DataCite

CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

DataCite Looking Ahead
Global Data Citation Corpus for All Data Citations

Matt Buys
Executive Director, DataCite
February 13, 2023

MAKE DATA COUNT

@datacite
@mjbuys

Our Challenge

5.8M+ tracked data citations within DataCite

99% from DataCite repositories

<1% from Publisher metadata

Fundamentally, the corpus addresses the major issue that known data citations exist in third-party systems and data citations are not compiled into a comprehensive, publicly accessible corpus that the community can use.

This work is supported by the Wellcome Trust grant # 226453/Z/22/Z

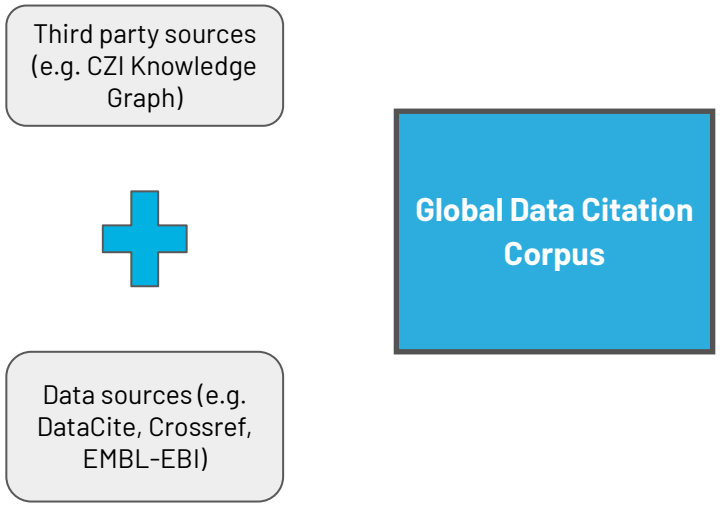
It takes a village

This requires collective action and coordination across partners and key stakeholders in the community is imperative. Some of the early collaborators include:

Populating the Corpus

The corpus will be populated through two channels:

- **Third party sources.** These sources aggregate or discover citations through various techniques, such as full text mining and curation.
- **Data sources.** These sources collect citations as part of their deposit workflow, such as Persistent Identifier (PID) authorities.



Corpus Data

The community needs a clear understanding of data reuse to monitor impact, inform future funding, and improve the dissemination of research. The development of a trusted central aggregate of all references to research data across articles, preprints, government documents, and other outputs will help achieve this goal.

- Dashboard. The dashboard will be an user interface that allows the community to view and understand data citations in the Corpus. The flexibility of this user interface will be extended through multiple iterations.

- Corpus data. The Corpus data will be openly available via data dumps and/or APIs. All data will be made available as CC0. We will have phased releases of the corpus data.

Our focus
working together

Our immediate focus is to build and launch prototype while working closely with stakeholders to gather input on the MVP design.

In the meantime;

1. Please share your user stories using the QR code or **link** in chat.
2. Follow **@makedatacount** and **@datacite**
3. Reach out to **info@datacite.org** if you have questions.

DataCite

CONNECTING RESEARCH,
IDENTIFYING KNOWLEDGE

info@datacite.org

pidforum.org

datacite.org
blog.datacite.org

support.datacite.org
support@datacite.org

@datacite

DataCite

@datacite