

Adoption of Open Science Practices in Psychology and Education Research

ANONYMOUS AUTHOR(S)

Open science practices like data and code availability are (what exactly). To what extent do researcher actually implement OSF-practices in their publications? Do institutions explicitly value these principles in their hiring process and research conduct?

We examine open science practices within the fields of Psychology and Education. Specifically, we analyse the frequency of open science statements in [N] published articles alongside the prevalence of open science *expectations and requirements* in job advertisements for professorships at German universities. We found that bla ...

1 Introduction

2 Analysis

rieser: Should we provide a sample of keywords here?

2.1 Job postings for professorships

We analyzed job postings for professorships from the *Academic Job Posting Service of the German Association of University Professors and Lecturers* [1] published between January 2024 and [tbd.] 2025. Out of a total of $N_A = 295$ postings, 18 of them explicitly mentioned keywords related to open science. While all 18 described open science as desirable, only [k=6] required a formal statement. Exactly 1 was a specifically created "W2 Open Science" professorship. We identified these postings by matching keywords, their mangled variants, and partial sentences and extracting their containing sentences; the final classification however was based on manual review of the corresponding paragraphs.

rieser: double check this

rieser: I need to add to the code here to make sure that all keywords are hit and all of them provide a paragraph. Else the desirable <-> requirement stat might be wrong.

2.2 Publications

We analysed [N=?] open access publications from the open-alex index [3]. If there was an open access pdf available, we downloaded it using the Unpaywall service. We used the open alex API to build an index of papers, which we filtered for pdf urls. We used a batch service downloading batches of [k=?] and then processing these batches using *grobid* [2] to obtain a structured text document.

3 Implementation

We extracted all *works* from the OpenAlex [3] corpus that were tagged as belonging to the fields Psychology or Education. From these records, we retained the ones that included at least one linked open access PDF. To manage this dataset, we implemented a database consisting of two tables: an **index** table containing the metadata for each entry, and a **pdfs** table storing information about the retrieved documents.

rieser: explain what this means or phrase differently

rieser: This is not done yet

Because of hardware limitations and the large volume of file , we implemented a batched processing pipeline. This pipeline retrieves and processes PDFs in batches of size $k = 200$, so we only have to store at most k PDFs simultaneously. After processing, the results are joined onto the bibliographic data from the index table for further analysis.

rieser: How many GBs?

4 Related Work

We could mention dataset here, even if we adapt it. Otherwise, fexrep [5].

5 Conclusion

Data Availability Statement

Our code, proofs and evaluations are available on GitHub [4] or on osf or on zenodo??.

References

[1] [n. d.]. Academic Job Posting Service. <https://www.hochschulverband.de/en/services/academic-job-posting-service>
[2] 2008–2025. GROBID. <https://github.com/kermitt2/grobid>. swb:1:dir:dab86b296e3c3216e2241968f0d63b68e8209d3c
[3] Jason Priem, Heather Piwowar, and Richard Orr. 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. doi:10.48550/arXiv.2205.01833 arXiv:2205.01833 [cs].
[4] Jeremias Rieser. 2025. Open Science Mentions. <https://github.com/jkr11/open-science-mentions>.
[5] Jian Wu, Rajal Nivargi, Sree Sai Teja Lanka, Arjun Manoj Menon, Sai Ajay Modukuri, Nishanth Nakshatri, Xin Wei, Zhuoer Wang, James Caverlee, Sarah M. Rajtmajer, and C. Lee Giles. 2021. Predicting the Reproducibility of Social and Behavioral Science Papers Using Supervised Learning Models. doi:10.48550/arXiv.2104.04580 arXiv:2104.04580 [cs].