

NERC EDS

DOIs for growing datasets: best practice guidelines



Description

This document sets out the NERC Environmental Data Service (EDS) position on assigning a single DOI to datasets that are growing, and the best practice guidance on how to implement this. This guidance should be viewed in conjunction with the [NERC EDS DOI policy](#), [Data citation guidance](#) and [NERC EDS data preservation and withdrawal policy](#).

NERC EDS position

Assuming certain conditions are met (i.e. the dataset is within scope) and the recommendations detailed below have been discussed, the EDS are able to assign a single DOI to a dataset that is growing. Upon deposit of the initial dataset, a DOI will be issued. Subsequent additional data can be added to the dataset without the need to issue new DOIs for each update. This position is in alignment with DataCite guidelines, where *minor* changes are permitted to a resource that has an associated DOI. A minor change refers to a correction or update that does not fundamentally alter the nature, interpretation or citation of the dataset. This provision has been implemented as a response to operational and user requirements.

Scope

The guidance detailed below should only be used if the following conditions apply:

- i. The dataset is expected to **grow over time** (updated regularly or irregularly), and additions are temporal. Spatial and variable additions are also permitted if they are concurrent with temporal additions.
- ii. Where the previous data and their associated scientific meaning as well as the mandatory DataCite metadata (see appendix) **DOES NOT CHANGE**.

- iii. A NERC EDS data centre must agree in advance of the initial data deposit that the data are within scope allowing them to offer this service. Best practice should be followed from the point it is agreed onwards with **no retrospective changes** allowed. For example, if a depositor has had yearly data published with the data centre previously, each year with its own DOI assigned, a single DOI for the growing data cannot subsequently be implemented to include the already published data. Going forward, if it is agreed, a new growing dataset can be assigned a single DOI and subsequent years data added to the growing dataset. Data previously published as a timebound stand-alone resource with a DOI cannot subsequently become an open-ended, growing dataset.
- iv. Depositors should be ‘trusted’ to provide reliable resources over time, i.e., both engaged in the deposit process, and have clear quality control procedures in place to ensure continued production of robust, high-quality data products, to minimise the risk of subsequent errors being found, or substantive changes in data product. Individual Data Centres will evaluate this prior to initial deposit and may require evidence from the depositor.
- v. Licensing arrangements are not affected by updates to the resource.
- vi. There must be a temporal attribute for all growing data. However, new data may add additional attributes. In these cases, already published data may require the use of a null value for these new attributes. Attributes may also cease to be provided (e.g., instrument failure) as part of an ongoing resource; this should be clearly indicated in the discovery metadata record and null values included in the data itself. Changes to attributes supplied must not change the scientific meaning of previously published data.

Examples of data that are in scope

- Time-series data coming from sensors in river catchments across the UK published monthly.
- Model simulation runs carried out monthly, quarterly etc.
- Timeseries data that locates and characterises a low-pressure region in the Amundsen Sea
- Data providing 93-day sea ice forecasts generated every day and published quarterly
- High frequency magnetic field induction coil data published yearly

Examples of data that are out of scope

- Data (digital) collections
- Data that are growing but where additions change previous data and/or mandatory DataCite metadata
- Data that are growing solely on a spatial or variable basis (there must be a temporal aspect)
- Amendments to already published data subsequently found to contain errors; errata should be issued instead.
- Data that are growing but the dataset subsequently splits (e.g., ensemble model runs). The split necessitates a step change and so closure of one dataset and requires a new dataset to be initiated.
- Data not held in a NERC EDS repository, or not suitable for a DOI e.g., they don't meet the NERC EDS requirements (see [NERC EDS DOI policy](#))
- Any dataset whose acquisition or end-user licensing is incompatible with EDS DOI policy
- Data deposited by a depositor who lacks engagement with the deposit process and Data Centre.

Recommendations

In all cases, depositors **must** contact the relevant data centre to discuss their request for a single DOI to be assigned to data that will grow in advance of first publication and, ideally, first deposit. The request must meet the conditions detailed above. The following recommendations around changes to data and metadata and their governance must be considered and agreed with the depositor before assigning a single DOI for a dataset that is growing.

Author names

An organisational/institutional group (e.g., delivery programme or monitoring scheme) can be used as the author/creator of the data. Individuals who have contributed to creation of the data can be acknowledged as a 'contributor'; however, they will not be included in the citation of the data and are less likely to receive credit for the output. The 'contributor' list can change over time as new data are added; however, the data centre will not retrospectively change the contributor list for data that have been published.

If named individual authors are to be included as the creators, a fixed author list for the growing data must be agreed prior to implementation and minting of the DOI. This list cannot change over time, and depositors should be made aware of this. If the list of authors does change over time, a new DOI must be issued, and the previous dataset may not include any additional data.

Dataset title

Data centres should avoid including date ranges in the titles of data that are growing. However, date information in the title is often useful to users. Data centres *could* consider using a start date (e.g., *Plant monitoring data collected annually from 2012*) if there are no plans to add data from previous years. Care should also be given to the use of organisational names in dataset titles as changes can also occur during the lifetime of the ongoing asset.

Citation recommendations

To support reproducibility, the data centre should provide citation recommendations to depositors upon publication, and to users on receipt of data. The recommendations should include the requirement to add an 'accessed on' date to the data citation.

Discovery metadata for growing data

Discovery metadata records for growing data must include the following elements:

Temporal extent(s) which accurately describes the extent of the currently published data resource

Maintenance note. Maintenance and update frequency (using a value taken from the ISO 19115 MD_MaintenanceFrequencyCode list) is already a mandatory field for GEMINI compliance. For growing data, a "maintenance note" must also be added which should include a brief explanation of the nature of the growing resource including an accurate account of the planned update frequency.

In addition, it is **strongly recommended** the following are completed:

The lineage should be updated to include the same information provided in the "maintenance note" - this will assist with human readability of the metadata.

Resource date (dateType lastUpdate)

Resource date (dateType plannedUpdate)

Updating of Datacite Metadata

Where catalogue metadata are updated to reflect the change in the data resources, associated Datacite metadata entries also should be updated.

Data subsequently found in error

Data centres should review the [NERC EDS data preservation and withdrawal policy](#) and the [NERC EDS DOI policy](#) if they are made aware that data they have published with a DOI is found to contain errors. In general, major modifications to the data due to rectifying the error will result in the incorrect version being withdrawn and a new version being published with a new DOI. The dataset would then 'grow' under the new DOI. Depositors must be made aware of this before the data centre agrees to assigning a single DOI for the dataset that will grow. Depositors must provide sufficient evidence/reassurance that the data they are submitting for publication are quality assured to reduce the chance of error.

Terminating a growing dataset

All growing datasets will come to an end eventually. Data Centres should have a documented process for terminating a growing resource which includes updating:

- Temporal extent field with an end date
- Frequency of update field
- Data status field
- Lineage text should also be updated to make it clear that updates to the resource have been halted, and no new data will be added.

Document control

Version: 1.0

Last updated: 02 June 2026

Authors: Kathryn Harrison, Phil Trembath, Graham Parton, Jo Bowie, Paul Breen and James Ayliffe

Approved by: NERC Data Operations Group

Document owner: NERC EDS

Review date: June 2028

Appendix

Discovery metadata which must not change

- Dataset title
- Dataset authors (creator)
- Dataset publisher
- Dataset publication date
- Resource type (commonly 'dataset')