Tracking Data Sharing in Publications from Major Funders

EBRAINS Summit 2025

Christoph Li[1], Josh Lawrimore[2], Dustin Moraczewski[1], Adam Thomas[1]

[1]Data Science & Sharing Team, National Institute of Mental Health, [2]Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research.

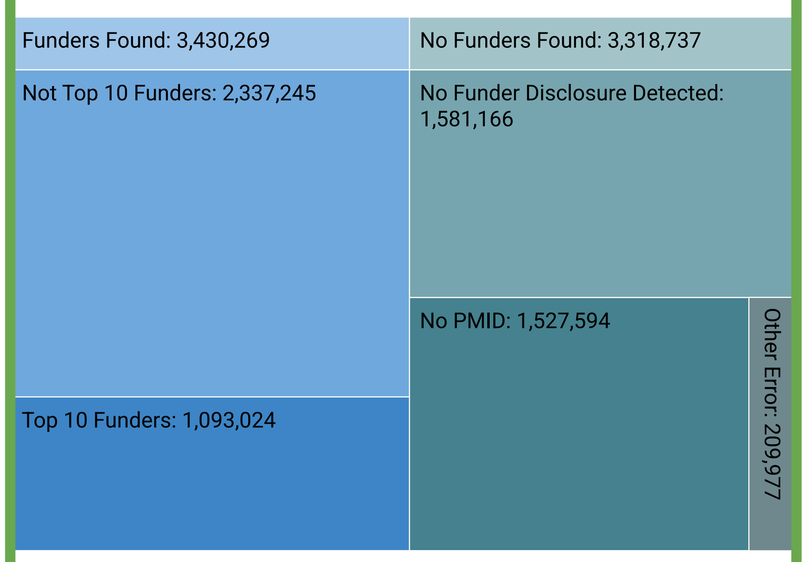
INTRODUCTION

Data sharing and transparency has been a priority for NIH and other major funders of biomedical science for over 25 years. However, bibliometric analyses show that a minority of publications report how underlying data can be obtained[1].

In this work we explore differences in data sharing reporting in PubMed Open Access publications across the ten largest funders of biomedical science in the world as well as within NIH. Results are presented here as well as on an online interactive dashboard that we intend to maintain and update.

METHODS

1. Downloaded 6.7 million XMLs from PubMed Central Open Access (PMC OA)
2. Ran data through *rtransparent* pipeline[1]
3. Identified publications acknowledging funding from one of ten largest funders of biomedical science
4. Built interactive dashboard
   a. All materials are available under CC0 license in the NIMH-DSST Open Science Metrics repository[2].

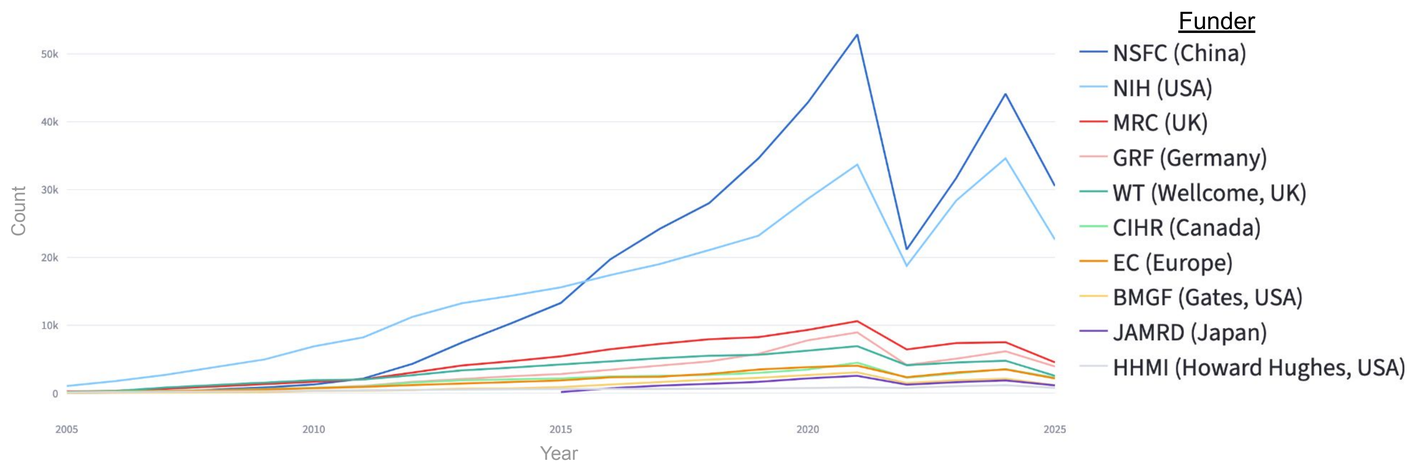


Distributions of funding information in 6.7M PMCOA publications
Of 3.43M with funder info, 32% were supported by the top 10 funders (1.09M) and 68% by others (2.34M). Of 3.32M with no funders, 48% had no funding disclosure (1.58M), 46% lacked a PMID (1.53M), and 6% failed in the *rtansparent* pipeline (209K).

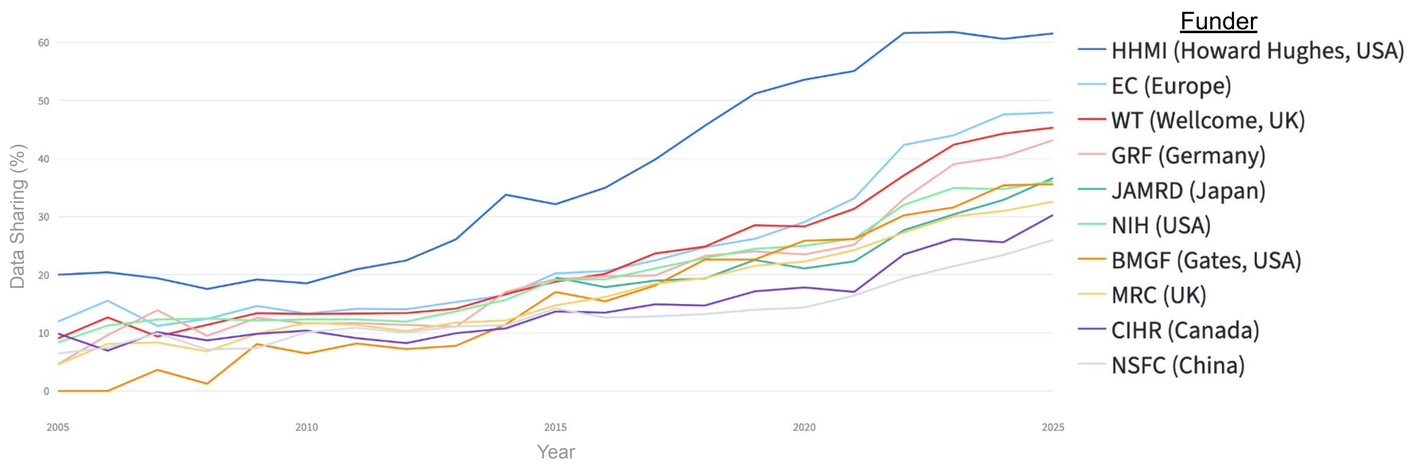
TAKEAWAYS

- Consistent difference in rates of data sharing exist between different funders.
- We hypothesize that funders' policies and practices influence grantees likelihood of data sharing

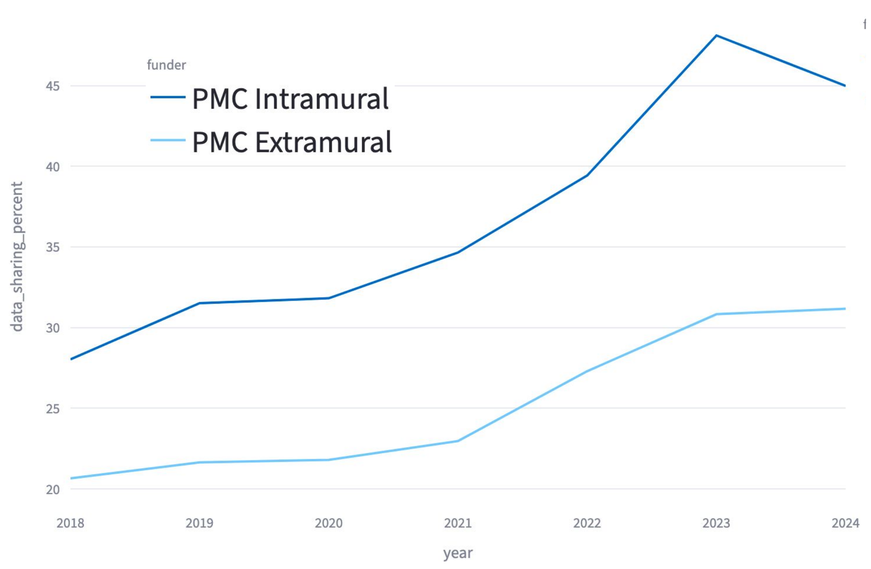
RESULTS



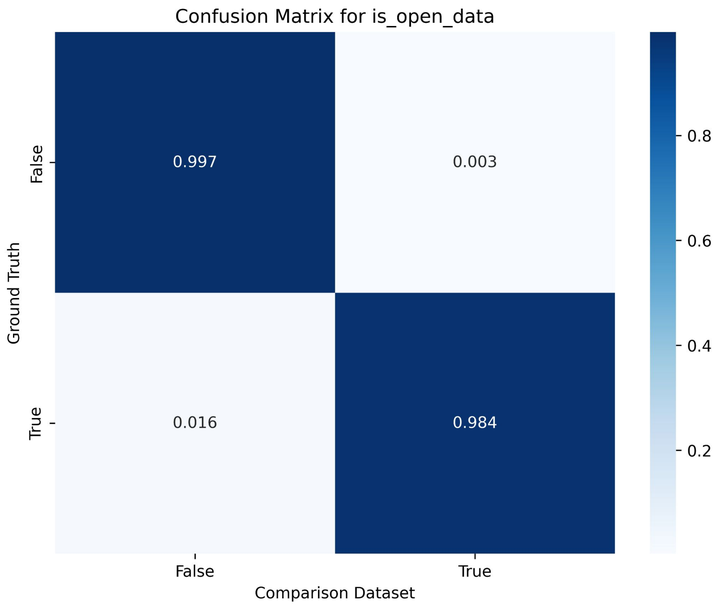
Total number of publications over time among the top ten biomedical funders, 2005-2025
Each line represents the number of publications in the PubMed Open Access subset from a given funder.



Data sharing rates over time among the top ten biomedical funders, 2005-2025
Each line represents the percentage of publications in the PubMed Open Access subset from a given funder in which information regarding how underlying data can be accessed was detected.



Trends in data sharing among NIH intramural and extramural publications, 2018–2024
PMC OA publications listed in annual reports from NIH intramural groups showed 10-20% higher rates of data sharing reporting than other PMC OA publications that acknowledged (extramural) NIH funding.



Open data confusion matrix; this study against Serghiou et al. 2021
For publications before 2021, our labelling showed high overlap with results from Serghiou et al. 2021

REFERENCES

[1] Serghiou, S.; Contopoulos-Ioannidis, D. G.; Boyack, K. W.; Riedel, N.; Wallach, J. D.; Ioannidis, J. P. A. Assessment of Transparency Indicators across the Biomedical Literature: How Open Is Open? PLoS Biol. 2021,19(3), e3001107.
[2] NIMH-DSST. GitHub - nimh-dsst/osm: OpenSciMetrics (OSM) applies NLP and LLM-based metrics and indicators related to transparency, data sharing, rigor, and open science on biomedical publications. GitHub. https://github.com/nimh-dsst/osm/
[3] PMC OA FTP Service - PMC. https://pmc.ncbi.nlm.nih.gov/tools/ftp/

DISCUSSION

- Our analysis suggests that funding source has significant impact on the rates of data sharing reporting in publications.
- Some funders who do not promote data sharing (NSFC) show low levels while other vocal advocates of data sharing show median levels (e.g. NIH, EC, Wellcome).
- NIH Intramural publications have a 10-20% higher rate of data sharing, between 2018 and 2024.

FUTURE

- Investigate why HHMI has a higher rate of data sharing.
- Compare data sharing in publications from HHMI Janelia Research Campus and its extramural program.

National Institute of Mental Health

EBRAINS 2.0 has received funding from the European Union's Research and Innovation Program Horizon Europe under Grant Agreement No. 101147319.

This project was funded by the Intramural Research Program of the NIMH, grant numbers ZICMH002960. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. 75N91019D00024. The content of this publication does not necessarily reflect the views of policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.