

DataCite

Linking Datasets to Publications

Matt Buys

June 27, 2024
CHORUS Forum

 [@datacite](https://twitter.com/datacite)

 @datacite@openbiblio.social

**Connecting
advancing**

**research,
knowledge**



The how



The what

Great progress in data sharing over the last decade



National and funder policies in support of data sharing have dramatically raised awareness among communities and driven adoption.

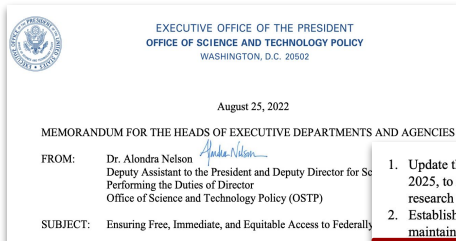
Infrastructure available to support preservation, publication, curation and controlled access to meet a wide range of needs.

Summary of open science practices in Horizon 2020 and Horizon Europe:

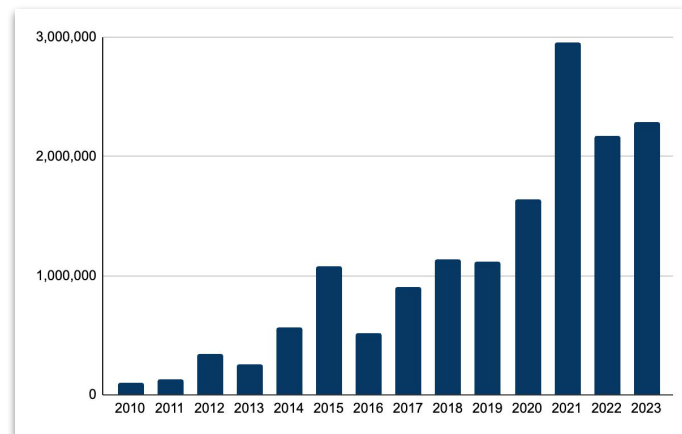
In Horizon 2020	Recommended or mandatory in Horizon Europe
Open access to research data under the principle "as open as possible, as closed as necessary"	Partially: Only for projects which are part of open Research Data Pilot Mandatory (but exploitation, protection of IPR, security and privacy rules have a higher priority)



Although the Plan S principles refer to peer-reviewed scholarly publications, cOAlition S also strongly encourages that research data and other research outputs are made as open as possible and as closed as necessary. The early sharing of research results through preprints is also strongly encouraged.



1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.



Over 15 million datasets registered with DataCite Commons commons.datacite.org

And still

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](https://arxiv.org/abs/2105.11825)

Research assessment frameworks focus heavily on publications & 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 L, Blondal E, MacDonald H (2020) The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. PLoS ONE 15(2): e0229182.

<https://doi.org/10.1371/journal.pone.0229182>

Cultivating Change Easy & Rewarding



Adoption of Open Research

by Hendricks & Buys (derived from Nosek: cos.io/blog/strategy-for-culture-change)

Researcher Perceptions

Researchers still perceive data sharing as a burden that brings little professional benefit

We need to **make it rewarding for researchers to share & cite data**

“I don't have time. [...] It's not incentivized either [...] so why would I spend some time on something where it really amounts to no recognition [...]. People who are in this community might think I'm a good person [...] and that's fine. I can accept that. But I'm not going to track it, necessarily, because it doesn't amount to anything that's recognized.”

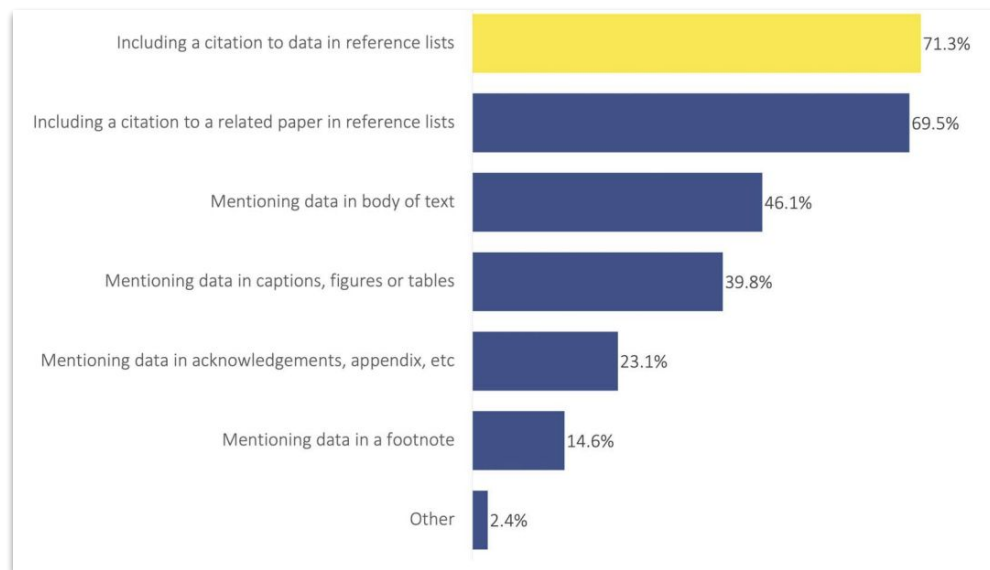
Professor, Anesthesiology, Canada

Data Citations

Provide a link between the dataset and another research object, pointing to the data being used or reused

A useful step toward evaluation of data usage:

- Valued by researchers
- Existing workflows available for repositories & publishers
- AI tools developed to identify data citations




Make Data Count



Build open infrastructure
and community-based
standards

Advocate through
outreach and adoption
campaigns

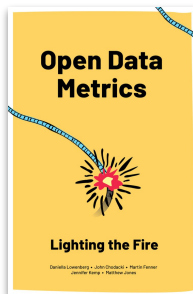
Contextualize with
evidence-based
bibliometric studies



The COUNTER Code of Practice for Research Data

The Code of Practice for Research Data Usage Metrics standardizes the generation and distribution of usage metrics for research data, enabling for the first time the consistent and credible reporting of research data usage. COUNTER welcomes input and feedback from the community on this first iteration, so that it can be further developed and refined.

A downloadable PDF is now available in the download section below.



Webinar

BEGIN:
metadata for meaningful
data metrics

Thu May 19, 2022 05:00 PM - 6:00 PM, CEST



Speakers: CHRISTINE EISENBERG, ROBERTO COSTA, NICOLA DI GIACOMO, DANIELA PERINO



RDA Data Usage Metrics WG Recommendations



RDA Recommendation

DOI: [10.15497/RDA00062](https://doi.org/10.15497/RDA00062)

Authors: Daniela Lowenberg, Thomas Jouneau, Ian Bruno

Published: 4th March 2022

Version: 1.0

Abstract: Research data are increasingly recognized as important outputs of scholarly research, yet there are currently no standardized or comprehensive metrics for research data as there are for articles. This Working Group was founded following a Birds of a Feather at RDA Plenary 10 hosted by the [Make Data Count](https://makedatacount.org/) initiative. Lending expertise from various projects and research

makedatacount.org/about-us/

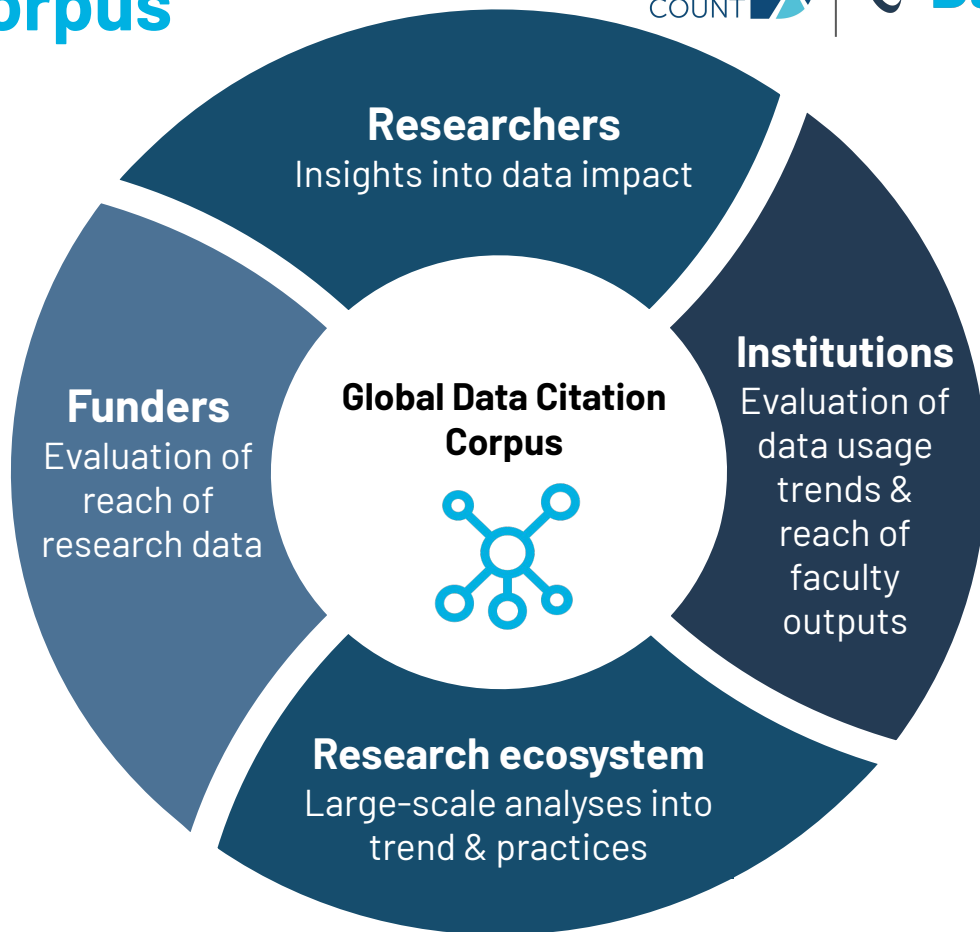
The Data Citation Corpus

MAKE
DATA
COUNT

 DataCite

Partnering across the community to support open and meaningful tools that provide insights on data usage

<http://corpus.datacite.org/>



Building on existing infrastructure

Immediate priorities

- We have to stop focusing on altering publisher behavior
- Repositories can and should add citation links to dataset metadata
- Text mining and NLP must be used to expedite surfacing data citation
- Leverage open metadata, persistent identifiers, aggregate linkages not exposed



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

**Join us for the Make Data Count
Summit on September 5-6, 2024 in
London, UK**

Live, in-person only event.



CONNECTING RESEARCH,
ADVANCING KNOWLEDGE



info@datacite.org



pidforum.org



datacite.org
datacite.org/blog



support.datacite.org
support@datacite.org



@datacite



@datacite



@datacite



DataCite