

The Expanded Use of DOI and Content Citation Granularity

CSE 2020 Virtual Annual Meeting

MODERATORS:

Nancy R Gough
Publishing Consultant
BioSeredipity, LLC
Elkridge, Maryland

SPEAKERS:

Midori Baer
Director Publishing Operations
PLOS
Washington, DC

Stacy Konkiel

Director Research Relations
Almetrics & Dimensions, Digital
Science
Minneapolis, Minnesota

Daniella Lowenberg

Dryad Product Manager & Make
Data Count Lead
California Digital Library
San Francisco, California

REPORTER:

Beverly Lindeen
Senior Managing Editor
Allen Press, Inc.
Lawrence, Kansas

It is pretty well known that DOIs (digital object identifiers) are used for articles that are published in an online format. However, they are also necessary for the various parts of an article, such as preprints, figures, tables, databases, data, methods, etc. Nancy Gough moderated this session looking at how DOIs for these items is used by publishers and indexing services.

In her presentation entitled “Expanded Use of DOIs at PLOS,” Midori Baer, Director Publishing Operations at PLOS, reviewed the ways in which PLOS uses DOIs as part of their larger organizational mission of “empowering researchers to accelerate progress in science and medicine by leading a transformation in research communication.” The continuum of research is important, as is the ability of the authors to tell the stories of their science.

Research needs to be discoverable. Article assets (supplementary information and datasets, tables, figures, peer-review reports, etc.) are assigned DOIs and are registered as component DOIs with Crossref. A link to the DOI appears in both the html and PDF views of all the article assets.

There are currently limitations. On the PLOS website, it would be beneficial to display contextualizing information, article details, and how-to-cite information. On Crossref, there should be information about the assets’ relationship to the parent article, as well as the titles of the assets. And, regarding peer-review reports, it would be best to update the schema so that PLOS can provide more robust metadata.

With preprint servers, the preprint version should be linked to the published article. Likewise, PLOS links to the preprint version of the paper.

Regarding data, PLOS requires authors to make all data available to the public without restriction at the time of publication. They have processes to manage compliance, and require that all data have persistent identifiers (PIDs).

PLOS also uses DOIs to facilitate access to laboratory methodology as part of their partnership with protocols.io. Researchers are encouraged to deposit lab protocols on protocols.io where DOIs are provided for the methods that will be included in the article itself.

To close her presentation, Baer listed some things that PLOS would like to do in the future: 1) Enhance and increase compliance for data and software citations. 2) Update DOI schema for peer-review reports. 3) Add contextualizing metadata to asset landing pages.

The second presentation, “From Idea to Impact: The Next Evolution in Linked Scholarly Information,” was from Stacy Konkiel with Dimensions. She discussed the Dimensions data approach.

Dimensions is a linked research discovery platform. It is an abstracting and indexing-like service with a broader information landscape. Data is pulled from many sources and carefully curated. Articles and all related data, preprints, grants, citations, patents, clinical trials, etc., are presented together. This makes tracing content types easier for the user. The general approach to data in Dimensions is inclusivity: They do not decide what information is relevant; rather, they enable publishers to include what they want with the article publication.

How does this citation granularity affect journal citations? Data shows that, in general, it is not negatively impacting article citations.

The final presentation was from Daniella Lowenberg with California Digital Library. In her presentation, “Make Data Count, Citing Dataset DOIs—A Revolution!” Lowenberg discusses data publishing. First, she indicates that storing data in supporting information files is not data publishing. Data publication consists of a few elements: 1) It must be citable with a PID that comes from the repository where it is stored. 2) There could be FAIR and data-specific metadata associated with the data themselves. 3) Data is much larger in size than articles, and repositories are equipped to handle these large files.

CONTINUED

Lowenberg then went on to talk about Make Data Count. This is an initiative started in 2014 by PLOS, the California Digital Library, and DataONE. It looked at what researchers value about their published data. In 2017, it transitioned to a project funded by the Sloan Foundation between DataCite, California Digital Library, and DataONE. They built the infrastructure for data use and data citation and wrote the COUNTER Code of Practice for Research Data.

Data citation is one of the components of making data count. In a recent blog post by Force11, Rachael Lammey and Helena Cousijn pointed out that “[d]ata citation needs to be a standard component of publication so that links from other research outputs to the data that supports them are comprehensive and helps the transparency and

reproducibility of research.” Examples of data citation are an article that cites a dataset, a dataset is derived from 2 other datasets, and then subsets of a dataset are generated.

There are concerns about making data citation possible. Lowenberg then shared an example of an article where the author cited data as a reference, which then caused the number of citations of that data to be 0 because it was not formatted correctly. Publishers need to play a role in indexing the citations properly. Scholix is a framework that helps with this.

How can publishers support the framework?

- Support FAIR data repositories and data curation
- Implement best practices for data citation indexing
- Educate authors on how to cite data in references