UKEOF
ENVIRONMENTAL OBSERVATION FRAMEWORK

Data Citation and Publication for Researchers

ADVICE NOTE 4

Why cite data?

Data are a fundamental part of the research process. Without them, it is impossible to verify the results of published studies or reproduce the science. Yet modern datasets are so large that it is not practical for the data to be published as part of the research article, or in the supplementary materials, and so the data must be kept and managed in a suitable repository. This separation of data from article means that if the scientific record is to be maintained, the links between the article and the dataset must be kept.

Data citation provides a method of obtaining academic credit for the work put into creating, managing and curating a dataset. With a formal data citation, it becomes possible to piggy-back on existing methods for counting the impact of journal papers, providing an indication of how cited (and therefore how used) a dataset is.

There is a rising feeling in the scientific community that data can and should be treated as a first class research object, and those scientists who create them should get an appropriate level of academic credit for their work. Unsurprisingly, this requires a substantial cultural change, with buy-in from all levels of the scientific community, including the research funders.

I have used a dataset created by another researcher in my work. How do I cite the data?

If the data is stored in a trusted data repository, then the repository should provide information on how to cite the dataset.

An example citation is:

- Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras]. (2009a): GBS 20.7GHz slant path radio propagation measurements, Sparsholt site. NERC BADC. http://dx.doi.org/10.5285/E8F43A51-0198-4323-A926-FE69225D57DD

At this time there are no de facto standards for data citation strings, though the form laid out in the DataCite Metadata Schema for the Publication and Citation of Research Data (http://schema.datacite.org/) is gaining ground.

The recommended DataCite format for a data citation is:
Creator (PublicationYear): Title. Publisher. Identifier

The dataset I have used isn't in a repository. How do I cite the data?

We strongly recommend that you deposit the dataset in a trusted digital repository as this will ensure the dataset's persistence, which is vital for it to be part of the scientific record.

There are several free digital repositories available in the Earth Science, including, but not limited to: Pangaea.de, figshare.com, zenodo.org and of course the NERC federation of Environmental Data Centres.

I have created a dataset and I want to cite it. Do I need a DOI?

It is possible to cite a dataset using a URL or database accession number. Many journals now request that if a dataset is to be cited then it should have a Digital Object Identifier (DOI), though this does vary according to scientific discipline. In the long term, we would anticipate that all datasets supporting research publications would have their own DOIs.

If a DOI is needed, contact the dataset's host repository in the first instance to get one assigned.

What sort of criteria have to be met for my dataset to get a DOI?

Different data repositories have different criteria. However, for a DOI to be assigned to a dataset, the DOI must resolve to an open and publically available landing page. The dataset may be openly and publically available from that page, or there may be restrictions on the dataset, which should be given on the landing page.

What are my responsibilities once a dataset has a DOI assigned?

Once your dataset has had a DOI assigned, your primary responsibility is to ensure that your dataset is cited in the literature formally, using the form of words suggested for the formal citation of the dataset and the DOI, and to encourage others to cite your dataset similarly.

How does DOI issuing fit in with my data management plan?

DOI issuing will happen at the end of the data collection and submission process, but it is useful to start thinking about citation when drafting the data management plan. For example, the data management plan allows key datasets that will be created by the project to be identified, and their suitability for reuse tracking through citation assessed. Licensing and author permission for DOI issuing can also be thought about when writing the data management plan, potentially avoiding problems later on when the dataset is ready to be cited.

How can I publish my data in a way that gives me academic credit?

This will depend on your research discipline. In some cases, academic credit can be gained by making your data freely available through a suitable repository. In other cases, publishing your data (and a data article describing it) in a data journal may be appropriate.

A Data Journal is a journal that published short descriptive papers linked (usually be DOI) to a dataset held in a trusted repository, and both the dataset and the paper are reviewed as part of the publication process. The data article describes the dataset and the how, where, when and why it was created, but doesn't go into the analysis and conclusions needed by a more traditional article.

An incomplete list of data journals can be found at http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList

What further information is there on data citation?

- DataCite Metadata Schema for the Publication and Citation of Research Data (http://schema.datacite.org/)
- Ball, A., Duke, M. (2011). 'Data Citation and Linking'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/briefing-papers/
- CODATA-ICSTI Task Group on Data Citation Standards and Practices, "Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data" Data Science Journal Vol. 12 (2013) p. CIDCR1-CIDCR75 http://dx.doi.org/10.2481/dsj.OSOM13-043
- Joint Declaration of Data Citation Principles https://www.force11.org/datacitation

Extra guidance on data citation and publication for repository managers

Do I need to assign DOIs to cite data in my repository?

The short answer is no – any identifier can be used for citation. However, the current feeling is that researchers are more likely to cite data if it has a DOI. Note that this may vary in different research domains – in some cases accession numbers or URLs/ARKs/ other permanent identifiers are commonly used for data citation.

How do I get the ability to mint DOIs?

As a first step, contact DataCite (http://www.datacite.org) who will be able to put you in touch with your local DataCite member (in the UK, this is the British Library). The DataCite member will then work with you to create a contract and give you the ability to mint DOIs.

Note that there is a charge associated with this service, along with terms and conditions that can be found in the DataCite Business Models Principles http://www.datacite.org/sites/default/files/Business_Models_Principles_v1.0.pdf

How do I gain "trusted repository status" for a data journal?

At this time there isn't any standardized, formal way of proving trusted repository status, so the best route is to liaise with the journal directly to determine their requirements.

Some of the questions the journal are likely to ask can be found in

- Sarah Callaghan, Jonathan Tedds, John Kunze, Varsha Khodiyar, Rebecca Lawrence, Matthew Mayernik, Fiona Murphy, Timothy Roberts, Angus Whyte, "Guidelines on Recommending Data Repositories as Partners in Publishing Research Data" 2014, International Journal of Digital Curation, Vol. 9, No. 1, pp. 152-163 doi:10.2218/ijdc.v9i1.309

UKEOF
ENVIRONMENTAL OBSERVATION FRAMEWORK

www.ukeof.org.uk

This is a series of advice notes prepared by UKEOF's Data Advisory Group.

UKEOF works to improve coordination of the observational evidence needed to understand and manage the changing natural environment. It is a partnership of public sector organisations with an interest in using and providing evidence from environmental observations. Contact us at office@ukeof.org.uk.