Building Trust: Data Metrics as a Focal Point for Responsible Data Stewardship

Iratxe Puebla Director, Make Data Count
Daniella Lowenberg Advisor, Make Data Count

22 May 2024

𝕏 @datacite
@makedatacount
@iratxepuebla

@danilowenberg
@makedatacount@openbiblio.social

I. Data stewardship beyond data access

Data stewardship needs to extend beyond data access

Open data has dramatically increased over the last decade, however...

Broad discoverability is not yet the norm for open data.

Datasets	Description of the datasets
7,182,036	Total profiles in ORCID
12,686	Total profiles in ORCID with at least one dataset
80,555	Total distinct datasets in ORCID
4,284	Total profiles in ORCID with at least one dataset indexed to Zenodo, Dryad, Figshare or DataCite

Exploring the relevance of ORCID as a source of study of data sharing activities at the individual-level: a methodological discussion. arXiv:2105.11825

Research assessment frameworks still focus heavily on publications and few include data.

"I think one barrier to data sharing is the merit review process within institutions for tenure and promotion; things such as 'how many people accessed your dataset' are not valued".

Perrier, Blondal & MacDonald (2020) The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. PLoS ONE 15(2). https://doi.org/10.1371/journal.pone.0229182

Data stewardship needs to extend beyond data access

We lack an understanding of which **data should be made open** and for what uses, **what data is highly used or underutilized**, and whether and **how data is translating into societal benefit**.

If we do not evaluate data usage, how can we build trust in our programs or goals of advancing science and society?

Evaluation of data usage should be supported by transparent and responsible data metrics

II. Data Metrics

Data metrics

Data metrics are **meaningful** and **contextualized** quantitative or qualitative measures of how open data sets are accessed or utilized

The responsible evaluation of data should be anchored on:

- Normalized counts across platforms
- A suite of metrics that address diverse perspectives:
    - disciplinary differences in production & usage of data,
    - different axes of value e.g. scarcity of datasets

"If it's done right, it's okay. If it's done, like, let's say the h-index then oh my god, please no."

Postdoc, HCI, Germany

Interviews by Stefanie Haustein & colleagues. Make Data Count Summit presentations. Make Data Count Summit, Washington DC. Zenodo. https://doi.org/10.5281/zenodo.8370593

III. Prioritizing Data Metrics

Make Data Count Summit '23

For the first time, convened government agencies and those involved in scholarly research to tackle nuanced issues about the importance of data metrics

IT'S TIME TO
MAKE **DATA** COUNT
TOGETHER.

Join us for the MDC Summit on September 5-6, 2024 in London, UK

Learn more →

Build on existing technical infrastructure

- **Data Usage Standards**: Broader implementation of COUNTER Code of Practice for RD

- **Bibliometric studies:** Identify missing or incomplete information to contextualize metrics

- **Aggregation Corpuses**: Leverage AI, persistent identifiers, aggregate linkages not exposed

Lane *et al.* (2024). An Invisible Hand for Creating Public Value From Data. Harvard Data Science Review, (Special Issue 4). https://doi.org/10.1162/99608f92.03719804

Build on existing cultural infrastructure

Broad discussions on evaluation

- Evidence Based Policymaking Act (US, 2018)
- OSTP Public Access Policy (US, 2022)
- HELIOS Open (US)
- CoARA (Europe)
- Barcelona Declaration

IV. It can be done, it will be done, it must be done

Immediate priorities for data metrics

Complete **bibliometrics studies**

Improve the **accessibility and discoverability of open data**,

Optimize **administrative and funder metadata**

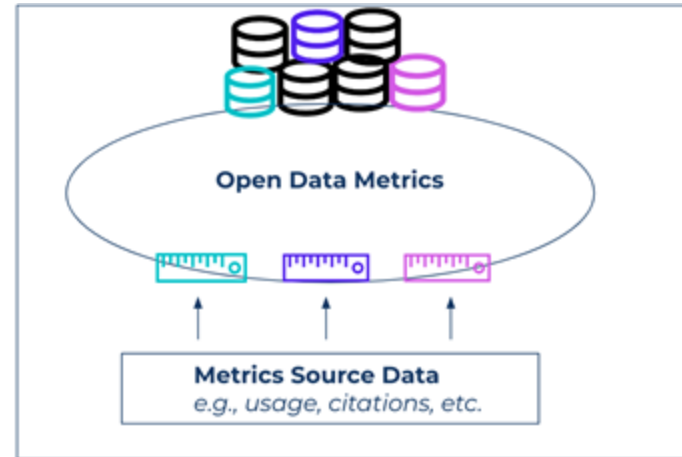Update **research and institutional assessment frameworks**

Prioritize improvements to **journal workflows**

Building trust

Data metrics needs to align to the same principles and values as open data

Build trust in data metrics by ensuring they are:
- ✓ **Complete**
- ✓ **Transparent & auditable**
- ✓ **Contextualized**

The development and broad use of open data metrics *can* happen because people and technology are ready, it *will* happen when resources are dedicated toward this need, and it *must* happen for the responsible stewardship of data as a public asset.

Thank you!

'Without an accurate means to track the downstream consequences of putting data into the public domain, as well as solid data quantifying the benefits, we will forever be working in the dark and depending on anecdote and good will. '

Scott Edmunds, Gigascience, member of the Make Data Count Advisory Group

iratxe.puebla@datacite.org
daniella.lowenberg@ucop.edu

Make Data Count:
makedatacount.org

Data Citation Corpus:
makedatacount.org/data-citation/

Join the Make Data Count listserv:
tinyurl.com/makedatacountupdates