

Current State

Starting from Mooney and Newton's¹ work on data citation we decided to examine what happens to a data set after it is set on its path as a piece of scholarly communication.

- How do the elements that are tracked in the modern data citation market match those key elements identified by Mooney and Newton¹?
- How are data sets indexed for purposes of discovery and reuse?
- How do data citations contribute to scholarly reputation and impact?

As demonstrated by Mooney and Newton¹, there is little consistency in the guidance or practice of data citation. Even so, they identified seven key data citation elements: Author, Title, Published Date, Publisher, Material designator, Electronic retrieval location (URI), Persistent identifier.

Briefly reviewing a selection of datasets, repositories, and platforms we found an uneven application of commonly accepted standards. Although guides for repositories such as Dryad or identifier registrars such as DataCite recommend inclusion of the key elements of Author, Title, Published date, and Publisher, there are two notable trends: to leave off the Material designator, leading to confusion when differentiating between data sets, and other data publications, and articles; and to unify the Electronic retrieval location and Persistent identifier. This unification trend is exemplified by Dryad's direction that "The DOI should be presented with the web prefix <http://dx/doi.org>."²

Dataset Indexing

Element	#Indexed By
Author	6 ++++++
Title	6 ++++++
Published Date	6 ++++++
Publisher	5 +++++
Persistent Identifier / URI	5 +++++
Categories / Subjects	4 ++++
Material designator	3 +++
Tags / Description	2 ++
Share count	1 +
View count	1 +
Citation count	1 +
Link to article	2 ++
Contributor	1 +
Resource Type	2 ++
Format / Filetype	2 ++
doi prefix	4 +++++

Services reviewed: DataCite, Figshare, Dryad, ORCID, CrossRef, Data Citation Index

Scholarly Impact

As measured by

Altmetrics:	Citations Usage: views & downloads Social mentions: CiteUlike, Mendeley, Twitter, Facebook Blogs & Media
Impact Story:	Citations Usage: views & downloads, Social mentions: Twitter, Facebook, Figshare, Mendeley (career stage, country, discipline, readers)

Figure here
re: datasets
& impact

N/A
Data
Unavailable -
Not Enough
Information

Services reviewed: Altmetrics, Impact Story

CALL TO ACTION

There is a lack of data available demonstrating how the development of datasets enhances scholars' reputation in the developing arena of impact beyond article publication.

To encourage scholars to develop and cite data, consistent practices must be promulgated for data citation and indexing. New, open API tools offer an easier path to fuller functionality, as well as usage statistics.

Note that many of the groups implementing these API tools have also embraced JSON, a language-independent standard for data encoding, transfer, and reuse, to serialize metadata.

Key questions that still remain around data citations and their use in determining scholarly impact include:

- What are the minimum metadata necessary for citability and tracking?
- How do we ensure that more robust, domain-specific metadata are recorded to enable data reuse?
- What are the best practices for developing a web services architecture around data?

Fruitful answers to these questions may be pursued by developing the conversation, activities, and best practices around:

- Metadata serialization
- Metadata indexing (possibly using SOLR)

Once we get the basics out there (preferably at an API with JSON output) we can iterate on the hard parts!

REFERENCES

- ¹ Mooney, H, Newton, MP. (2012). The Anatomy of a Data Citation: Discovery, Reuse, and Credit. *Journal of Librarianship and Scholarly Communication* 1(1):eP1035. <http://dx.doi.org/10.7710/2162-3309.1035>,
²<http://datadryad.org/pages/depositing#howFormat>

Robert Hilliker, Digital Repository Manager
Columbia University Libraries
Website: academiccommons.columbia.edu
Email: rhilliker@columbia.edu
Twitter: [@ResearchAtCU](https://twitter.com/ResearchAtCU)
Phone: 212-851-7339



Amy Nurnberger, Research Data Manager
Columbia University Libraries
Website: scholcomm.columbia.edu/data-management/
Email: anurnberger@columbia.edu
Twitter: [@DataAtCU](https://twitter.com/DataAtCU)
Phone: 212-851-2827

