science has in addressing societal challenges[20].

Surely, it is incredible that a gap of 956.089.546 citations exists between these two fields, only considering the datasets here discussed which - no matter how up-to-date - cannot entirely represent all the existing citations. In fact, our source dataset ERIH-PLUS doesn't contain all the venues that exist in Meta: our guess is that a considerable number of citations are not included in this research even if they are associated with a SSH discipline.

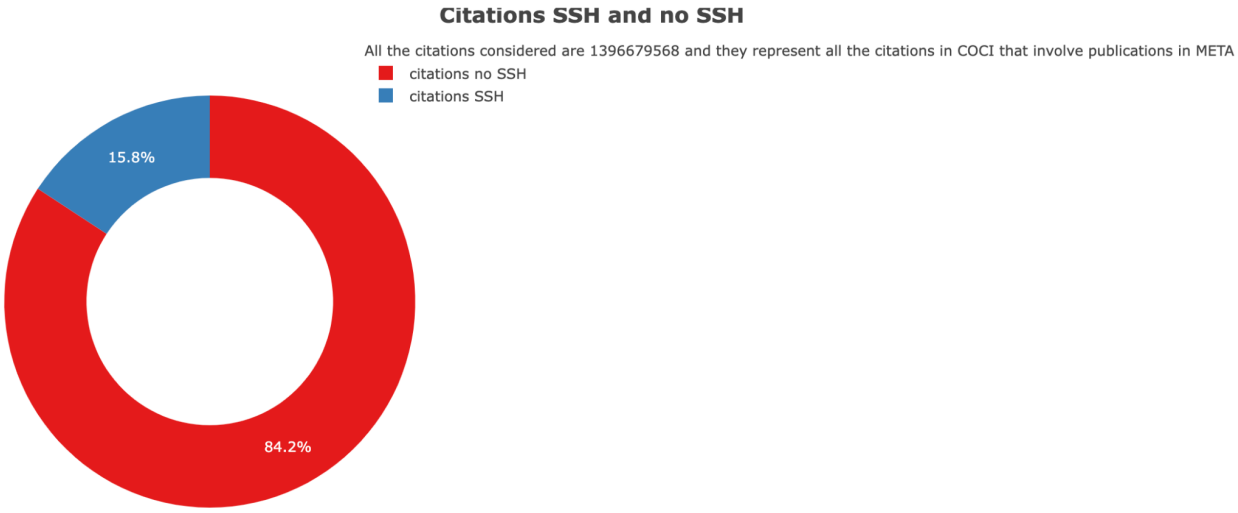


*Fig.1: Citations SSH vs citations not SSH according to our results*

We have decided to generate with our software some additional files to highlight these findings, *dataset_ssh* and *dataset_no_ssh*, through which it is possible to understand respectively whether a citation is totally (or just partially) included or excluded in the SSH field. The datasets also allow us to sustain the answers to our first and third question, because the same results can be obtained by counting the citations contained in these files.

We have also noticed that not all the DOIs included in COCI are included in Meta, with the consequence that "partial citations" can be found by cross-analysing the datasets, meaning that in a citation the citing DOI, but not the cited, may be present in Meta (and vice versa). Why this happens is out of the scope of this research. Further experiments can be made by using the additional files we have produced, *excluded_dois_from_meta*, that also give some information about the nature and the role of each DOI in the citation, i.e. if it is citing or cited.

[20] Ibidem

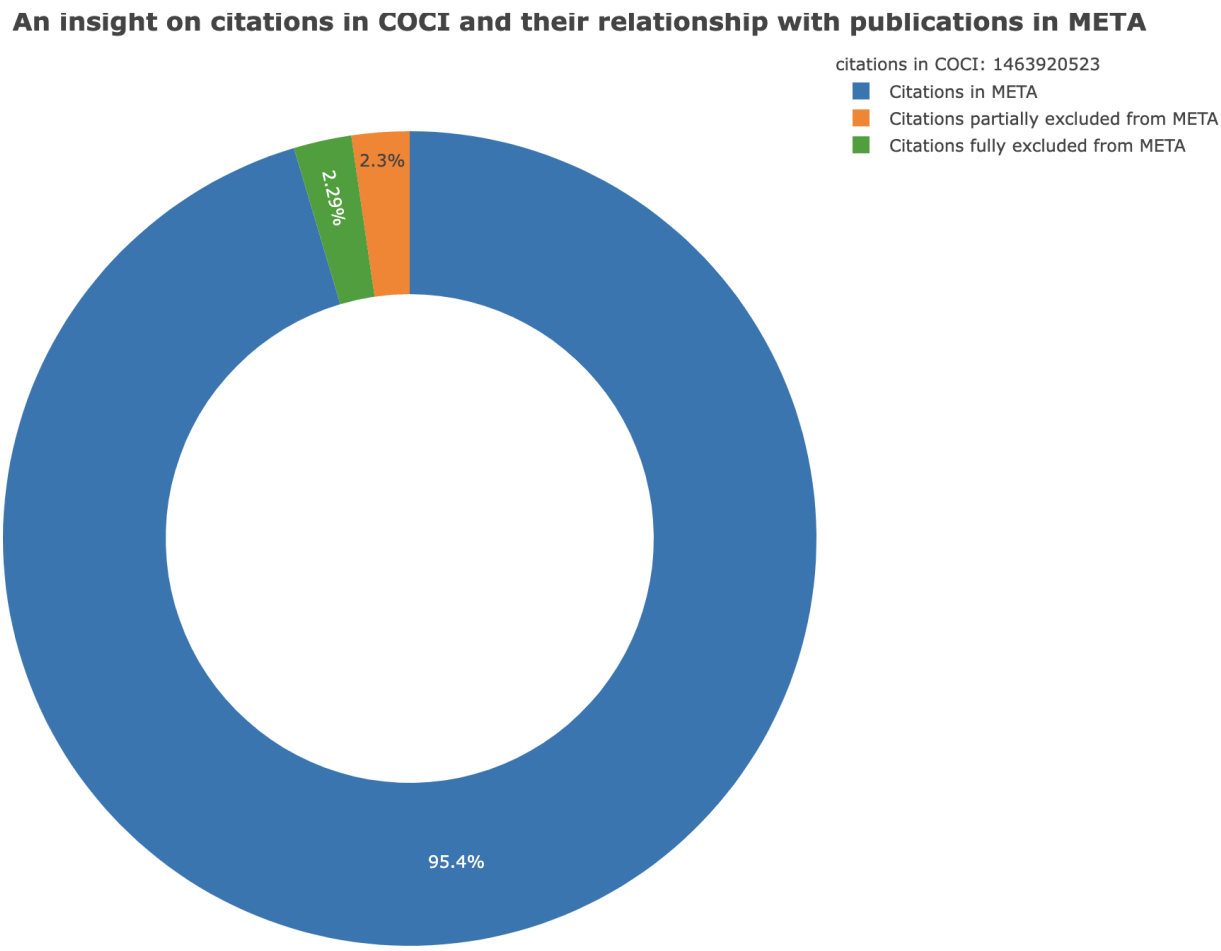


*Fig.2: Overall percentage of citations in META - Overall percentage of citations excluded from META*

For what concerns Psychology as the most citing and most cited discipline, we found this result fascinating and we tried to understand why.

Psychology is the science of behavior and mind, and includes the study of both conscious and unconscious phenomena (like feelings and thoughts). As social science, it aims to understand individuals and groups by establishing general principles[21].

If we consider how important the knowledge of the human factor and the psychological system is for developing healthy human relationship in society, it seems fair to state that the study of the mechanisms of improving the techniques of social thinking is one of the most pressing issues[22] that must be widely discussed. This seems to be confirmed by the second place, which is occupied by "Anthropology", both as citing and cited discipline.

[21] Abdullah MQ. (2019). Contemporary Issues in Psychological Sciences. https://doi.org/10.19070/2332-3000-1900045

[22] Mahmudjonov Ibrohimjon. (2022). THE POSITION OF SOCIAL PSYCHOLOGY IN THE WORLD TODAY. https://doi.org/10.5281/zenodo.6819322

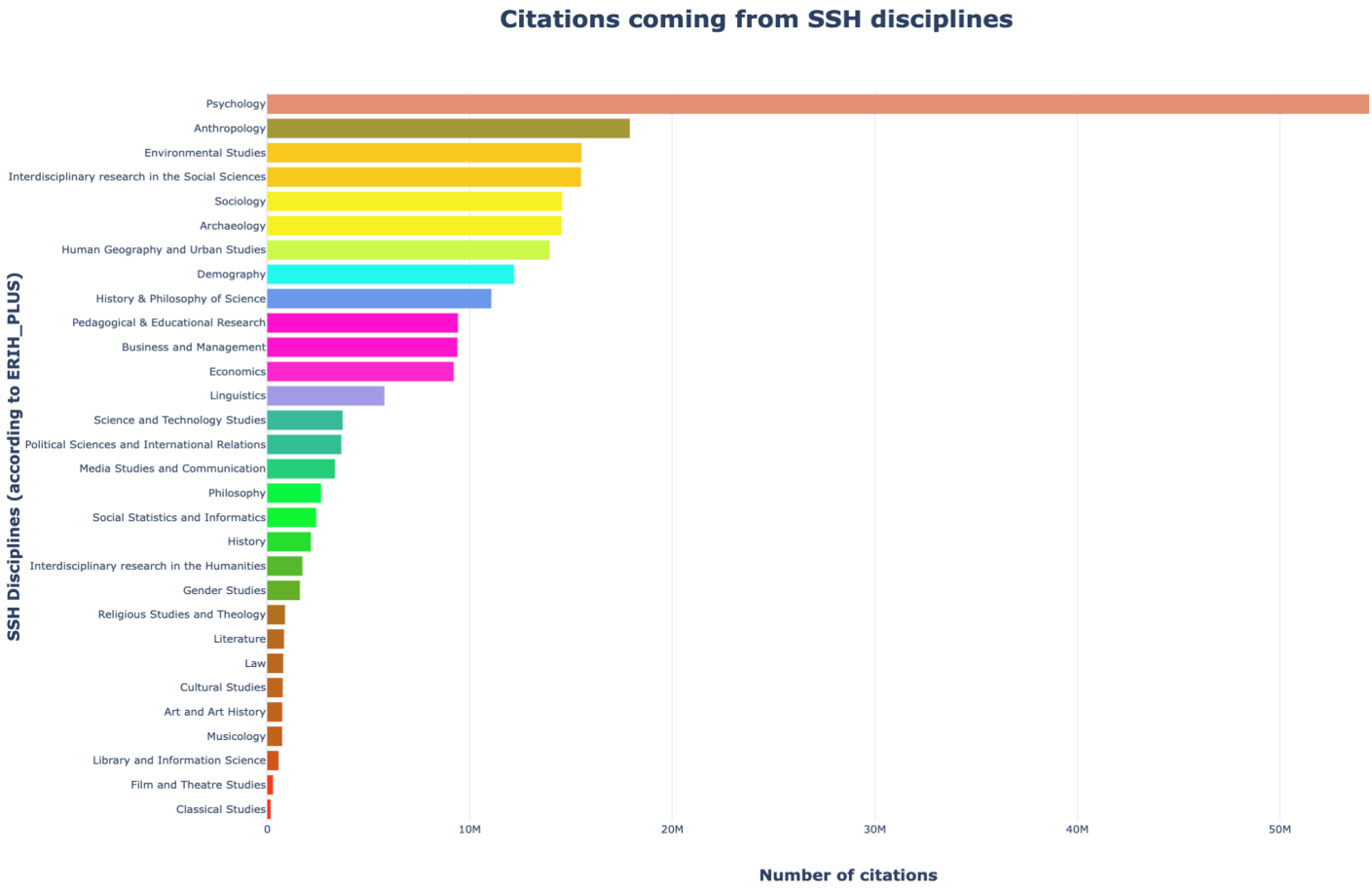


*Fig.3: Ranking of citing disciplines according to our results*

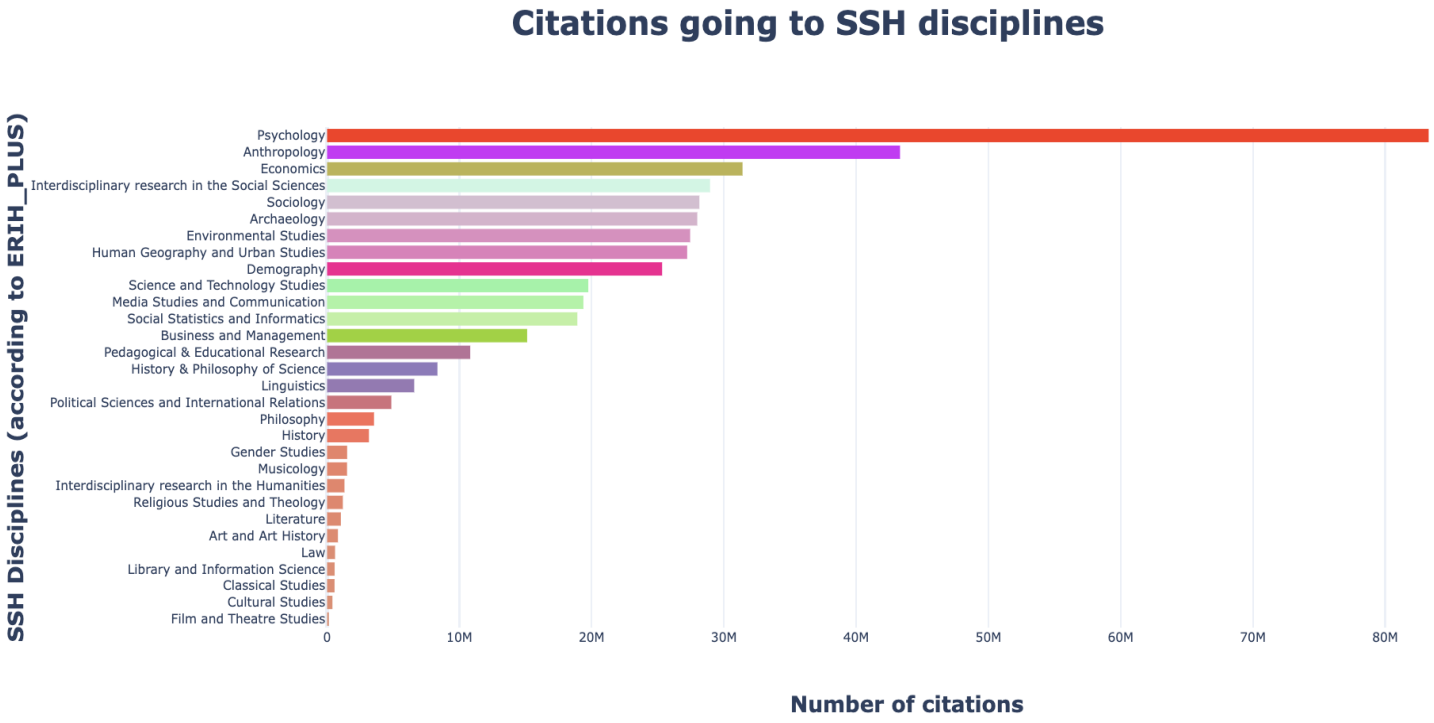


*Fig.4: Ranking of cited disciplines according to our results*


*Fig.5: First and second most citing discipline*

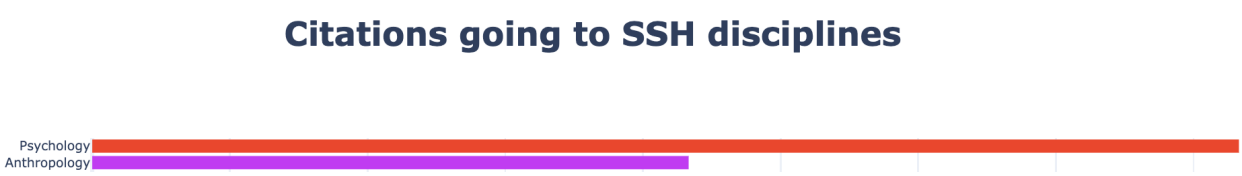


*Fig.6: First and second most cited discipline*

By looking at Fig.5 and Fig.6 we can point out two facts: first, a quantitative observation shows for both Anthropology and Psychology that the number of citations going to SSH disciplines is higher than the number of citations coming from SSH disciplines. We hypothesize that this phenomenon happens because of the high number of non-SSH journals citing SSH journals, an assumption which seems to be validated by the previous results (Fig.1).

Secondly, it seems that in our society there is a focus on the "human" and, particularly, on the "human behavior". Ellwood, in *Social Psychology and Social Science*[23], highlights how psychology doesn't deal with physiology, but with purposes, desires and emotions[24]. The human being is not just flesh that needs to be studied, but also mind, an aspect that appears to be equally important.
A study conducted by Dariusz Doliński, published in 2018, underlines how psychologists seem to be more interested in explaining why people display certain reactions more than demonstrating the conditions under which people display these reactions[25]. Doliński found the reason in a researchers' preference regarding the spread of statistical analysis applied to empirical data, which produces a more quantitative than qualitative vision of the discipline itself.
By analyzing the publication years of our results we have observed that for what concerns the citing DOIs (Fig.7), the year with the higher number is 2021, while the peak of cited DOIs (Fig.8) is in 2009.

[23] Ellwood, C. A. (1921). Social psychology and the social sciences. https://doi.org/10.1037/h0072869
[24] Ibidem
[25] Doliński, Dariusz. (2018). Is Psychology Still a Science of Behaviour?. Social Psychological Bulletin, 13((2)), e25025. https://doi.org/10.5964/spb.v13i2.25025

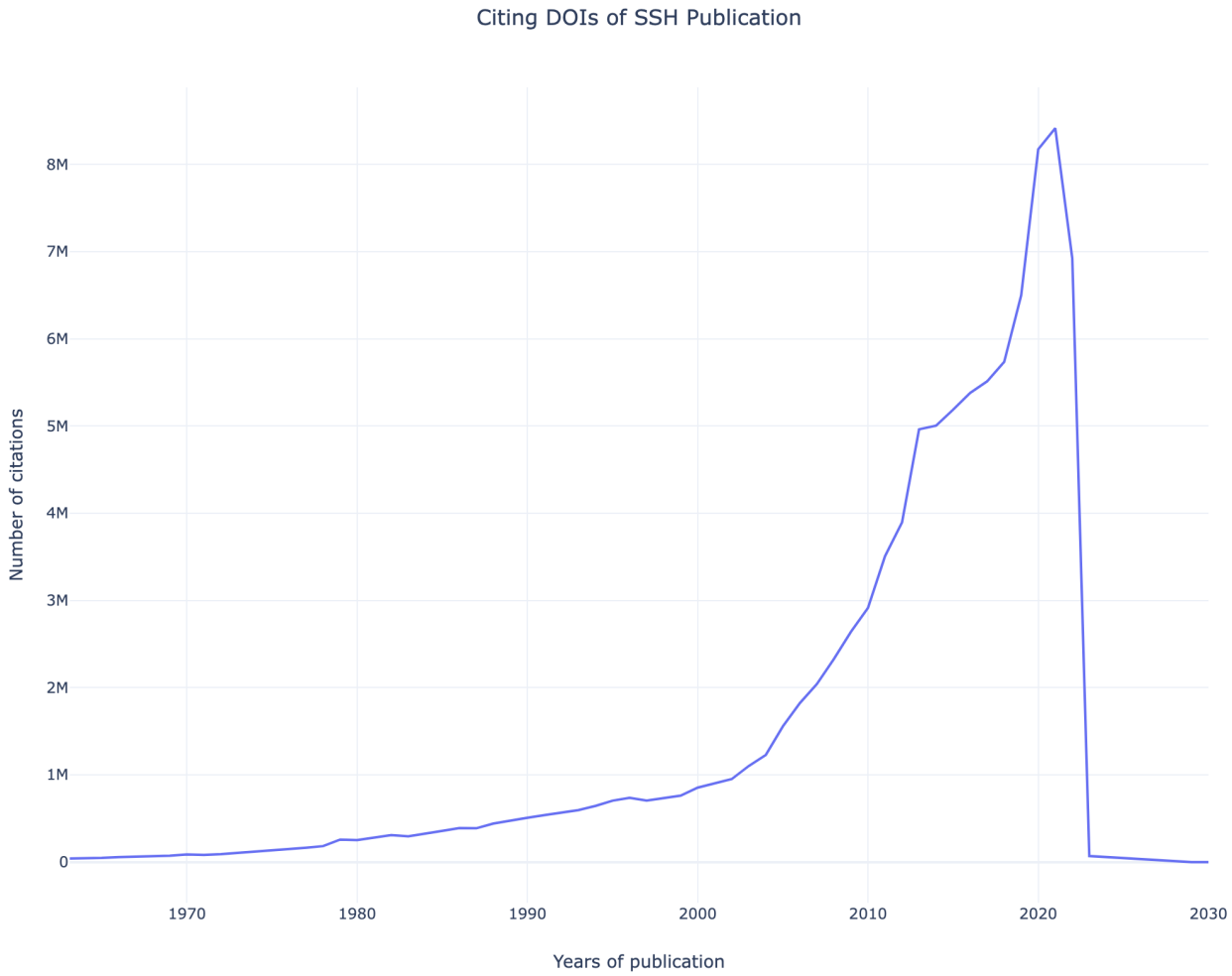


*Fig.7: Citing DOIs of SSH publications - years of publications*

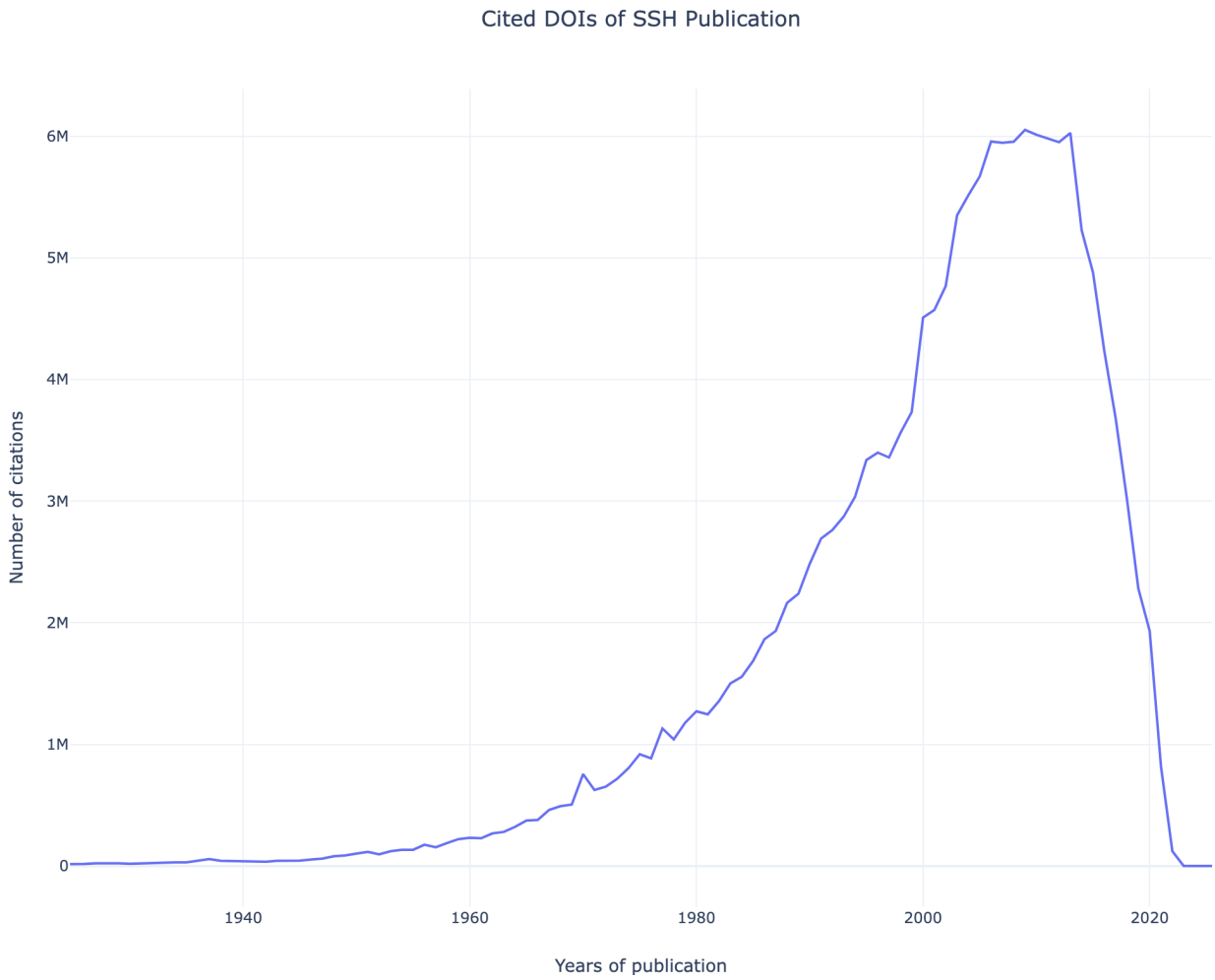


*Fig.8: Cited DOIs of SSH publications - years of publications*

Even if we didn't highlight the years in which there are psychology citations, we can guess that Doliński's supposition finds a match in our results. According to his theory, Psychology seems to be "a more scientific" discipline, and in addition to the huge impact on the society that it seems to have, it's reasonable why this discipline results as the most citing and the most cited one.

With this research we tried to provide as much materials as possible to guarantee future investigations in this field. We conducted our experiments approaching them mainly with a quantitative point of view, but a qualitative evaluation could return interesting results.

**Acknowledgement**: We acknowledge Silvio Peroni for providing us with the starting research questions.

# References


***Papers***

Abdullah MQ. (2019). Contemporary Issues in Psychological Sciences. https://doi.org/10.19070/2332-3000-1900045

Castellano, C., Radicchi, F. On the fairness of using relative indicators for comparing citation performance in different disciplines. Arch. Immunol. Ther. Exp. 57, 85–90 (2009). https://doi.org/10.1007/s00005-009-0014-0

Dassa Michèle, Kosmopoulos Christine et Pumain Denise, « JournalBase - A Comparative International Study of Scientific Journal Databases in the Social Sciences and the Humanities (SSH) », Cybergeo: European Journal of Geography [En ligne], Science et Toile, document 484, mis en ligne le 08 janvier 2010, consulté le 27 mai 2023.URL:http://journals.openedition.org/cybergeo/22862;DOI: https://doi.org/10.4000/cybergeo.22862

Doliński, Dariusz. (2018). Is Psychology Still a Science of Behaviour?. Social Psychological Bulletin, 13((2)), e25025. https://doi.org/10.5964/spb.v13i2.25025

Ellwood, C. A. (1921). Social psychology and the social sciences. https://doi.org/10.1037/h0072869

Fecher, Benedikt, Sokolovska, Nataliia, Kuper, Freia, & Fenton, Alex. (2021). Impact of social sciences - (How) Can it be measured?. https://doi.org/10.5281/zenodo.5704639

Heibi, I., Peroni, S., & Shotton, D. (2019). Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. Scientometrics, 121(2), 1213–1228. https://doi.org/10.1007/s11192-019-03217-6

Mahmudjonov Ibrohimjon. (2022). THE POSITION OF SOCIAL PSYCHOLOGY IN THE WORLD TODAY. https://doi.org/10.5281/zenodo.6819322

***Links***

Crossref:
https://www.crossref.org/
OpenCitation Meta:
http://opencitations.net/meta
ERIH-PLUS:
https://kanalregister.hkdir.no/publiseringskanaler/erihplus/
https://en.wikipedia.org/wiki/ERIH_PLUS

***Softwares***

Current Research Software:
Olga Pagnotta, Sara Vellone, Marta Soricetti, & Lorenzo Paolini. (2023). Uncovering the Citation Landscape: Exploring OpenCitations COCI, OpenCitations Meta, and ERIH-PLUS in Social Sciences and Humanities Journals - SOFTWARE (Version 1). Zenodo. https://doi.org/10.5281/zenodo.7978742

OpenCitation *Preprocessing*:
https://archive.softwareheritage.org/swh:1:dir:2480b5ba7cbecfd13bc7bffc4b8d5601b035540d;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/

OpenCitation *Preprocessing/base.py*:
https://archive.softwareheritage.org/swh:1:cnt:2faf157225885e5420cdd740bee5311649c1b1a1;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/base.py

OpenCitation *Preprocessing/pubmed.py*:
https://archive.softwareheritage.org/swh:1:cnt:e1cec205850fa2e58fc639f8ae3ce5981535ede0;origin=https://pypi.org/project/oc-preprocessing/;visit=swh:1:snp:b429746305d915b577b0ed022b2650b70ecf5dc2;anchor=swh:1:rel:44fb3b0a058877ea4ef15490a499391c910a384f;path=/oc_preprocessing-0.0.5/preprocessing/pubmed.py

OpenCitation *Preprocessing/csvmanager.py*:
https://archive.softwareheritage.org/swh:1:cnt:8362d20804ab87cf6862ffe37f2624e26634eff2;origin=https://github.com/opencitations/oc_meta;visit=swh:1:snp:dcbd1d5baf07b98504292230c53215f1e080e782;anchor=swh:1:rev:8510cdd12e3693202231fef27b0311ca7b88414d;path=/oc_meta/lib/csvmanager.py