

The Making of Machine-Readable Metrics

Data metrics



Data metrics are meaningful and contextualized quantitative or qualitative measures of how open datasets are accessed or utilized. We can collect information on data usage via (for example):

- Views e.g. metadata, 3D models, images displayed on the landing page
- Downloads, file level or dataset level
- Citations, references to data, in the same way researchers provide a bibliographic reference to other scholarly resources

This slide is adapted from: Puebla, I. (2023, October 25). The Global Data Citation Corpus. Workshop on Open Citations & Open Scholarly Metadata 2023 (WOOC), Bologna, Italy. Zenodo. <https://doi.org/10.5281/zenodo.10039542>

12 BEST PRACTICES for RESEARCH DATA Sharing

Intended for
STAKEHOLDERS
from across the spectrum of
scholarly communications:

Researchers
Data repositories
Publishers
Funders
Research organizations
Policymaking institutions

1 Use trustworthy data repositories when publishing results

Deposit research data in repositories that assign persistent identifiers — ideally DOIs.

2

2 Enable FAIR practices for data sharing and citations

Foster sharing of research outputs with high-quality, complete metadata.

4

4 Set instructions for data citations

Offer guidance for authors to include data citations with persistent identifiers in the references.

1

3 Utilize persistent identifiers to cite data research

Provide attribution by citing datasets in reference sections using persistent identifiers.

3

5 Establish clear data policies for journals

Detail the way data is to be shared alongside a published article.

5

6 Integrate data availability statements in published articles

Include both human and machine-readable data availability statements in published articles where appropriate.

8

8 Incentivize data sharing through guidance and tracking

Offer guidance on open science practices, track policy compliance and incentivize researchers to share, cite and link research data.

10

10 Collectively develop tools, processes and incentives

Build systems and products that work together to enable easy, clear, and efficient data sharing.

12

Content drawn from the
Joint Statement on
Research Data created by:



7 Include data citations in article metadata

Include data citations with persistent identifiers (DOIs if possible) in the metadata registered with Crossref.

7

9 Connect articles and datasets through persistent identifier connections

Establish persistent identifier connections in metadata and reference lists.

9

11 Collaborate across stakeholders to align on FAIR research data policies

Work together to align FAIR data policies and guidelines.

11

12 Develop research assessment practices including data sharing and citation

Consider data sharing and citation in reward and recognition systems for research assessment.

“It is not sufficient for scientific journals merely to announce a data sharing requirement...without diligent enforcement, a toothless journal data policy appears to produce the same result as no policy at all; few authors post their data.”

Christensen, G., Dafoe, A., Miguel, E., Moore, D. A., & Rose, A. K. (2019). A study of the impact of data sharing on article citations using journal policies as a natural experiment. PloS one, 14(12), e0225883.

<https://doi.org/10.1371/journal.pone.0225883>

We have broad
stakeholder agreement on
standards & best
practices. We face a
challenge of adoption.



The poster is framed by a green border with playing card symbols at the corners: a heart at top-left, a club at top-right, a diamond at bottom-left, and a spade at bottom-right. The RDA logo, featuring a globe and the letters 'RDA', is centered at the top. Below it, the title 'RDA Data Usage Metrics WG Recommendations' is written in green. The section 'The Challenge:' is followed by a paragraph explaining the lack of standards for normalizing data usage. At the bottom, a 3D illustration of a red circular base with three white cards (Ace of Spades, King of Hearts, Queen of Diamonds) is shown. The footer includes the text 'Produced by: Data Usage Metrics WG' and a URL.

RDA

**RDA Data Usage Metrics WG
Recommendations**

The Challenge:

Standards for normalizing data usage (views, downloads) had not existed and community conversations around data metrics had been centered around data citation. It is essential for the development of data metrics and assessing research data that the community prioritize ways to reliably count and report data usage at repositories.

Produced by: **Data Usage Metrics WG**
<https://www.rd-alliance.org/groups/data-usage-metrics-wg>

Global Data Citation Corpus



A comprehensive corpus that incorporates data citations from different sources into a centralized, publicly accessible community resource

Supported by the Wellcome Trust

Collaborative effort to incorporate data citations from diverse sources:

- Persistent Identifier (PID) authorities (e.g. Crossref, DataCite) that collect citations as part of their metadata deposit workflow.
- Additional sources that aggregate or discover citations through various techniques, such as machine learning and curation of full-text in articles.

This slide is adapted from: Puebla, I. (2023, October 25). The Global Data Citation Corpus. Workshop on Open Citations & Open Scholarly Metadata 2023 (WOOC), Bologna, Italy. Zenodo. <https://doi.org/10.5281/zenodo.10039542>