



DATA NOTE

MapOSR - A mapping review dataset of empirical studies on Open Science [version 1; peer review: 2 approved, 2 approved with reservations]

Jana Lasser ^{1,2}, Jürgen Schneider ³, Thomas Lösch ⁴, Ronny Röwert⁵, Tamara Heck ⁴, Clemens Bluemel⁶, Maike Neufend⁷, Isabel Steinhardt⁸, Stefan Skupien ⁹

¹Complexity Science Hub Vienna, Vienna, 1080, Austria

²Institute for Interactive Systems and Data Science, Graz University of Technology, Graz, 8010, Austria

³Tübingen School of Education, University of Tübingen, Tübingen, Germany

⁴Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

⁵Hamburg University of Technology, Hamburg, Germany

⁶German Center for Higher Education Research and Science Studies, Berlin, Germany

⁷Open Access Office Berlin, Freie Universität Berlin, Berlin, Germany

⁸University of Paderborn, Paderborn, Germany

⁹Berlin University Alliance, Berlin, Germany

V1 First published: 18 May 2022, 11:535

<https://doi.org/10.12688/f1000research.121665.1>

Latest published: 18 May 2022, 11:535

<https://doi.org/10.12688/f1000research.121665.1>

Abstract

Research that investigates respective researchers' engagement in Open Science varies widely in the topics addressed, methods employed, and disciplines investigated, which makes it difficult to integrate and compare its results. To investigate current outcomes of Open Science research, and to get a better understanding on well-researched topics and research gaps, we aimed at providing an openly accessible overview of empirical studies that focus on different aspects of Open Science in different scientific disciplines, academic groups and geographical regions. In this paper, we describe a data set of studies about Open Science practices retrieved following a PRISMA approach to compile a literature review. We included studies from the Scopus and Web of Science databases with keywords relating to Open Science between the years 2000 and 2020, as well as a snowball search for relevant articles. Studies that did not investigate any aspect of Open Science, or weren't peer-reviewed were excluded, resulting in a total of 695 remaining studies. The data set was collaboratively annotated to ensure intercoder reliability of the coded data.

Keywords

Open Science, Open Access, Open Data, Open Education, Open Evaluation, Open Methods, Open Participation, Open Policies, Open Software, Open Tools

Open Peer Review

Approval Status ✓ ? ? ✓

1 2 3 4

version 1 ✓ ? ? ✓
18 May 2022 view view view view

1. Katja M. Mayer , University of Vienna, Vienna, Austria
2. Judith J. de Haan , Utrecht University, Utrecht, The Netherlands
3. Elena Šimuković , ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland
4. Andrew M. Cox , University of Sheffield, Sheffield, UK

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Research on
Research, Policy & Culture gateway.

Corresponding author: Ronny Röwert (ronny.roewert@tuhh.de)

Author roles: **Lasser J:** Data Curation, Software, Visualization, Writing – Original Draft Preparation; **Schneider J:** Data Curation, Validation, Writing – Original Draft Preparation; **Lösch T:** Conceptualization, Data Curation, Project Administration, Writing – Original Draft Preparation; **Röwert R:** Conceptualization, Data Curation, Project Administration, Writing – Original Draft Preparation; **Heck T:** Data Curation, Writing – Original Draft Preparation; **Bluemel C:** Data Curation, Writing – Original Draft Preparation; **Neufend M:** Data Curation, Writing – Original Draft Preparation; **Steinhardt I:** Data Curation, Writing – Original Draft Preparation; **Skupien S:** Conceptualization, Data Curation, Project Administration, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: JL has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101026507 – STAY. The open access publishing is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 491268466 and the Hamburg University of Technology (TUHH) in the funding programme Open Access Publishing.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Lasser J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Lasser J, Schneider J, Lösch T *et al.* **MapOSR - A mapping review dataset of empirical studies on Open Science [version 1; peer review: 2 approved, 2 approved with reservations]** F1000Research 2022, 11:535
<https://doi.org/10.12688/f1000research.121665.1>

First published: 18 May 2022, 11:535 <https://doi.org/10.12688/f1000research.121665.1>

Introduction

Open Science is still only vaguely defined. Different initiatives are subsumed under the label of Open Science, coming from different communities which share the goal of making sciences more open and transparent. The Open Science communities have made attempts to define key elements of Open Science - which are also referred to as the pillars of Open Science, that is, open access to publications, open data and open source (FOSTER Taxonomy of Open Science¹). Findings from bibliometric studies indicate that research dealing with Open Science as a phenomenon either by exploring its concepts, by assessing Open Science initiatives both at the national or international level or by exploring Open Science research practices (Levin *et al.*, 2016) have increased (Blümel & Beng, 2018). Yet, research that investigates engagement in Open Science varies widely in the topics addressed, methods employed, and disciplines investigated. This makes it difficult to integrate and compare results and get deeper insights on how Open Science and related practices evolved in science, or if the Open Science movement has any impact on research practices (Christensen *et al.*, 2020). To get a better understanding of Open Science research and investigate aspects of Open Science, we are providing an openly accessible overview of peer-reviewed empirical studies that focus on the attitudes, assessments, and practices of Open Science among individuals, communities, and organizations.

With this approach, we intend to clarify the current understanding of Open Science. Empirical studies capture diverse aspects of Open Science: among others, different disciplines, practitioner groups, geographical scopes and user groups are investigated. For instance, numerous empirical survey-based studies have asked similar questions, but often to different groups of respondents. Therefore, a complementary overview of existing studies will allow us to identify which user groups are less covered in the current research landscape.

Empirical studies were collected following a Preferred Reporting Items for Systematic Reviews (PRISMA) workflow and then annotated along five categories. The collected data serves three purposes, among others: First, other researchers in the field of Open Science may use the data for further in-depth analysis and synthesize data in a systematic review or any other format synthesizing research. Second, the data can be used as an annotated literature corpus that allows for a curated introduction in literature on Open Science. Third, Open Science practitioners (*e.g.* librarians, Open Science officers at universities, funding bodies) can use the data as a source of information on Open Science studies.

Methods

Data collection protocol

We designed the study as a mapping review. Our aim was to identify all empirical studies concerned with Open Science or any of its key elements, to be used as a basis for deeper investigations. Research on Open Science and its concrete concept varies, and the term “Open Science” is not new. However, Open Science as a movement of new research practices enabled by means of technical innovations on the Internet, has been discussed for over twenty years (Bartling & Friesike, 2014). Following the Foster taxonomy of Open Science, mapping in this study covered research related to key elements of Open Science (Open Data, Open Access, and Open Source) which also guided our search strategies. Our aim was to map research investigating these key elements of the Open Science movement, and not to identify the extent of Open Science concepts discussed, as a scoping review might aim for (*cp.* Grant & Booth, 2009). Moreover, in this first step, we did not synthesize any results like in a systematic review, but annotated the publications with five key features to give a better overview of the nature of the studies. With these settings and restrictions, we consider our study as a mapping review of empirical studies on defined Open Science elements.

Literature search and screening

Considering the recommendations on literature reviews (Gough *et al.*, 2017), we carried out a systematic search, included and excluded publications based on factual criteria, and annotated the relevant publications to characterize main study design and the key features regarding the covered Open Science aspects, study method, disciplinary focus, targeted group and geographical scope. We did not pre-register the review because the idea was developed out of a research group that first collected Open Science studies, and then went on to expand the work with a systematic search and annotation. This first snowball search went over a period of six months. Researchers made an announcement on Twitter and [researchgate.net](https://www.researchgate.net) in June 2020 and invited colleagues to contribute to the collection of empirical studies on Open Science. The results were included in the project’s publicly available Zotero-Library². The first entry was made on 30 June 2020, the last on 16 March 2021. The snowball search yielded 126 publications.

In addition, we conducted a systematic literature search on January 26th and 27th, 2021. We searched in the Web of Science (all indices) and Scopus databases. The search query consists of two blocks: a) terms of the Open Science elements, b) terms describing any empirical study (see Table 1).

¹See https://www.zotero.org/groups/2526436/meta-research_on_os-related_surveys/library

²<http://www.fosteropencience.eu/>

Table 1. Original search queries applied in the Web of Science and Scopus.

Web of Science Indexes=SCI-EXPANDED, SSCI, A&HCI, ESCI Timespan=2000-2020 Note: Title (shown here), abstract and author keywords for block b searched separately in original search	(TI=(“open science” OR “open access” OR “open scholar*” OR “open research data” OR “open data” OR “open evaluation” OR “open code” OR “open peer review” OR “open method*”) AND TI=(“action research” OR “case stud*” OR “Ethnograph*” OR “Evaluation Method*” OR “Evaluation Research” OR “EXPERIMENT” OR “Focus Group*” OR “Field Stud*” OR “fieldwork*” OR “Interview” OR “Mixed Methods Research” OR “Survey” OR “observation” OR “Participatory Research” OR “Qualitative Research” OR “Question*” OR “Statistical Analysis” OR “Statistical Stud*” OR “empiric*” OR “qualitative method*” OR “quantitative method*”) AND DOCUMENT TYPES: (Article OR Book OR Book Chapter OR Proceedings Paper)
Scopus Note: publication date limited to 2000-2020, not shown in string here	TITLE (“open science” OR “open access” OR “open scholar*” OR “open research data” OR “open data” OR “open evaluation” OR “open code” OR “open peer review” OR “open method*”) AND TITLE-ABS-KEY (“action research” OR “case stud*” OR “Ethnograph*” OR “Evaluation Method*” OR “Evaluation Research” OR “EXPERIMENT” OR “Focus Group*” OR “Field Stud*” OR “fieldwork*” OR “Interview” OR “Mixed Methods Research” OR “Survey” OR “observation” OR “Participatory Research” OR “Qualitative Research” OR “Question*” OR “Statistical Analysis” OR “Statistical Stud*” OR “empiric*” OR “qualitative method*” OR “quantitative method*”) AND (LIMIT-TO (DOCTYPE, “ar”) OR LIMIT-TO (DOCTYPE, “cp”) OR LIMIT-TO (DOCTYPE, “ch”) OR LIMIT-TO (DOCTYPE, “bk”))

We deliberately excluded terms relating to open education, like open educational resources (OER) and open educational practices. Although OER are mentioned as part of Open Science (see *e.g.* FOSTER), research on them as well as educational practices span a different research field quite separated from discussions on Open Science (Scanlon, 2013). Including OER and similar research would therefore have resulted in a very large corpus, which was beyond the scope of this study. Similarly, we excluded citizen science from our search. Block b was necessary to limit the retrieved publications to a manageable number, as block a alone would have resulted in a large number of non-empirical studies discussing Open Science. After testing several search strings and checking the results, we decided to search Open Science elements in the title field only. Terms describing empirical studies were searched in title, abstract, and (author) keywords. Additionally, we limited results to the document types article, book, book chapter, and proceedings paper. As the term “Open Science” is rarely mentioned in the research literature before 2000 (Blümel & Beng, 2018), the date range was specified from 2000 to 2020. This search yielded 3651 publications. Table 1 shows the original queries for the Web of Science and Scopus.

The snowball search and the systematic search in the two databases resulted in 3777 publications. From these, we removed 842 duplicates (see Figure 1). The titles and abstracts of the remaining 2935 publications were independently screened by three coders (all authors of this study) according to the following inclusion criteria:

- The publication deals with any aspect of Open Science (excluding OER and citizen science).
- The study focused on Open Science inside academia (*e.g.* exclusion of topics such as open government data or industry-based research).
- The study includes the collection of empirical data.
- The publication is written in English, German, Italian, French or Spanish. Unfortunately, languages had to be limited according to the coders’ language skills.

Disagreement was resolved through discussion. The screening resulted in 2101 publications being excluded.

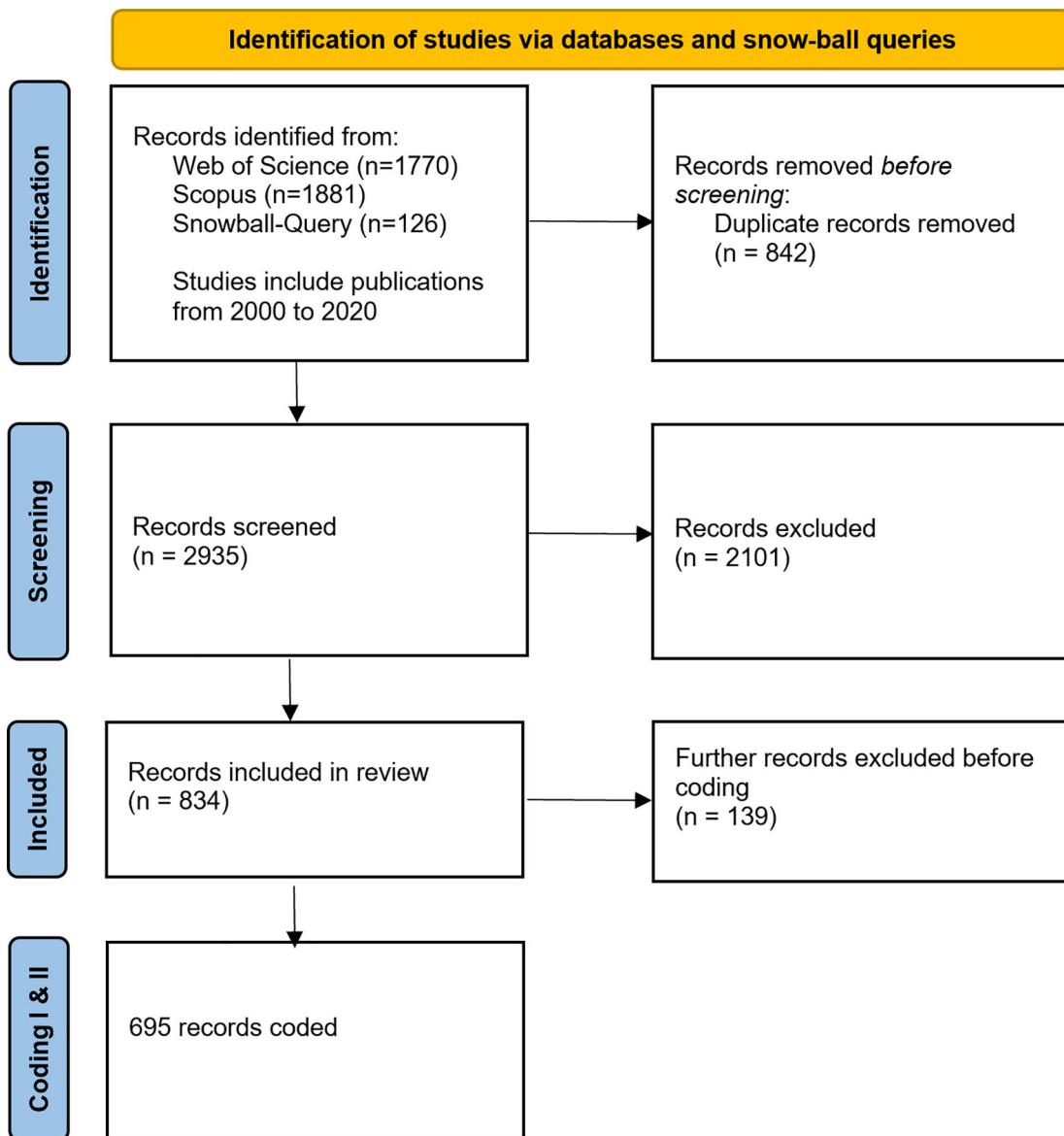


Figure 1. PRISMA Flowchart depicting the search and selection process based on [Page et al. \(2021\)](#).

Annotating key features

The annotations of key features followed several iterative steps. The 834 publications were coded along five categories as described in the codebook (see [Table 2](#) and [mapOSR_codebook_V4.csv](#) in the data repository, *Extended data*, [Lasser et al., 2022](#)) Action, method, discipline, group and geo scope. Within all categories, several labels could be awarded at the same time. We distributed the publications randomly across nine coders. These coders were trained in the codebook through joint development, refinement, and discussion in two rounds of coding.

During the coding process, we excluded another 139 studies that did not meet the inclusion criteria upon closer inspection, e.g. for some of the studies the empirical design was not clear. Most exclusions resulted mainly from duplications between conference papers and corresponding publications. The final sample of coded publications therefore included n=695 publications. We adapted the codebook during our process with regard to the following aspects: In the *action* category we assessed which aspect of Open Science was targeted in the publication. We adapted the labels within the category based on the FOSTER taxonomy and added ‘open education’ and ‘open participation’ (categories are explained in [Table 2](#)). We note that while we did not include Open Education and Open Participation in our database search, we still included them in our codebook, to leave room for future extensions of our approach to these categories. Furthermore, we converted ‘open

Table 2. Condensed version of categories and labels from the codebook.

Category	Label	Description
Action	openaccess	Open Access to research publications
	opendata	Open Data as data of a particular research, research data management practices
	openmethod	Open Methodology - Collecting open practices and processes throughout the research process, including reproducibility/replicability studies
	openevaluation	Open Metrics and Impact, Open Peer Review
	openpolicies	Policies governing OS-practices of organizations and subjects
	opensoftware	Open software/code available to members of research process or created by them
	opentools	Open Hardware
	openeducation	Educational resources created as open for the use in teaching
	openparticipation	Citizen Science, inclusion of other than the involved researchers in the research process
	openscience	If targeted in general without reference to any of the practices above
Method	survey	survey
	interview	qualitative approaches, including focus group discussions
	biblio	bibliometrical studies, e.g. studies based on articles and journals
	document review	e.g. literature review, policy document analysis
	othermeth	e.g. action research, models, experiments, data visualisation, altmetrics
Discipline	natscie	Natural Sciences
	engtech	Engineering and Technology
	med	Medical and Health Sciences
	agric	Agricultural Sciences
	socscie	Social Sciences
	hum	Humanities
	nonspecificdisc	No specific discipline targeted
Group	researcher	scientists mainly occupied with research in all career stages
	librarian	employees of university libraries or other libraries
	university	University as an organization
	unisupportstaff	Staff at universities not including researchers and librarians
	publisher	publishing house as organizations, editors
	policy	e.g. policy-maker, government officials, legislators
	funder	funding organisations that fund science and Higher Education
	business	any commercial (profit/non-profit) actor involved in Open Science
	practitioner	any practitioner outside of the academy, e.g. NGO
Geo scope	Alpha 3-Codes for countries, 2-digit continent codes if only continents are mentioned	
	nonspecificgeo	Use nonspecific if study does not target specific country or region

reproducible research' into a broader 'open methodology'. The second category describes the methods that are applied to empirically study the chosen aspect of Open Science, such as bibliometric studies or surveys. In the third category we coded the disciplines that are targeted with the study, such as engineering or social sciences. The selection of labels for this category is based on the OECD-Frascati Manual ([OECD, 2015](#)). The fourth category describes the group under investigation, such as researchers or librarians. In the last category we recorded the geographical scope of the empirical study according to design and included cases. The labels in this category were based on the ISO 3166-1 alpha-3 codes for countries.

Data validation

After manual annotation, we performed an automated data cleaning step to correct misspelled labels. The code used to perform the data cleaning is publicly available (see file `clean_data.ipynb` in the code repository; Lasser & Schneider, 2022). This included replacing two-digit country codes with three-digit country codes where necessary, replacing “missing” and “none” with NaN values and unifying label names such as policies, which was mapped to “openpolicies”. A list of all encountered misspellings is provided in the data cleaning code accompanying this publication. In addition, the letter and “=” symbol preceding each label was stripped from the entries. The consistency of the data was then checked by comparing the labels present in each category (action, method, discipline and group) to the labels allowed by the coding scheme. Country codes in the data set were manually checked for consistency.

Since each coded category was not exclusive, each entry could contain a list of labels, separated by a semicolon. Entries were first automatically split into a list of entries. Categories were then split into as many columns as possible, labels allowed in them and dummy-coded to only contain boolean values. For example, the category “method” was therefore split into five columns with the column names “method_biblio”, “method_documentreview”, “method_interview”, “method_survey”, and “method_other”. An entry that would originally read “m=biblio; m=survey” would be split into the following column entries: “method_biblio=True”, “method_documentreview=False”, “method_interview=False”, “method_survey=True”, “method_other=False”.

An overview of the development of publication numbers between the years 2000 and 2020 for the category “Action” is shown in Figure 2. The categories “Method”, “Discipline”, “Group” and “Geo Scope” are summarized in Figure 3. Code to reproduce the figure is publicly available (see file `create_visualizations.ipynb` in the code repository; Lasser & Schneider, 2022).

Interrater reliability

Interrater agreement was calculated for each label within the five categories of the codebook. For this purpose, we double-coded 63 of the 697 publications (9%). The coders were evenly distributed across the data underlying the computation of the interrater agreement. The occurrence for several of the dichotomous labels (dummy transformed from the categories) was strongly imbalanced. An example of this is the occurrence frequency of certain countries in the geo category that were never or very rarely coded. Cohen’s kappa, the standard measure of agreement for dichotomous categorical variables, leads to biased values for skewed variables and was therefore not appropriate in this case (Xu & Lorber, 2014). We therefore resorted to simple percentage agreement values for all labels.

Based on the results, we adapted category labels for geo and discipline, *i.e.* we recoded the labels of “geo=none” and “geo=all” to “geo=unspecific”, and “discipline=none” and “discipline=all” to “discipline=unspecific” again due to the skewed distribution. A reason for the coders’ disagreements for these category labels was that empirical studies do not always explicitly state their geographical or disciplinary focus. For example, bibliometric studies usually investigate

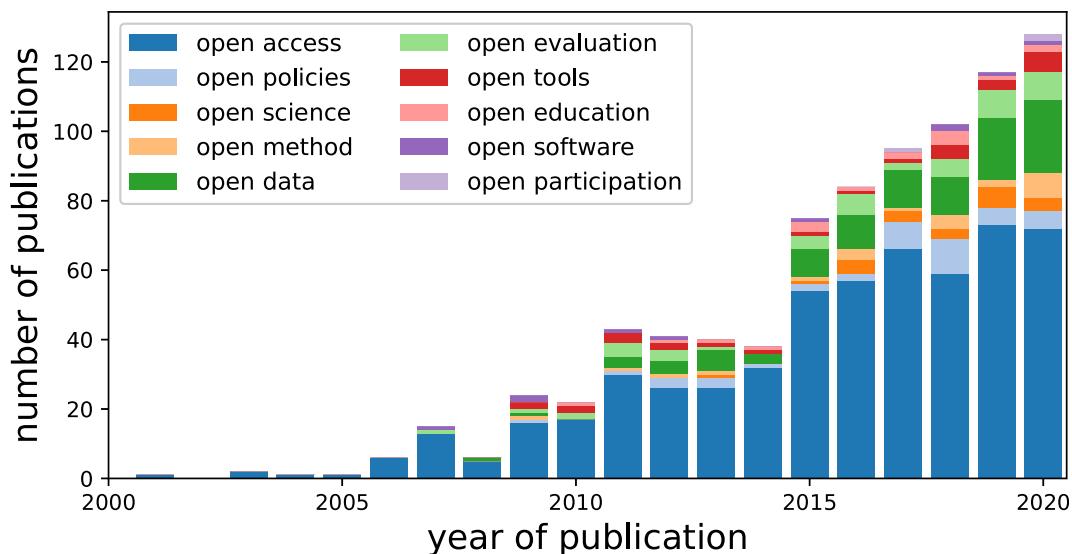


Figure 2. Overview over the number of studies by Open Science subfield published between 2000 and 2020.

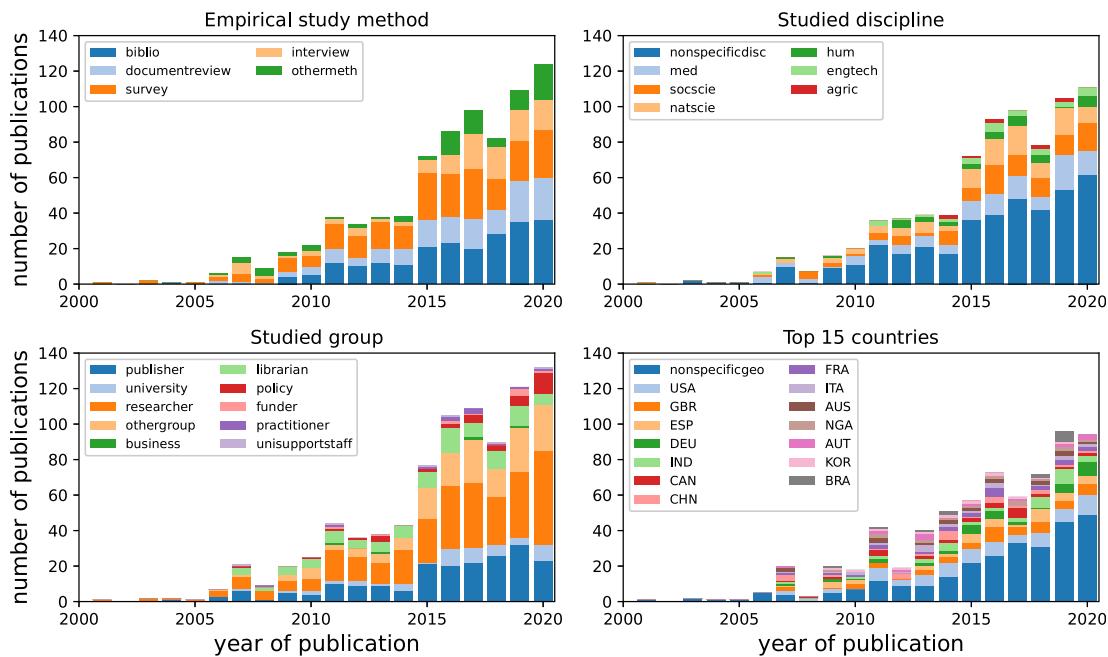


Figure 3. Overview of the number of studies by category published between 2000 and 2020. The top left panel shows the empirical study method, the top right panel shows the studied discipline, the bottom left panel shows the studied group, and the bottom right panel shows the distribution over the top 15 countries present in the dataset.

Table 3. Summary of the percentage agreement of coders on the five categories of the codebook.

Category	min	max	median
Action	85.92	100.00	92.96
Method	80.28	91.55	83.10
Discipline	74.65	92.96	90.85
Group	77.46	98.59	91.55
Geo scope	84.51	100.00	100.00

publications from determined journals. Here, some coders labeled “geo=None” or “discipline=None” as they did not deviate geographical or disciplinary focus from a journal sample. Other coders annotated “all” to the categories for the same reason, *i.e.* the journal sample does not deliberately limit geo or discipline in any way.

We calculated the percent agreement for each of the 36 labels from the five categories that were double coded. In this section we are only reporting a summary of the agreement (see Table 3); details for the data transformation, recoding, and results are reported in the documentation (see file “reliability.html” in the code repository; Lasser & Schneider, 2022).

Risk of bias and limitations

The following limitations should be considered with any use of the data set: despite the snowball search, which led to relevant results for the mapping review, we only did the systematic search in two databases, due to time constraints. As Web of Science and Scopus do not include all research literature and are biased towards specific criteria like publications (journal articles), languages (English-focused), and journals that are published in the United States, we lack other relevant peer-reviewed publications not covered in the two databases. We did not explicitly search for further gray literature to complement the results from the database search, therefore our data set may be susceptible to publication bias. Furthermore, in the inclusion criteria, we specify English, German, Italian, French or Spanish as the languages of publications due to the language skills of the authors and coders involved in our study. This systematically excludes publications in other languages and thus regions investigated. Also, the terms used in the search query were not translated to German, Italian, French, Spanish. Therefore, the database search only returned publications in these languages if an abstract or title was available in English. We invite native speakers of other languages to apply the selection criteria and coding system to other databases and searches in their language and thus contribute to the expansion of the data set.

Long-term data maintenance plans

The current review has the character of a pilot study, which we will build on. Three long term data maintenance plans are currently developed: first, annual data will be added following the year 2020 using the same selection criteria, coding and databases to keep the data and its value for research, teaching and science policy up to date, and to follow empirical research trends on Open Science practices. Second, we currently plan to include comparable data on literature about open educational resources and inclusive science practices such as citizen science or transdisciplinary approaches. Hence, the data will be expanded to further Open Science practices. Third, as a midterm goal - dashboard with visual analytical features will be programmed to allow for immediate usability of the data and to showcase the scoping efforts to a broader public.

Data and software availability

Underlying data

Zenodo: MapOSR - A Mapping Review Dataset of Empirical Studies on Open Science, <https://doi.org/10.5281/zenodo.6491891> (Lasser *et al.*, 2022)

This project contains the following underlying data:

- mapOSR_data_V5_9_3_220419_coded_clean.csv
- mapOSR_codebook_V4.csv
- mapOSR_references.bib
- mapOSR_interrater_reliability_clean.csv
- mapOSR_info.csv

Extended data

PRISMA checklist and flow chart for "MapOSR - A Systematic Mapping Review of Empirical Studies on Open Science" are deposited on Zenodo: <https://doi.org/10.5281/zenodo.6491891>

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Analysis code available from: <https://github.com/JanaLasser/mapping-open-science-research/tree/v1.0>

Archived analysis code as at time of publication: <https://doi.org/10.5281/zenodo.6491829>

License: MIT

References

-
- Bartling S, Friesike S: **Towards Another Scientific Revolution**. Bartling S, Friesike S, editors. *Opening science*. New York: Springer Verlag; 2014; pp. 3-15.
[Publisher Full Text](#)
- Blümel C, Beng F: **Open Science – a loosely coupled discourse? Comparing Open Science and Open Innovation from a bibliometric point of view**. *STI 2018 Conference Proceedings*. 2018; 369-377.
- Christensen G, Wang Z, Levy Paluck E, *et al.*: **Open science practices are on the rise: The state of social science (3s) survey**. 2020.
[Reference Source](#)
- Gough D, Oliver S, Thomas J: **Introducing Systematic Reviews**. Gough D, Oliver S, Thomas J editors. *An Introduction to Systematic Reviews*. 2nd ed. London: Sage; 2017; pp. 1-17.
- Grant MJ, Booth A: **A typology of reviews: an analysis of 14 review types and associated methodologies**. *Health Inf. Libr. J.* 2009; 26(2): 91-108.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lasser J, Schneider J: **mapOSR dataset 2000-2020 (v1.0)**. 2022.
[Publisher Full Text](#)
- Lasser J, Schneider J, Lösch T, *et al.*: **MapOSR - A Mapping Review Dataset of Empirical Studies on Open Science (V5_9_3_220419) [Data set]**. Zenodo. 2022.
[Publisher Full Text](#)
- Levin N, Leonelli S, Weckowska D, *et al.*: **How do scientists define openness? Exploring the relationship between open science policies and research practice**. *Bull. Sci. Technol. Soc.* 2016; 36(2): 128-141.
[PubMed Abstract](#) | [Publisher Full Text](#)
- OSCD Frascati Manual: *In The Measurement of Scientific, Technological and Innovation Activities*. OECD; 2015.
[Publisher Full Text](#)
- Page MJ, McKenzie JE, Bossuyt PM, *et al.*: **The PRISMA 2020 statement: an updated guideline for reporting systematic reviews**. *Int. J. Surg.*

2021; **88**: 105906.

[PubMed Abstract](#) | [Publisher Full Text](#)

Scanlon E: Scholarship in the digital age: Open educational resources, publication and public engagement. *Br. J. Educ. Technol.* 2013; **45**(1):

12–23. Wiley.

[Publisher Full Text](#)

Xu S, Lorber MF: Interrater agreement statistics with skewed data:

Evaluation of alternatives to Cohen's kappa. *J. Consult. Clin. Psychol.*

2014; **82**(6): 1219–1227.

[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 28 November 2022

<https://doi.org/10.5256/f1000research.133556.r154173>

© 2022 Cox A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Andrew M. Cox

Information School, University of Sheffield, Sheffield, UK

The paper explains the creation of a bibliometric dataset of peer reviewed publications about open science for 2000-2020. Exclusion of OER explained. Keywords used explained (I wonder if "open research" should have been a term?) A number of limits on the dataset are appropriately acknowledged, namely: limited range of databases searched, languages covered and exclusion of gray literature.

It would have been useful to discuss the implications of the database and language limitations. Given the quantity of material published in Chinese excluding material from this region is significant. Open access journals in South America would not be included in the search. Given the distinctive traditions around open science in that region this is again an issue. As an area of practice/policy I think including gray literature would have been helpful in increasing the value of the dataset.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Open data, research data management, open science from an information

science perspective. I am not a scientometrician.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 18 November 2022

<https://doi.org/10.5256/f1000research.133556.r154172>

© 2022 Šimukovič E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Elena Šimukovič

ZHAW Zurich University of Applied Sciences, Winterthur, Switzerland

The aim of this data note, as outlined in the Methods section, was "to identify all empirical studies concerned with Open Science or any of its key elements", which is "to be used as a basis for deeper investigations". For this purpose, the authors compiled a list of peer-reviewed publications from two commercial databases and publicly shared the results of their collaborative coding exercise.

In summary, this approach tackles an important gap that might be useful for research and teaching alike, as well as for related policy work. However, in my view, the broad aims as announced in the Introduction section (e.g. "to clarify the current understanding of Open Science") are somewhat overestimated. In particular, this applies to the authors' choice to limit their literature search to a few elements of Open Science only (i.e. Open Data, Open Access, and Open Source) and to include only peer-reviewed publications from two particular databases. As the authors state themselves, there are plenty of risks and limitations related to the chosen approach, especially with regard to inherent biases inscribed in these methods. Since Open Science is a highly relevant topic to various practitioners (incl. professionals at libraries, funding bodies, publishers, science policy-makers, research consulting firms, etc.), which have produced a huge number of reports on related issues themselves, excluding grey literature entirely appears to me as highly problematic.

Because this contribution is meant as a "data note" and not a full research article (or other type of publication), it satisfies the requirements imposed on this type of work. However, I would suggest that the authors bring their announced (very broad) objectives into better accordance with the (necessarily limited) outcomes in their data note and related data sets. Also, the list of keywords might be misleading, since only a few of the named Open Science sub-areas are dealt with in more detail.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Open Access, Open Science, research data management, Open Research Data, Science and Technology Studies, science policy, scholarly communication

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 17 November 2022

<https://doi.org/10.5256/f1000research.133556.r154176>

© 2022 de Haan J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Judith J. de Haan

Open Science Programme, Utrecht University, Utrecht, The Netherlands

This article describes a mapping of literature on various open science practices in the period of 2000 to 2020. There has been a lot of attention and progress in the open science movement, but there is no clear overview of the empirical studies that support open science practices. The authors of this paper made the first start by creating such an overview.

The strength of this paper is that they are trying to create a mapping of the empirical studies as a starting point for others to work on further and to add other papers as well. Though the dataset is now available in csv format (accessible and reusable, which is great), it would be great if this would be available in a workable, open way. Like in a living library, where people can add and annotate papers, to start working on a richer and more complete overview. And more up-to-date because the newest studies can immediately be added as well. An example of such a living library can be found here: <https://living-library-uu.web.app/>. This is a prototype from the Freudenthal Institute. The manuscript of this library will be shared soon, as the open source code on github. For questions in the meantime, livinglibrary@uu.nl can be used.

Open science contains a lot of different practices and the authors state that they focus on only a part of these practices: open publications, open data, and open source. It could be described clearer which practices are included and which are not. In the search string in Table 1, it also

includes 'open peer review' and 'open material'. On the other hand, it does not include the terms 'open source', 'open software', or 'FAIR data', which I think are also relevant if you want to search for open source and open data. The selection of the search terms is not clearly defined, this could be defined better and also state in the title and introduction better what the search includes. This is not about open science but about a selection of open science practices.

The authors also include a search block to narrow down the number of papers, because it would not be possible to go through that many papers. This does mean that it probably will miss studies on open science.

Some small remarks, mostly about the extra explanation of the selection that was used for this search and how this paper was written:

- Why online peer-reviewed? What about preprints, PCI, or preregistrations?
- In the introduction it is written: 'Open Science is still only vaguely defined', maybe broadly defined would be a better term.
- The goal of open science is not only to be more open and transparent, but this is a means to an end - improve the quality of research and improve the impact of the research.
- The explanation for no preregistration is a bit limited. You can always still (pre)register your study design and follow-up steps, and state that you started data collection already.
- Explain better what is in this review, no OER, no citizen science, but also no public engagement, no open materials, no pre-prints, no pre-registration, no open proposals
- Why is research from the industry removed? Mention this earlier and explain why.
- Can you specify the articles that were excluded (N=... for language, N=... not empirical data)
- How many disagreements were there in the selection part?
- 'We invite native speakers of other languages to apply the selection criteria and coding system to other databases and searches in their language and thus contribute to the expansion of the data set' - I think this is an important sentence. This study can become impactful if others contribute and it will be a living, annotated dataset.
- 'The current review has the character of a pilot study, which we will build on.' - Mention this earlier. Relevant for the reader.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Partly

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: Part of the steering committee of a preregistration platform for animal research, preclinicaltrials.eu

Reviewer Expertise: Open science, preregistration, systematic reviews

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 08 July 2022

<https://doi.org/10.5256/f1000research.133556.r138248>

© 2022 Mayer K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Katja M. Mayer 

University of Vienna, Vienna, Austria

The "data note" assessed here is the first to provide an open annotated literature corpus that provides an overview of empirical studies on open science, specifically on open access, open data, and open source. The dataset has annotated references to 695 studies from 2000 to 2020 found through Web of Science searches, Scopus, and snowballing in the OS community. The rationale for creating the dataset is clearly described, protocols are appropriate and overall the work is technically sound. Datasets are in a useable and accessible format. The data paper describes in detail not only the dataset but also the annotation and category work, as well as the data validation processes. This is particularly noteworthy, as not common, and very helpful not only to reuse the dataset but also to learn about the process. The document is, therefore, also excellently suited for teaching. It is already apparent that I rate the paper and the dataset very highly in any case. Both also score well with regard to the dynamics of the field that is described: due to the excellent preparation for re-usability, as well as transparency of the production conditions, the dataset can be continued in versions and forms, a good starting point for an important repository of basic knowledge for further research.

It would be very interesting to know how the results of the search for studies differ between the indices and the answers from the community: were there significant differences along the five categories in terms of what could be found?

Regarding the problem that perhaps more such studies can be found in the grey literature area: what outlook on this would the authors allow themselves on the basis of their experience? After

all, Open Science is precisely about opening up the boundaries of traditional scientific communication channels, so a few more sentences on this would certainly be in order. The discussion of this dataset and papers could also contribute to the development of a metadata standard for precisely the recording of such "grey literature".

Perhaps the article - although understandably not wanting to get into a discussion of the concepts around Open Science - could briefly note with regard to the search activities that it would probably also be possible to take up other terms such as "data sharing", "e-infrastructures" etc., especially with regard to the time before the widespread use of Open Science terms. Since this would certainly require a lot of additional effort, this dimension could be thought of in a further study or as an extension of the dataset in the future.

The Zotero list is quite wonderful, but it would be even more awesome if the 695 studies were tagged as such, and also the distinction of whether it was found from an index or through community snowballing could be tagged there. The Zotero Online group could then adopt these tags as well.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: STS, critical data studies

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research