

Publishing data with a DOI (Digital Object Identifier) is becoming increasingly popular, and many researchers are being asked by journals to publish the data underlying the research results described in the article. Where does the data come from and how do I cite it correctly?

Since the science ministers of the G8 countries signed the Open Data Charter in 2013, the first clear ministerial statement in favour of open data, work has intensified worldwide on practices and guidelines to improve the availability of publicly funded scientific research data and software and how to provide appropriate credit for the data providing researcher or institution. This demand has now become a central component of science policy, following the maxim "as open as possible, as closed as necessary" (Council of the European Union 2016, p. 8), and is embedded in many data policies of research funders, scientific institutions and journals.

As early as 2015, for example, the German Research Foundation (DFG) adopted guidelines on the handling of research data (DFG, 2015), in which it pledges its support for the management of research data and makes a clear recommendation for the quality-assured storage of research data, collected or generated with DFG funds, in a suitable domain repository for research data.

The "Coalition for Publishing Data in the Earth and Space Sciences" (COPDESS 2015, 2018; <https://copdess.org/>) is

a landmark initiative for geoscience journals and publishers: In two declarations, leading scientific publishers (including Springer Nature, Elsevier, Science, AGU, EGU, Wiley, Copernicus Publications) have committed themselves to make the provision and publication of the data on which the scientific results of articles are based a condition of publication in journals. This should no longer be done in the form of an attached data supplement, but as an independent data publication (with DOI) and ideally via domain repositories for research data. Moreover, DOI-referenced data publications should be cited and included in the reference lists of scholarly literature, in brief treated exactly like other references. Many publishers provide clear instructions on their websites on which objects this applies to and also give recommendations on the use of domain repositories (see Fig. 1).

There are still many questions about citing research data: How do I cite data correctly and what does a citation look like? Where should I cite the data? Why can't I just put the DOI in the data availability statement of the article? At this point, we would like to clear up some of the confusion and provide some background information.

Publishing data is not just about putting a table in a repository; it is also about describing the data so that others can use it. These descriptions are called "metadata" and are used to explain the data itself (e.g.

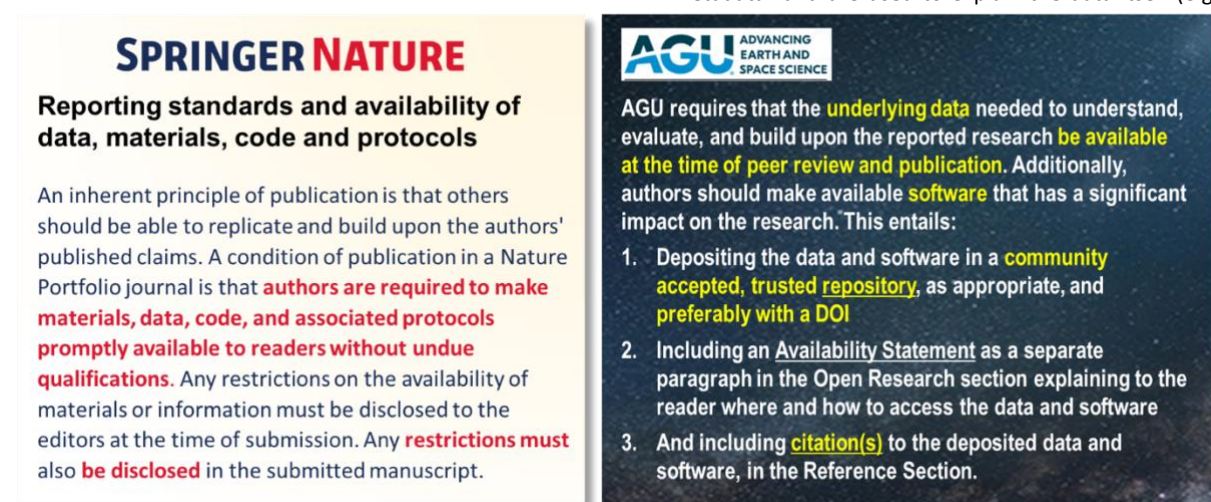


Figure 1: Examples of open data policies from the publishers Springer Nature and AGU. Common to both is the clear request to publish data, software and also sample descriptions on which the scientific results are based with DOI and to cite these in the reference list of the articles. Sources: (<https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards> und <https://www.agu.org/Publish-with-AGU/Publish/Author-Resources/Data-and-Software-for-Authors> (last access on 13 Oct 2022)

what variables are in the table and in what unit were they measured, where and when?) On the other hand, metadata are also used to unambiguously assign the data to the authors, i.e. the researchers who measured, collected or generated the data, and to be able to cite them. Just like scientific articles, data publications have "authors" (also called "creators"), a descriptive title, a year of publication and are published by a repository (analogous to a scientific journal or publishing house). These and some other information, such as disciplinary keywords or geographic coordinates, are mainly used to find and cite the data on the Internet. Citing a data publication is very similar to citing a scientific article:

Citation of a data publication:

Klinkmüller, M.; Schreurs, G.; Rosenau, M. (2016): GeoMod2008 materials benchmark: The sieve dataset. GFZ Data Services. <https://doi.org/10.5880/GFZ.4.1.2016.003>

Citation of a scientific article:

Klinkmüller, M., Schreurs, G., Rosenau, M., & Kemnitz, H. (2016). Properties of granular analogue model materials: A community wide survey. *Tectonophysics*, 684, 23–38. <https://doi.org/10.1016/j.tecto.2016.01.017>

A dataset can and should be cited in the same way as an article. Figure 2 clearly illustrates this: First, the dataset is cited in the text of the article, as only references cited in the text may appear in the reference list (step 1). An ideal place to cite the data is in the "Data Availability Statement" required by many journals; but the data can

also be cited elsewhere in the text. The full reference of the data publication (with DOI) is then included in the reference list of the article (step 2). Ideally, the data publication can be accessed directly from there with a single click. Of course, the scientific article will also be linked to the DOI landing page of the data publication (the page to which the DOI link leads), thus completing the cycle and ensuring permanent links between the data and the article (step 3).

The citation of data publications in the reference lists of the articles is important, as citation analysis tools only look for citations in this place. A DOI number in the body text of the article cannot be found (and is therefore not counted). So far, the value of data publications in the scientific value catalogue is still low, but it is increasing. There is currently much discussion about how "other forms of publication", including data or software publications, should be recognized as scientific output. The first institutes are also including data publications in the official publication lists of researchers (e.g. GFZ).

On 1 September 2022, the German Research Foundation (DFG) published the "Package of Measures to Change the Culture of Scientific Evaluation". It states that "the performance evaluation [of proposals], which is based on content-related and qualitative criteria, should explicitly include that the entire spectrum of scientific output should be equally represented and recognized in funding proposals and CVs. In addition to a maximum of ten publications in the most common publication formats, up to ten other research results and findings published in different publication formats can now be listed in the CV.

How to cite data correctly?

3. The data publication cites the scientific article and links to it.

1. Citation of a data publication in the text.

2. Entry of the complete reference (with DOI) of the scientific article in the reference list, which links to the data publication.

Figure 2: Visualization of the citation of data publications in scientific articles and linking of the articles on the DOI landing page of the data publication. Sources: : Klinkmüller et al., 2016: (<https://doi.org/10.1016/j.tecto.2016.01.017>, scientific article); Klinkmüller et al., 2016 (<https://doi.org/10.1016/j.tecto.2016.01.017>, data publication)

These can be, for example, articles on preprint servers, datasets or software packages." On an international level, there are several initiatives developing robust approaches to research assessment with a much broader scope than the current focus on the h-index (e.g. the San Francisco Declaration on Research Assessment DORA, the European Coalition for Advancing Research Assessment CoARA). These are all important steps towards recognizing data as an independent and recognized scientific achievement. A simple way to support this is to always cite data correctly.

FID GEO offers various consulting services on the topic of data publication (individual consulting, workshops for groups, presentations in colloquia or working groups) and the affiliated GFZ Data Services domain repository, is a suitable partner for publication requests from the community.

--

GFZ German Research Centre for Geosciences

Kirsten Elger & Melanie Lorenz, Potsdam

--

info@fidgeo.de

www.fidgeo.de

--

References

CoARA: European Coalition for Advancing Research Assessment: <https://coara.eu/>

COPDESS (2015) Statement of Commitment of the Coalition for Publishing Data in the Earth and Space Sciences. URL: <https://copdess.org/statement-of-commitment/>

COPDESS (2018) Enabling FAIR Data Commitment Statement in the Earth, Space and Environmental Sciences. URL: <https://copdess.org/enabling-fair-data-project/commitment-statement-in-the-earth-space-and-environmental-sciences/>

Council of the European Union (2016), The transition towards an Open Science system - Council conclusions (adopted on 27/05/2016). URL: <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>, p. 8

DFG (2015) Leitlinien zum Umgang mit Forschungsdaten. URL: https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/leitlinien_forschungsdaten.pdf

DFG (2022) Information für die Wissenschaft Nr. 61 | 1. September 2022: Maßnahmenpaket zum Wandel der

wissenschaftlichen Bewertungskultur. Deutsche Forschungsgemeinschaft. URL: http://www.dfg.de/foerderung/info_wissenschaft/2022/info_wissenschaft_22_61

DORA: Declaration on Research Assessment: <https://sfedora.org/>

Open Data Charter: <https://opendatacharter.net/g8-open-data-charter/>