UCLA School of Education & Information Studies

School of Information Sciences

Revealing Contributions on Research Data the Potential of Data Paper

Tianji Jiang — School of Education and Information Studies, UCLA

Shuhan Ai — School of Education and Information Studies, UCLA

Kai Li — School of Information Sciences, University of Tennessee, Knoxville

The 2024 MET-STI Workshop

Nov 6, 2024

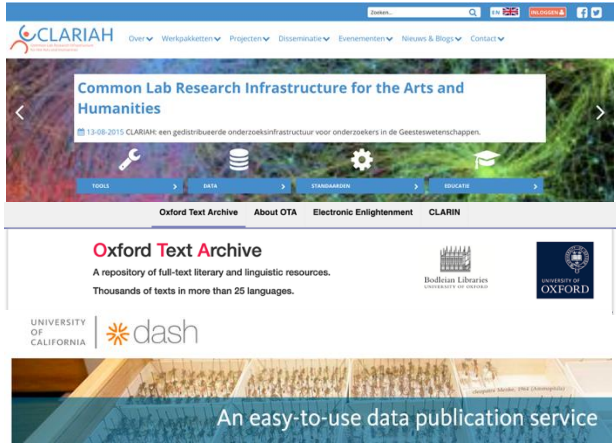
Background: Publishing Research Data

- Releasing data in a form that can be used by other individuals (Pasquetto et al., 2017).
- Link datasets to journal article or publication.
- Deposit datasets in a digital data archive.
- Publish data documentation (Borgman & Bourne, 2022).
  - Research protocols
  - Codebooks
  - Software
  - Algorithms
- There are lots of benefits to share data
  - Increase efficiencies and reproducibility in scientific research
  - Maximize the use of resources and investments
  - Allow for an expansion of innovation, escalating collaboration
  - Give researchers extra credits

DANS

HARVARD Dataverse

ICPSR A PARTNER IN SOCIAL SCIENCE RESEARCH | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH



Background: Publishing Research Data

- But there's no such thing as a free lunch …
  - Make research data publicly available also has challenges and costs.
  - Describe and publishing data is a costly proposition in and of itself.
  - **Work on data publishing fail to receive deserved recognitions in scholarship**.
- To encourage more researchers to publish and share their data, infrastructures should be established to ensure that researchers receive recognition for their contributions to data and data sharing.
- **Data paper** can be a resolution.

Background: Data Paper

- Data paper is a type of academic publication that focuses on describing and sharing research data rather than presenting new research findings.
- It primarily aims to make datasets accessible for other researchers to reuse, verify, or expand upon in future studies. It provides detailed descriptions of data collection, processing, and verification methods.
- Journals: *Scientific Data*, *Journal of Physical and Chemical Reference Data*, *Data in Brief, …*
- By publishing a data paper, the authors will:
  - Receive credit through indexing and citation of the published paper
  - Increase the visibility, usability and credibility of the data
  - Track the usage and citations of the data more effectively

Research questions

- How many authors publish a data paper after previously publishing a research paper related to the same project?
- What changes in publishing practices might occur for data-relevant contributors after they publish a data paper?

Research Design

- Data paper provides a unique approach to make observation;
- We resolve the question by comparing the author information of a data paper and its associated research paper;
- Data papers were retrieved from *Data in Brief* (published between 2014-2022), a multidisciplinary and peer-reviewed journal that publishes data papers and provides access to open research data;
- The research papers associated with the selected were also retrieved. Each data paper was associated with one research paper;
- The data used in this study were collected from Scopus and OpenAlex.

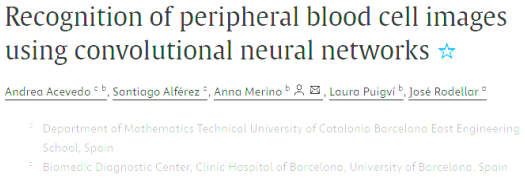
Research Design: Key Concepts

- Contributions to data are disclosed when the authorship of a data paper is different from its associated research paper. A key piece of information revealed is the identification of research contributors who were not authors of the associated research paper. We resolve the question by comparing the author information of a data paper and its associated research paper;


Data Article

A dataset of microscopic peripheral blood cell images for development of automatic recognition systems

Andrea Acevedo [c b], Anna Merino [a], Santiago Alférez [c], Ángel Molina [a], Laura Boldú [a], José Rodellar [c]


Recognition of peripheral blood cell images using convolutional neural networks ☆

Andrea Acevedo [c b], Santiago Alférez [c], Anna Merino [b], Laura Puigví [b], José Rodellar [a]

[c] Department of Mathematics Technical University of Catalonia Barcelona East Engineering School, Spain

[b] Biomedic Diagnostic Center, Clinic Hospital of Barcelona, University of Barcelona, Spain

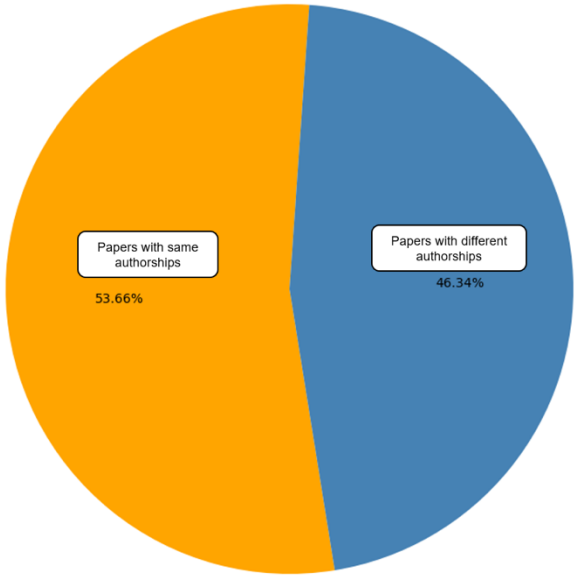
- **BA**: The researcher who share authorship in both the research paper and the data paper.
- **DA**: The researcher who share authorship in both the data paper.
- **RA**: The researcher who share authorship in both the research paper.

Preliminary results

- Among the **4,025** cases of data papers and associated research papers, excluding the **76** cases authored solely by the same person, **2,119** had the same authorship information, **290** had the same authors in a different order, **510** had the data paper removed at least one author from the research paper without adding new authors, and the remaining **1,030** cases had at least one new author in the data paper.

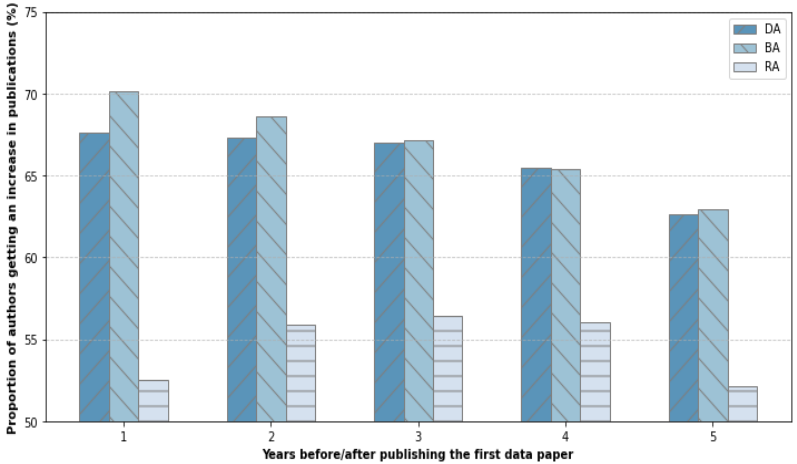
- We focused on the **1,030** cases that had at least one new author in the data paper in the analysis.

- We identified 9,147 authors, including 5,371 BAs, 2,025 RAs, and 1,734 DAs in the selected cases.



Preliminary results

- Approximately 65% of the DAs and BAs experienced an increase in their publication numbers after publishing the data paper. In contrast, only about 53% of the RAs, who did not publish the data paper as the DAs and BAs did, saw an increase within the same period.
- The average annual increase in publication numbers for DAs (1.18) and BAs (1.56) was significantly higher than that of the RAs .
- It indicates that publishing a data paper may encourage researchers to highlight their contributions and achievements permanently, particularly those related to data in scientific research. This, in turn, leads to an increased publication output over the long term.



Preliminary results

- We analyzed the publication history of authors who had at least one publication after publishing the data paper, and see the difference before and after these authors publish the data paper.
- Compared to the RAs, both the BAs and DAs had many more authors who had never published a data paper or dataset before the data paper in *Data in Brief*.
- For authors who had published a data paper or dataset before then, the BAs and DAs also showed a much higher increase in the publication rate.
- These findings indicate that authors who publish data papers tend to actively participate in opening and introducing their data. And after they published a data paper in Data in Brief, they tend to involve in publishing more data relevant academic publications in the future.

$$Increase = \frac{1}{E}\sum_{i=1}^{E}\frac{\left(\sum_{N=1}^{P} A_{ip} - B_{ip}\right)}{\sum_{N=1}^{P} B_{ip}}$$

Type	Publication	New	E	Increasing
DA	Article	219	1083	-0.155
	Data Paper	361	1043	0.673
	Dataset	10	373	0.909
BA	Article	398	3747	0.863
	Data Paper	396	3747	0.640
	Dataset	21	1334	2.533
RA	Article	194	1309	0.291
	Data Paper	6	122	0.187
	Dataset	18	422	0.271

Discussions

- As research activities become more sophisticated, research teams may include a variety of contributors beyond traditional academic roles, supporting specific aspects of a research project.
- Unlike academic workers such as faculty members or graduate students, these contributors usually do not participate in the paper writing process in projects and, of course, are not recognized through publications . It will be very interesting to know if data papers can include the researchers who usually don't share authorships for these reasons in a collaborative research project.
- We found that collaborative research groups tend to reassign authorship roles when disseminating a data paper within a considerable number of instances. This implies a tendency among researchers to reconsider and alter the order or presence of authors when transitioning from presenting original research to publishing the associated data.
- We also noticed that researchers tend to have more data relevant publications after they published a data paper, indicating that data paper can be a motivator for researchers to make their contributions to data visible through publications.
- Data paper may be an infrastructure to allow researchers receive recognition for their contributions to data and data sharing.

Conclusions

- Our work is an exploratory study to figure out investigating whether the researchers' publishing behaviors change after they publish their first data paper and examine how this affects their academic visibility.
- The finding indicates that publishing a data paper may encourage researchers to highlight their contributions and achievements permanently, particularly those related to data in scientific research. This, in turn, leads to an increased publication output over the long term.
- The enhanced visibility not only amplifies the impact of the project associated with the data paper but also benefits future work and the whole research community.

References

1.Ahn, C P Alexandroff, R Allende Prieto, C Anderson, S F Anderton, T Andrews, B H et al. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *Astrophysical Journal* 203: 21. https://doi.org/10.1088/0067-0049/203/2/21

2.Borgman, Christine L., & Bourne, Philip E. (2022). Why it takes a village to manage and share data. *Harvard Data Science Review*, *in press*. http://arxiv.org/abs/2109.01694

3.Culina, A., Crowther, T. W., Ramakers, J. J. C., Gienapp, P., & Visser, M. E. (2018). How to do meta-analysis of open datasets. *Nature Ecology & Evolution*, 2(7), 1053–1056. https://doi.org/10.1038/s41559-018-0579-2

4.Pasquetto, I., Randles, B., & Borgman, C. (2017). On the Reuse of Scientific Data. *Data Science Journal*, 16(0), 8. https://doi.org/10.5334/dsj-2017-008

5.Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. https://doi.org/10.7717/peerj.175

6.Popkin, G. (2019). Data sharing and how it can benefit your scientific career. *Nature*, 569(7756), 445–447. https://doi.org/10.1038/d41586-019-01506-x

7.Staunton, C., Barragán, C. A., Canali, S., Ho, C., Leonelli, S., Mayernik, M., Prainsack, B., & Wonkham, A. (2021). Open science, data sharing and solidarity: who benefits? *History and Philosophy of the Life Sciences*, 43(4), 115. https://doi.org/10.1007/s40656-021-00468-6

8.Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, 8(7), e67332. https://doi.org/10.1371/journal.pone.0067332

9.Walport, M., & Brest, P. (2011). Sharing research data to improve public health. *The Lancet*, 377(9765), 537–539. https://doi.org/10.1016/S0140-6736(10)62234-9

Q&A

Thank you for watching!