

Data Citation Metrics: Are You Ready for Us?

CHORUS Forum: Are We Ready for Data Citation Metrics?

2021-10-28

Amy L. Nurnberger (she/hers)
MIT Libraries, Program Head, Data Management Services
Interim Head, Data & Specialized Services
Teachers College, Columbia University, Assistant Adjunct
Professor in Learning Analytics

@ANurnberger | ORCID: [0000-0002-5931-072X](https://orcid.org/0000-0002-5931-072X)

* The views expressed herein are those of the presenter and not necessarily her institution.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

**MIT
Libraries**

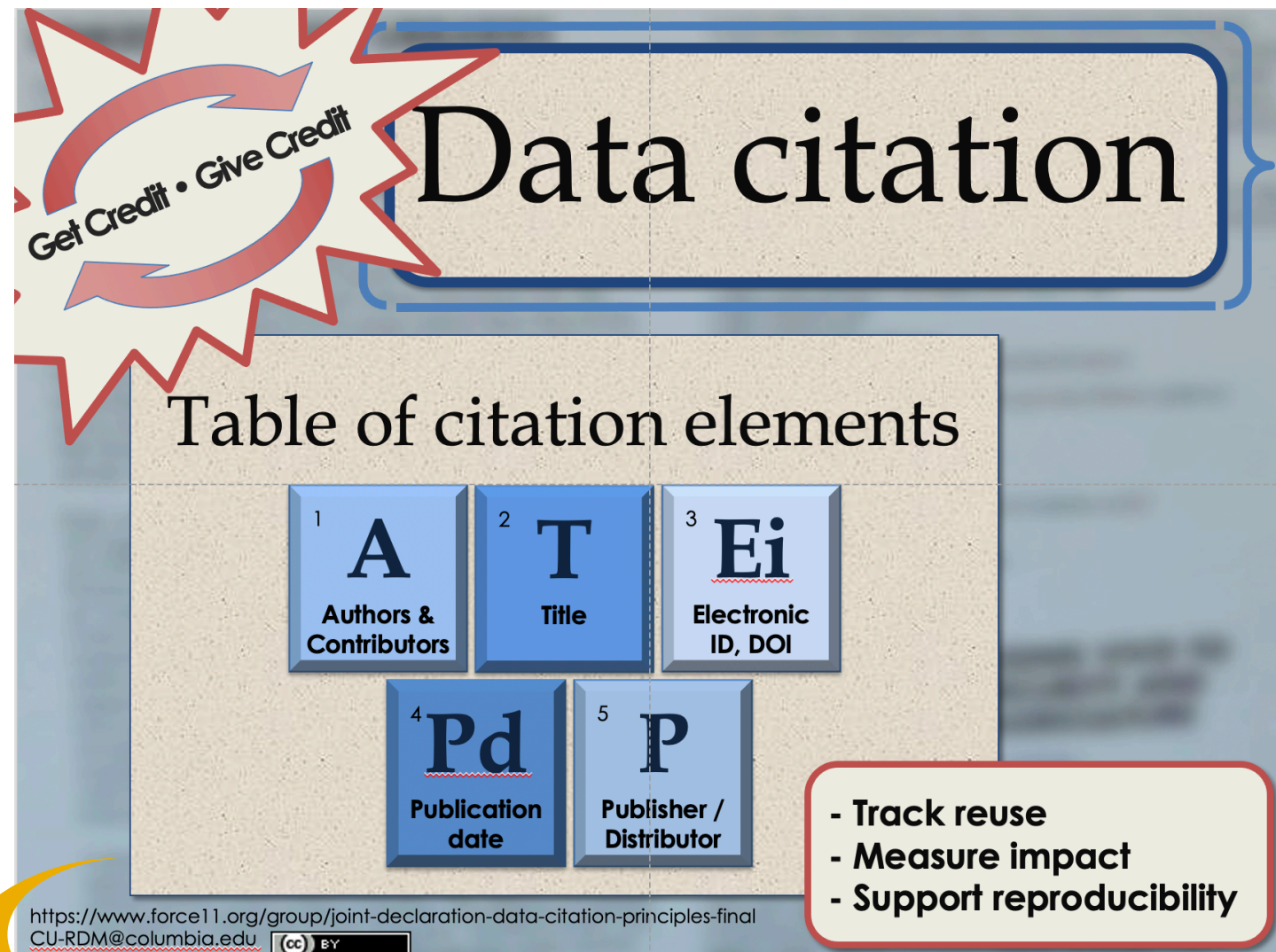
Step #1: Data Citation We've been waiting...

1985, NRC: “*Recommendation 12. Journals should require full credit and appropriate citations to original data collections...*” [1]

2007, Altman & King: “A critical component of this community is the common language of and the universal standards for scholarly citation, credit attribution, and the location and retrieval of articles and books. We propose a similar universal standard for citing quantitative data...” [2]

2012, Mooney & Newton: “the majority of journal articles failed to include an adequate citation to data used in secondary analysis studies” [3]

Just the other day, Data professional: “... require a PID and indexing if you want to change anything.” [4]



Slide I've been giving since ~2014.

What we have:

Including data availability statement (PLOS) [5]

Necessary, but insufficient

Appropriate citation of publicly available data sets (ACS) [6]

Closer!

What we need:

Joint Declaration of Data Citation Principles
(force11) [7]

That's the one!

Step #2: Metrics...we have a problem

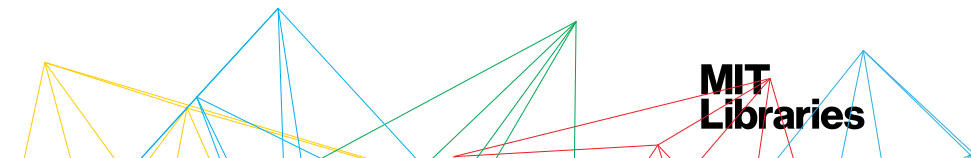
- misuse, misrepresentation, misinterpretation
- biases
 - inclusion and representation of diverse communities around the world in these metrics?
- inequities
 - accounting for geographic and social inequities in underlying infrastructures that support deployment of metrics?



Let's do better

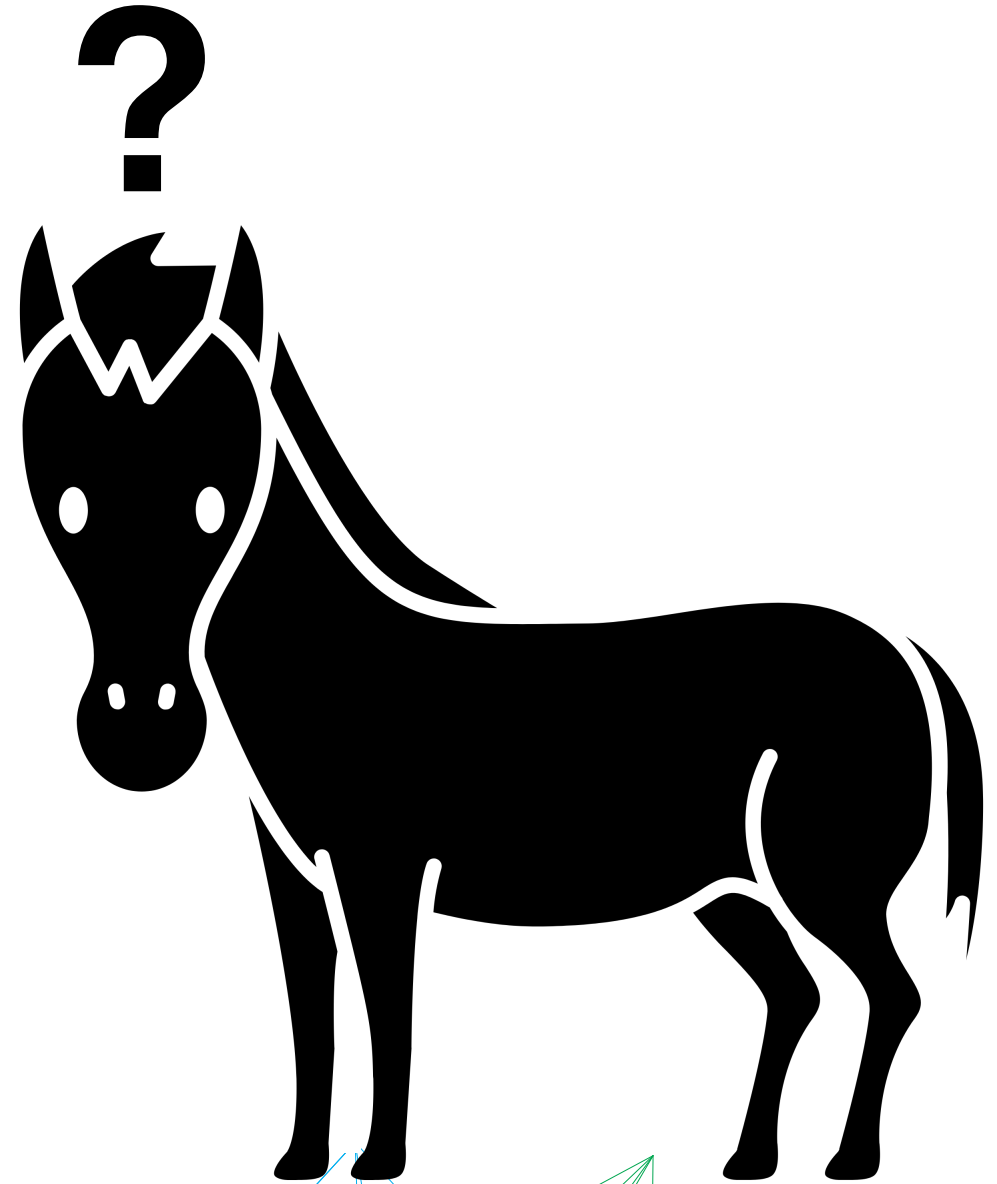
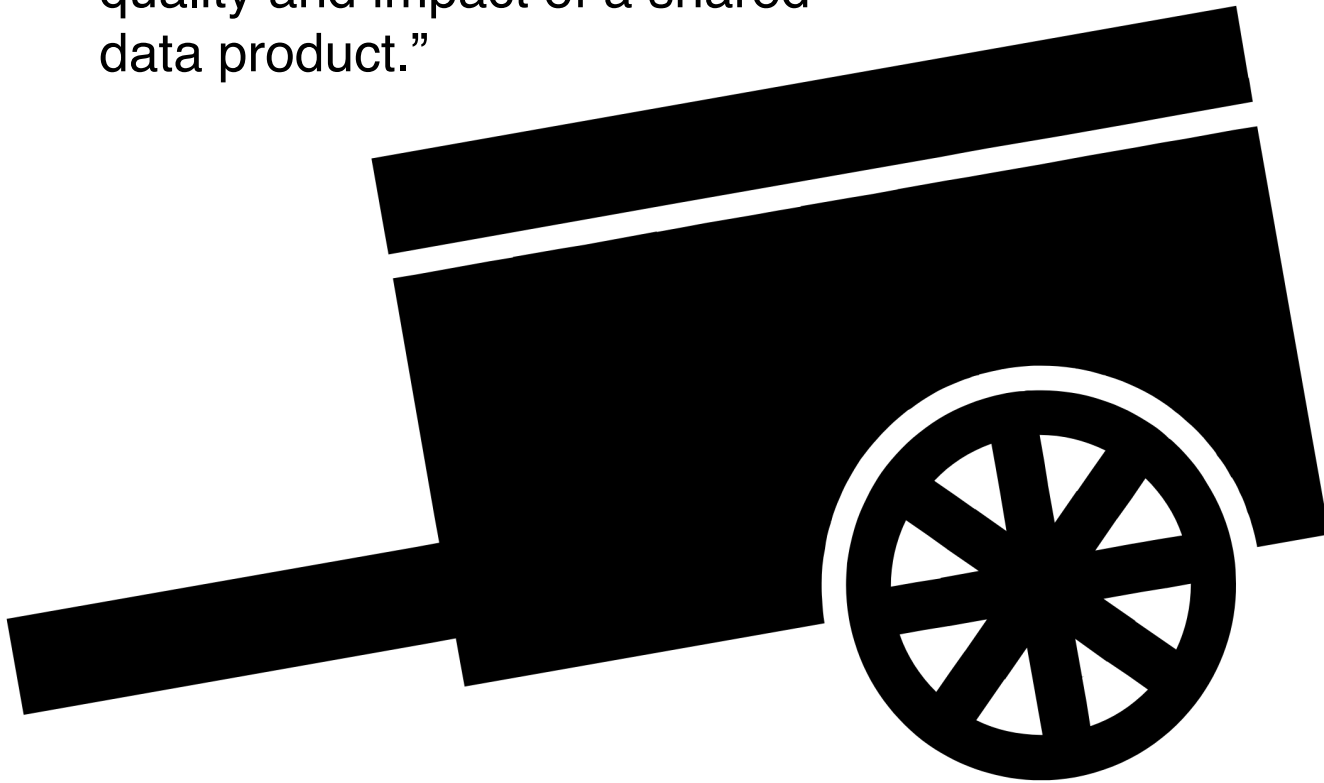
Clearly lay out data citation metrics' role and meaning for:

- Quality and impact of research products
- Incentives, rewards & recognition for researchers and institutions
- Inequities and biases of research and scholarship systems



Quality & impact

“In treating data as a first-class research product, we need clear measures that describe the quality and impact of a shared data product.”



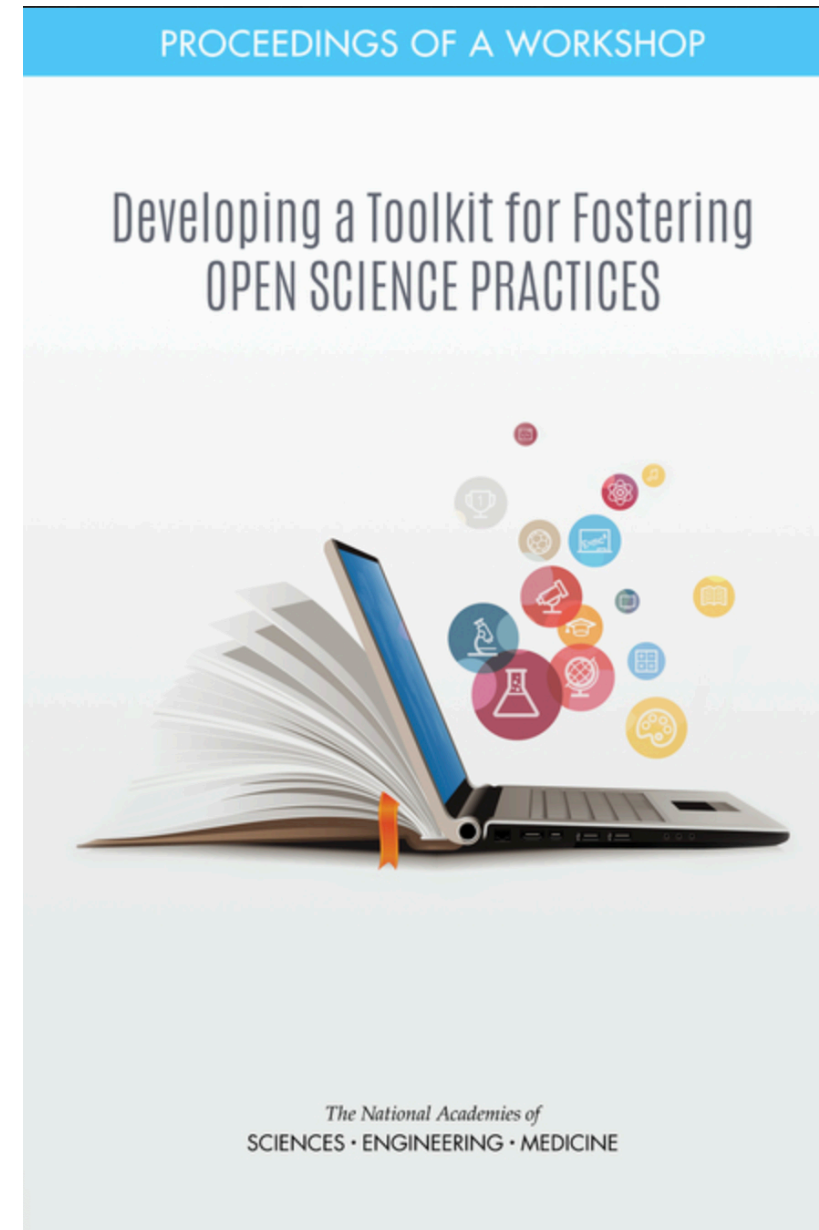
Incentives, rewards, & recognition

Data citation metrics are necessary, but to what purpose?

How do we

- incentivize
- reward
- recognize

behaviors that contribute to healthy, productive, sustainable scholarship systems?



National Academies of Sciences, Engineering, and Medicine. 2021.
Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop. Washington, DC: The National Academies Press.
<https://doi.org/10.17226/26308>

Inequities and biases

Let's not replicate those that already exist in:

- evaluation and credit systems
- (and between) research cultures
- costs
- intellectual property
- ethical perspectives
- resources and infrastructure

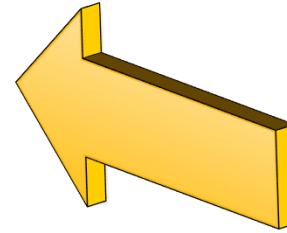
[8]



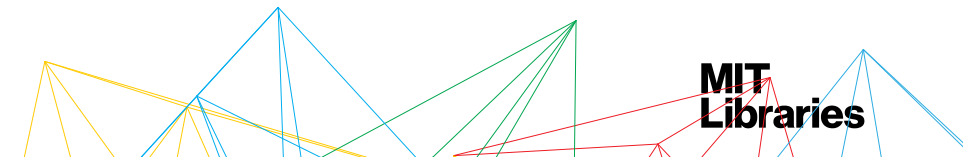
Print. By Kurt Komoda <https://www.flickr.com/photos/39573281@N06/7042782421> CC BY-NC-ND 2.0

Before we get to data citation metrics

- normative practice of data citation
- shared understandings of
 - what data citation metrics mean
 - what data citation metrics incentivize
 - what systems data citation metrics support
- infrastructure to support desired outcomes



**let's start
here**



Sources & Resources

Please respect the copyrights and licenses of the creators

- [1] "Conclusions and Recommendations." National Research Council. 1985. *Sharing Research Data*. p. 31. Washington, DC: The National Academies Press. doi: <https://doi.org/10.17226/2033>
 - [2] Altman, M., & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine*, 13(3/4). doi: <https://doi.org/10.1045/march2007-altman>
 - [3] Mooney, H. & Newton, M. P., (2012) "The Anatomy of a Data Citation: Discovery, Reuse, and Credit", *Journal of Librarianship and Scholarly Communication* 1(1), p.eP1035. doi: <https://doi.org/10.7710/2162-3309.1035>
 - [4] Anon. 20211001. Personal communication.
 - [5] Hrynaszkiewicz, I., & Cadwallader, L. (2021, September 27). A survey of funders' and institutions' needs for understanding researchers' open research practices. <https://doi.org/10.31219/osf.io/z4py9>
 - [6] ACS Research Data Policy. (2021). https://web.archive.org/web/20211005214909/https://publish.acs.org/publish/data_policy
 - [7] Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [<https://www.force11.org/group/joint-declaration-data-citation-principles-final>].
- National Academies of Sciences, Engineering, and Medicine. 2021. *Developing a Toolkit for Fostering Open Science Practices: Proceedings of a Workshop*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26308>
- [8] Leonelli, Sabina, Bezuidenhout, Louise, Schuster, Doug, & Stall, Shelley. (2021, October 1). Democratization of Data. Zenodo. <https://doi.org/10.5281/zenodo.5546659>

All icons unless otherwise noted licensed through [Noun Project](#)

- horse cart. By Firza Alamsyah <https://thenounproject.com/search/?q=cart+horse&i=3929852>
- Horse. By Slidicon <https://thenounproject.com/search/?q=horse&i=4193219>
- Elephant. By Amethyst Studio <https://thenounproject.com/search/?q=elephant&i=4132687>

Print. By Kurt Komoda <https://www.flickr.com/photos/39573281@N06/7042782421>
CC BY-NC-ND 2.0

