protocols.io

🌐 Retrieving SSH Journals Citation Information from three datasets (COCI, META and ERIH-PLUS) - Workflow

Marta Soricetti[1], Sara Vellone[1], Olga Pagnotta[1], ellepuntopi.[1]

[1]University of Bologna

Sara Vellone

ABSTRACT

Purpose: we want to find out

1. by looking at citations data contained in COCI, the number of citations included in Meta which refer to publication in SSH (Social Sciences and Humanities) journals indicated in ERIH-PLUS
2. the disciplines citing the most VS the disciplines cited the most
3. the citations from/to publication contained in Meta which are not included in SSH journals We want to create a connection between these three different datasets for having an overall view of the citations present in each of them.

Methodology: we approach the problem from a computational point of view, by building a python software able to analyse the data, querying them in order to retrieve the info needed and to present the results in a clear and understandable way.

Findings: for what concerns the findings, we can see that there are no meaningful differences in the number of citations coming from different disciplines, since it is related to the subject of the study, while the ones cited the most belong to psychology, health and science studies.

Originality/Value: the research conducted by us can be defined as very valuable, since it adds information to existing resources with the aim of facilitating their use and allowing the users to have a clearer view of the data contained in each dataset. Further development can be made, for example analysing other disciplines, to have the same overview as the one created by us but related to other fields.

APR 12, 2023

Protocol status: Working
We use this protocol and it's working

Created: Apr 03, 2023

Last Modified: Apr 12, 2023

PROTOCOL integer ID:
79935

Reading Input Data

1   We started to analyse the datasets using pandas:

- Meta: csv dataset of Open Citations Meta
- COCI: COCI dump
- ERIH-PLUS: list of approved journals

Dataset	
META	NAME
https://opencitations.net/meta	LINK

Dataset	
COCI	NAME
https://opencitations.net/index/coci	LINK

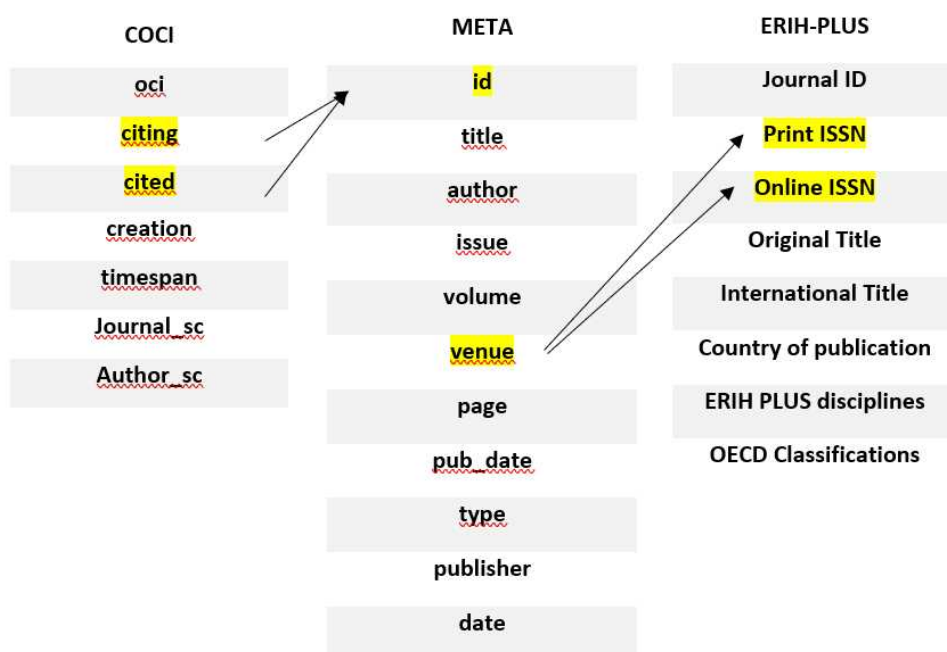| Dataset	
ERIH-PLUS	NAME
https://kanalregister.hkdir.no/publiseringskanaler/erihplus/periodical/listApproved	LINK

Processing of Input Data

2   We tried to define a mapping of the datasets and this is the result:

Mapping of the three datasets and relevant columns

3 We processed the data by cleaning them, keeping only the relevant information for our purpose.

Merged dataset

4 We merged all our input datasets for creating a new one with unified columns.

Analyzing dataset

5 We performed data analysis in order to retrieve information for answering to the following questions:
1. How many citations (according to **COCI**) involve, either as citing or cited entities, publications in SSH journals (according to **ERIH-PLUS**) included in **OpenCitations Meta?**
2. What are the disciplines that cites the most and those cited the most?
3. How many citations start from and go to publications in OpenCitations Meta that are not

included in SSH journals?

5.1 In order to answer to the first question we performed operations that returned the number of citation involving publications in SSH journals included in Meta.

5.2 In order to answer to the second question we retrieved the names of the disciplines that cites the most and the most cited disciplines.

5.3 In order to answer to the third question we extracted the number of citations that start to/go from Meta but are not included in SSH journals.

Results

6