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

Funders throughout the world have sought to increase data sharing and transparency from their grantees using different approaches and incentives. In this project we compare the relative frequency of data sharing between major funders of biomedical science as measured using text mining in the full text from 6.7M publications available in the PubMed Open Access (PMCOA) collection.

METHODS

- 6.7 million publications from PubMed Central's Open Access (PMCOA) June 2025 baseline (comm & noncomm licensing) were processed with rtransparent (*1*)
- Non-research articles types (e.g. review, errata, commentary) were excluded leaving 5.5M pubs which were processed with OddPub v7.2.3 (*2*)
- 375K research articles with open data (as determined by OddPub) were identified, approximately 6.8%
- 57 canonical funder names were extracted from these open-data research articles using Named Entity Recognition (NER) & a 4σ statistical threshold.
- The total number of publications both with and without open data for each identified funder was counted across publication years 2010 to 2024 and used to calculate year-wise percentages of publications with open data.
- All code used for this analysis is available under CC0 license on GitHub (*3*).
- These results are also available via an online dashboard at https://opensciencemetrics.org
- Claude Code was used extensively throughout the analysis

DISCUSSION

- The current version of OddPub (v7.2.3) provides a lower and more accurate (*4*) estimate of the rate of data sharing in biomedical literature than earlier versions and previous reports (6.8% vs. 15%) (*1*),
- Recent mandates for data availability statements have decreased the sensitivity of tools like OddPub (*5*)
- Our result can be used by funders to better understand what policies best incentivize data sharing.

REFERENCES

1. S. Serghiou *et al.*, *PLoS Biol.* **19**, e3001107, DOI 10.1371/journal.pbio.3001107 (Mar. 2021).
2. N. Riedel *et al.*, *oddpub: Detection of Open Data & Open Code statements in biomedical publications*, 2025, DOI 10.5281/ZENODO.3741403.
3. A. Thomas *et al.*, *github.com/nimh-dsst/osm-2025-12-poster-incf: Code and details for Open Science Metrics poster presented at EBRAINS / INCF December 11, 2025*, en.
4. D. G. Hamilton *et al.*, en, *BMJ* **382**, e075767, DOI 10.1136/bmj-2023-075767 (July 2023).
5. A. Iarkaeva *et al.*, en, *PLoS One* **19**, e0302787, DOI 10.1371/journal.pone.0302787 (May 2024).

TAKEAWAYS

- We demonstrate that data sharing rates vary consistently by funder.
- Our analysis can help funders learn what policies and practices are most effective in incentivizing data sharing from their grantees

RESULTS

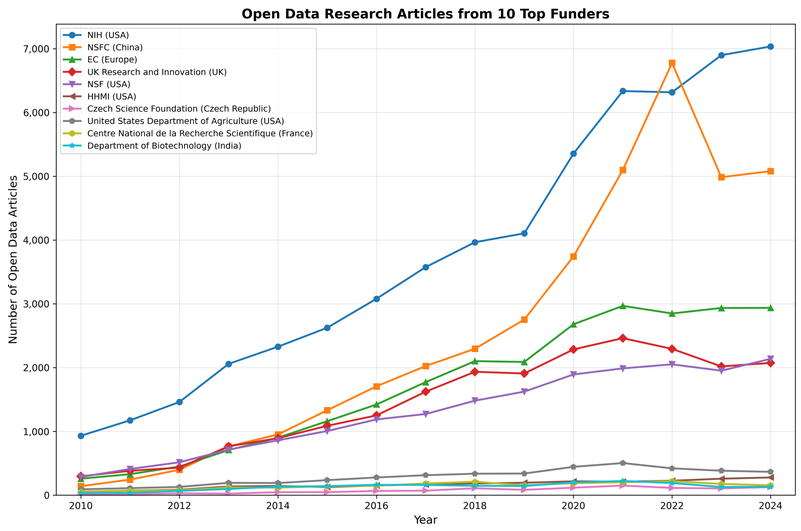

Figure 1: Total number of research articles with open data among 10 biomedical funders 2010–2024.

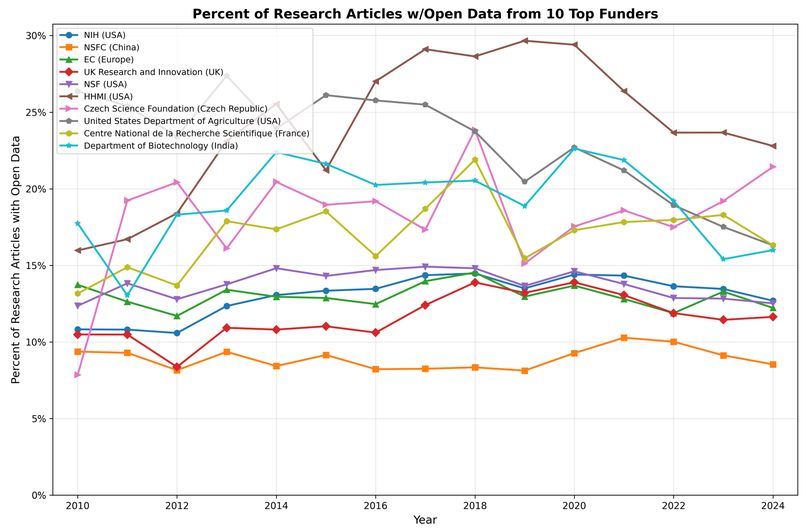

Figure 2: Percentage of research articles with open data among 10 top funders 2010–2024.

Funder Name	Country	Research Pubs 2010–2024	Pubs w/ Data Sharing	% Pubs w/ Data Sharing
Howard Hughes Medical Institute	USA	10,189	2,503	24.6%
United States Department of Agriculture	USA	20,060	4,328	21.6%
Department of Biotechnology	India	10,188	1,990	19.5%
Czech Science Foundation	Czech Republic	5,897	1,103	18.7%
Department of Energy	USA	22,368	4,011	17.9%
Agence Nationale de la Recherche	France	30,382	5,304	17.5%
Max Planck Society	Germany	10,143	1,760	17.4%
Centre National de la Recherche Scientifique	France	12,729	2,207	17.3%
Royal Society	UK	12,652	1,929	15.2%
Wellcome Trust	UK	68,683	9,873	14.4%
Associazione Italiana per la Ricerca sul Cancro	Italy	7,641	1,080	14.1%
Swiss National Science Foundation	Switzerland	24,562	3,431	14.0%
Research Foundation Flanders	Belgium	9,776	1,344	13.8%
National Science Foundation	USA	141,132	19,368	13.7%
Natural Sciences and Engineering Research Council of Canada	Canada	28,642	3,895	13.6%
Australian Research Council	Australia	26,111	3,511	13.4%
National Institutes of Health	USA	426,267	57,220	13.4%
Austrian Science Fund	Austria	15,381	2,014	13.1%
European Commission	EU	196,678	25,542	13.0%
Deutsche Forschungsgemeinschaft	Germany	82,116	10,660	13.0%
Russian Science Foundation	Russia	11,734	1,506	12.8%
Netherlands Organisation for Scientific Research	Netherlands	17,223	2,149	12.5%
Japan Agency for Medical Research and Development	Japan	15,385	1,872	12.2%
UK Research and Innovation	UK	179,659	21,694	12.1%
Ministerio de Economia y Competitividad	Spain	29,285	3,490	11.9%
Swedish Research Council	Sweden	29,409	3,486	11.8%
Fundacao de Amparo a Pesquisa do Estado de Sao Paulo	Brazil	19,810	2,335	11.8%
National Key Research and Development Program	China	48,760	5,732	11.8%
Bundesministerium für Bildung und Forschung	Germany	26,596	3,091	11.6%
Ministry of Education, Culture, Sports, Science and Technology	Japan	28,406	3,270	11.5%
China Scholarship Council	China	15,005	1,708	11.4%
European Regional Development Fund	EU	56,374	6,396	11.3%
Japan Science and Technology Agency	Japan	13,996	1,570	11.2%
Fundacao para a Ciencia e a Tecnologia	Portugal	17,108	1,844	10.8%
National Science Centre	Poland	14,256	1,528	10.7%
American Heart Association	USA	11,428	1,204	10.5%
Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior	Brazil	29,668	3,123	10.5%
Italian Ministry	Italy	19,005	1,947	10.2%
Japan Society for the Promotion of Science	Japan	66,236	6,258	9.4%
China Postdoctoral Science Foundation	China	24,691	2,328	9.4%
National Natural Science Foundation of China	China	420,824	38,274	9.1%
Bill and Melinda Gates Foundation	USA	22,670	2,046	9.0%
Ministry of Science and Technology	Taiwan	57,698	5,186	9.0%
National Health and Medical Research Council	Australia	30,425	2,667	8.8%
Canadian Institutes of Health Research	Canada	35,617	2,908	8.2%
National Research Foundation of Korea	Korea	84,068	5,625	6.7%
National Institute for Health Research	UK	47,528	2,865	6.0%

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.

10.5281/zenodo.NotYetRegistered