

Guidance for researchers wishing to have DOIs assigned to data resources they have generated

Introduction

This document provides information for researchers wishing to have a Digital Object Identifier (DOI) assigned to a data resource they have generated through NERC funding. Data resources must be deposited to one of the five NERC Data Centres which make up the NERC Environmental Data Service (EDS):

Data Centre	Area of interest	Contact
British Oceanographic Data Centre (BODC)	Marine science	enquiries@bodc.ac.uk
Centre for Environmental Data Analysis (CEDA) ¹	Atmospheric, earth observation, solar and space physics	support@ceda.ac.uk
Environmental Information Data Centre (EIDC)	Terrestrial & freshwater science, hydrology and bioinformatics	info@eidc.ac.uk
National Geoscience Data Centre (NGDC)	Earth sciences	ngdc@bgs.ac.uk
UK Polar Data Centre (PDC)	Polar and cryospheric science	polardatacentre@bas.ac.uk

Benefit of obtaining a DOI for data resources

NERC believes that data should be treated as a first-class research object, and those researchers who create it should get an appropriate level of academic credit for their work. This requires a substantial cultural change, with buy-in from all levels of the scientific community.

NERC fully support the citation and publication of data, as explicitly stated in their data policy^[1]:

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

“Help in the formal publication of data sets, as well as enabling the tracking of their usage through citation and data licences”

The NERC EDS supports data citation using persistent identifiers (DOIs) as a method of obtaining academic credit for the work put into creating, managing and curating a data resource. With a DOI, formal citation becomes possible and along with it, the ability to measure how well cited (and therefore how well used) a data resource is.

¹ The Centre for Environmental Data Analysis (CEDA) is run jointly with the Science & Technology Facilities Council

DataCite

In order to register DOIs, the NERC EDS works with DataCite^[2] in collaboration with the British Library^[3]. DataCite is a leading, global non-profit organisation which provides services to make it easier to discover and share datasets and to connect them with the broader research ecosystem. For example, by supporting data citation analytics.

How to get a DOI for data resources you have generated

NERC-funded researchers

All data generated with NERC funding must be offered to a NERC EDS Data Centre for long-term management and dissemination. Any DOIs assigned to data generated as a result of NERC funding must be issued by the NERC EDS. When depositing data in the Data Centre, you will usually be offered the option of obtaining a DOI. In order for the DOI to be provided, data must meet a minimum set of requirements (see below). Your NERC EDS representative will check that the data resource meets these criteria during the deposit process.

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

Non-NERC-funded researchers

Although the NERC EDS is principally for NERC-funded research outputs, we may also accept data resources that are funded by other means. If you would like further information, please contact the Data Centre associated with your scientific domain (see table above).

Criteria required in order to obtain a DOI for a dataset

- 1. The DOI request must come from an author of the dataset**
The request for a DOI should come from an author of the dataset or the researcher responsible for its management (as outlined in the Data Management Plan).
- 2. The dataset must be stable and complete**
That is, no additions or changes are expected to be made to it. For long-term datasets, 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 data resource into smaller pieces ('snapshots') and assign DOIs to those pieces. It is also possible to group the snapshots under one 'umbrella' record which may be assigned a DOI itself, even though it may be added to over time.
- 3. The dataset should be in an appropriate format**
Appropriate formats are defined by the Data Centre.
- 4. The dataset should be permanent**
By assigning a DOI the NERC EDS is committing to make the dataset available in perpetuity. The ownership of the data has to therefore be clearly defined, as do any licensing terms to allow the NERC EDS to distribute it.

5. **The dataset should be of good technical quality**

The data should be appropriately structured, reusable and well described. The metadata describing the dataset should be complete.

The NERC EDS accepts and manages a diverse range of data resources from different scientific disciplines. For this reason, we strongly recommend liaising with your NERC EDS representative to ensure that these criteria are met in a timely manner.

If there are any disputes about the technical quality of a dataset, with reference to assigning a DOI, the NERC Data Citation group can be contacted as a mediator. Contact your NERC EDS representative for details on how to do this.

Responsibilities

Researchers

When data has been 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.

It is the researcher's responsibility to ensure that any data resource referred to in subsequent publications is cited formally, using the form of words suggested by the NERC EDS ^[4, 5]. Researchers should also encourage others to similarly cite the data.

NERC EDS

When a data resource is assigned a DOI, the NERC EDS confirms that:

- It will be made available for the foreseeable future
- There will be no additions or deletions of files/records
- Any change made to a data resource will result in a new version of that resource with a new DOI.
- The NERC EDS will provide a data description page (landing page) which will appear when any user clicks on the DOI hyperlink

DOIs and Data Management Plans

Increasingly, journals require any data referred to in a journal publication to be deposited in an approved repository and have an associated DOI. In order for NERC researchers to meet this requirement in good time, proper data management planning is required. The NERC EDS will always do their best to provide DOIs for datasets in a timely fashion. However, researchers must allow sufficient time to prepare the data they wish to deposit and for which they want to obtain a DOI.

Whilst data submission to a NERC EDS Data Centre commonly occurs towards the end of a project, it is useful to start thinking about data citation when drafting the data management plan. For example, a data management plan should identify key datasets that will be created by the project and their suitability for citation should be assessed. Authoring, licensing and rights issues should also be considered when writing the data management plan, thus avoiding problems at a later stage when the dataset is ready to be deposited.

Further information

- [1] NERC Data Policy <https://nerc.ukri.org/research/sites/environmental-data-service-eds/policy/> [retrieved 20 Feb 2021]
- [2] DataCite <https://datacite.org/> [retrieved 20 Feb 2021]
- [3] The British Library and DataCite <https://www.bl.uk/datacite> [retrieved 20 Feb 2021]
- [4] Ball, A. & Duke, M. (2015). 'How to Cite Datasets and Link to Publications'. DCC How-to Guides. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets> [retrieved 20 Feb 2021]
- [5] Citation formatting tool <https://citation.crosscite.org> [retrieved 20 Feb 2021]. Different journals and publications may require different citation styles. With the DOI Citation Formatter, you can create more than 500 different citation styles