

Guidance document for NERC scientists wishing to have DOIs assigned to their datasets

Introduction:

This document takes the form of a set of questions and answers (with the data producers asking the questions), providing information for scientists wishing to have a DOI assigned to their datasets stored in any of the NERC environmental data centres (EDCs).

Why should I get a DOI¹ for my dataset and how does this benefit me?

The NERC EDCs are proposing dataset citation using DOIs as a method of obtaining academic credit for the work put into creating, managing and curating a dataset. With a formal citation and DOI, 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 it 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.

NERC fully support the citation and publication of data, as explicitly stated in their data policy:

“NERC has a policy on data in order to: ...

- a. Help in the formal publication of data sets, as well as enabling the tracking of their usage to be tracked through citation and data licences.”

I want to use a DOI to cite my dataset – how do I get one assigned?

NERC-funded scientists:

If your dataset is stored in one of the NERC EDCs then you should contact your data centre representative with your request. They will check that the dataset meets the DOI citation criteria, and if it does, will get a DOI minted.

Note that all NERC-funded data should be published and assigned a DOI through a NERC EDC for archiving.

Legacy datasets may have DOIs assigned to them retrospectively. Please contact your EDC representative in the first instance to discuss this. In some cases, EDC data scientists might contact dataset authors to request permission to assign DOIs to legacy datasets that the EDC views as particularly significant.

Non-NERC-funded scientists:

If your data is suitable for archiving but is not stored in a NERC EDC, then please contact the data centre associated with your scientific domain (see table 1).

¹ Digital Object Identifier - The Digital Object Identifier (DOI) System is for identifying content objects in the digital environment. DOI names are assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change. See <http://www.doi.org/> or <http://www.datacite.org>

Data Centre	Area of interest
British Atmospheric Data Centre (BADC)	Atmospheric science
National Geoscience Data Centre (NGDC)	Earth sciences
NERC Earth Observation Data Centre (NEODC)	Earth observation
British Oceanographic Data Centre (BODC)	Marine Science
Polar Data Centre (PDC)	Polar Science
Environmental Information Data Centre (EIDC)	Terrestrial & freshwater science, Hydrology and Bioinformatics
UK Solar System Data Centre (UKSSDC)	Solar terrestrial physics and chemistry

Table 1: List of the NERC Environmental Data Centres and their scientific areas of interest

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

1. The DOI request must come from an author of the dataset. In cases where authorship has not been previously established via the data ingestion process, the responsible party for the data should provide authorization for the request. (The responsible party for the data should be named in the data management plan).
2. The dataset must be stable and complete, i.e. no more additions or changes are expected to be made to it.
 - a. For datasets which are long term, e.g. instruments which have been in place for years and are anticipated to be collecting data for the foreseeable future, it is possible to break the dataset into smaller pieces and assign DOIs to those pieces. When the dataset is completed in the future, a single DOI can then be issued for the whole data set.
3. The dataset should be in an appropriate format (as defined by the EDC)
4. The dataset should be permanent, by being deposited in a NERC EDC.
 - a. By assigning a DOI the EDC is committing to make the dataset available for posterity. To do this, the ownership of the dataset has to be clearly defined, along with the licensing terms to allow the EDC to distribute it.
5. The dataset should be of good technical quality.
 - a. The metadata describing the dataset should be complete.
 - b. By assigning a DOI the EDC are giving it a “data centre stamp of approval”.

How can I meet these criteria?

Because datasets come in many different shapes and sizes, from different scientific disciplines and instruments, it's not possible to apply a one-size-fits-all policy for dataset criteria. For this reason, we strongly recommend liaising with your EDC representative to ensure that these criteria are met. For the most cases, once the (stable and complete) dataset has been accepted by the EDC, it meets the minimum technical quality for assigning a DOI. However, the EDC representative might request (for example) additional metadata to be provided before assigning a DOI.

If there are any disputes about the technical quality of a dataset, with reference to assigning a DOI, the NERC Data Citation Issuers [NERC-DOI-ISSUER@JISMAIL.AC.UK] can be contacted as a mediator.

How will I know when my dataset is assigned a DOI?

When your dataset is assigned a DOI, an email will be sent to the requester confirming the DOI and the form of words for the formal citation for the dataset.

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.

What are the data centre responsibilities to my dataset once a DOI has been assigned?

When a dataset is assigned a DOI, the EDC confirms that:

- The dataset will be made available for the foreseeable future
- There will be no additions or deletions of files/records
- There will be no changes to the directory structure in the dataset “bundle”
- Changes to the data or format of the data will result in new versions of the datasets with new DOIs.
- The EDC will provide a full catalogue page (the landing or splash page) which will appear when any user clicks on the DOI hyperlink

How does citation fit in with my data management plan?

Dataset citation 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 citation assessed. Licensing and author permission for citation issues can also be thought about when writing the data management plan, potentially avoiding problems later on when the dataset is ready to be cited.

Who can I talk to about DOI assignment in the data centre?

Data Centre	Contact
British Atmospheric Data Centre (BADC)	badc@rl.ac.uk
NERC Earth Observation Data Centre (NEODC)	
UK Solar System Data Centre (UKSSDC)	
British Oceanographic Data Centre (BODC)	enquiries@bodc.ac.uk
Environmental Information Data Centre (EIDC)	eidc@ceh.ac.uk
National Geoscience Data Centre (NGDC)	enquiries@bgs.ac.uk
Polar Data Centre (PDC)	polardatacentre@bas.ac.uk

What further information is there on DOIs/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/>

Glossary

EDC – Environmental Data Centre

DOI – Digital Object Identifier - The Digital Object Identifier (DOI) System is for identifying content objects in the digital environment. DOI names are assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI name will not change. See <http://www.doi.org/> or <http://www.datacite.org>

PI – Principal Investigator