

Self-Reported Data Reuse Associated with Highly Cited Journal Articles

Heidi J Imker (presenter)

with Hoa Luong, Bill Mischo, Mary Schlembach, and Chris Wiley

University of Illinois at Urbana-Champaign

Observations from the field

- **University of Illinois at Urbana-Champaign**
 - ~2800 faculty & 16,000 graduate students
 - \$637M (US) in research expenditures
 - Very strong STEM
- **Research Data Management Professionals & STEM Librarians**
- **Researcher Reactions to Data Sharing**
 - Highly variable
 - Other reasons for not sharing or reusing data
 - Concept of data sharing seems abstract
- **Uneven adoption of data sharing principles and practices**

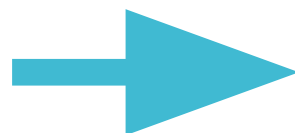
Research Motivations

- We wanted to hear from researchers across the spectrum
 - from enthusiastic to skeptical
- We wanted a frame data sharing/reuse in more concrete terms and ground it in something specific
 - highly cited articles for reflection
 - relevant in the context of publisher policies
- We wanted to incorporate the other reasons heard for not sharing/reusing data
 - data cannot (always) address other research questions

Survey A – Initial Authors

20 articles
per
discipline

Biochemistry
Chemistry
Computer Science
Earth Science
Engineering
Environmental Science
Materials Science
Medicine
Physics



Survey B – Citing Authors

all citing
articles from
Survey A
responses

Biochemistry
Chemistry
Computer Science
Earth Science
Engineering
Environmental Science
Materials Science
Medicine
Physics

Study Design

Survey A – Initial Authors

20 articles
per
discipline
(180 total)

Biochemistry
Chemistry
Computer Science
Earth Science
Engineering
Environmental Science
Materials Science
Medicine
Physics

48
= responses
(27%)

(A) For the research presented in highly cited articles, did the **initial authors reuse data themselves?** And if not, why not?

Survey A – Initial Authors

20 articles
per
discipline
(180 total)

Biochemistry
Chemistry
Computer Science
Earth Science
Engineering
Environmental Science
Materials Science
Medicine
Physics

48
= responses
(27%)

all citing
articles
(1698 total)

Survey B – Citing Authors

Biochemistry
Chemistry
Computer Science
Earth Science
Engineering
Environmental Science
Materials Science
Medicine
Physics

240
= responses
(14%)

(B) Did the citing authors reuse data associated with the initial article? And if not, why not?

Survey

A

Initial Authors

48 total responses

39 articles had associated data

35 answered if **reused data themselves**

14 Yes (40%)

21 No (60%)



*Number 1 reason
why not?*

12 (40%) – could not address the
research question

Survey

B

Citing Authors

240 total responses

240 answered if **reused initial data**

35 Yes (15%)

205 No (85%)



*Number 1 reason
why not?*

116 (58%) – could not address the
research question

Thought
Provoking
Comments

No, we did not reuse data.

"Our contribution was a new study. Why would we have used their data?"

Thought
Provoking
Comments

No, we did not reuse data.

"Data was publicly available and could be tailored to my analysis."

Thought
Provoking
Comments

Yes, we did reuse data.

"We routinely digitize data points from published figures."

Thought
Provoking
Comments

Yes, we did reuse data.

"This survey makes a wrong assumption. Data is used insofar as it refers to conclusions. My paper used other peoples' conclusions, drawn from those other peoples' data, but I had no need for their raw data."

Concluding Remarks

- A surprising lack of mutual interest in the underlying data & a major rationale seems to be that those data are not *useful* (vs. usable)
- A surprising presence of data sharing and reuse that do not “qualify” as such based on Open Science norms
- Preliminary – not confirmational
- Pause on generalizations

More Info



imker@illinois.edu

- **API Scripts (need access to Scopus API and to create a database)**
 - Mischo, William (2021): Scopus API Scripts for Data Reuse Project. University of Illinois at Urbana-Champaign.
 https://doi.org/10.13012/B2IDB-0988473_V1
- **Survey Data, Scripts, and Materials**
 - Imker, Heidi J; Luong, Hoa; Mischo, William H; Schlembach, Mary C; Wiley, Chris (2021): Data for: An Examination of Data Reuse Practices within Highly Cited Articles of Faculty at a Research University. University of Illinois at Urbana-Champaign.
 https://doi.org/10.13012/B2IDB-2087785_V1
- **Forthcoming Article**
 - Imker, H.J., Luong, H., Mischo, W.H., Schlembach, M.C., Wiley, C.A. 2021 An "Examination of Data Reuse Practices within Highly Cited Articles of Faculty at a Research University" *Journal of Academic Librarianship* (accepted)