



Data publication activities in the Natural Environment Research Council

A. Leadbetter (1), S. Callaghan (2), R. Lowry (1), G. Moncoiffé (1), S. Donnegan (2), S. Pepler (2), N. Cunningham (3), P. Kirsch (3), L. Ault (4), P. Bell (4), R. Bowie (4), K. Harrison (5), B. Smith-Haddon (5), A. Wetherby (5), D. Wright (5), and M. Thorley (6)

(1) British Oceanographic Data Centre, Liverpool, United Kingdom, (2) British Atmospheric Data Centre, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, United Kingdom, (3) British Antarctic Survey, High Cross, Madingley Road, Cambridge, United Kingdom., (4) British Geological Survey, Kingsley Dunham Centre, Keyworth, Nottingham, United Kingdom., (5) Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, United Kingdom., (6) Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, United Kingdom.

The Natural Environment Research Council (NERC) is implementing its Science Information Strategy in order to provide a world class service to deliver integrated data for earth system science. One project within this strategy is Data Citation and Publication, which aims to put the promotion and recognition stages of the data lifecycle into place alongside the traditional data management activities of NERC's Environmental Data Centres (EDCs).

The NERC EDCs have made a distinction between the serving of data and its publication. Data serving is defined in this case as the day-to-day data management tasks of:

- acquiring data and metadata from the originating scientists;
- metadata and format harmonisation prior to database ingestion;
- ensuring the metadata is adequate and accurate and that the data are available in appropriate file formats;
- and making the data available for interested parties.

Whereas publication:

- requires the assignment of a digital object identifier to a dataset which guarantees that an EDC has assessed the quality of the metadata and the file format and will maintain an unchanged version of the data for the foreseeable future
- requires the peer-review of the scientific quality of the data by a scientist with knowledge of the scientific domain in which the data were collected, using a framework for peer-review of datasets such as that developed by the CLADDIER project.
- requires collaboration with journal publishers who have access to a well established peer-review system

The first of these requirements can be managed in-house by the EDCs, while the remainder require collaboration with the wider scientific and publishing communities. It is anticipated that a scientist may achieve a lower level of academic credit for a dataset which is assigned a DOI but does not follow through to the scientific peer-review stage, similar to publication in a report or other non-peer reviewed publication normally described as grey literature, or in a conference proceedings.

At the time of writing, the project has successfully assigned DOIs to more than ten legacy datasets held by EDCs through the British Library acting on behalf of the DataCite network. The project is in the process of developing guidelines for which datasets are suitable for submission to an EDC by a scientist wishing to receive a DOI for their data.

While maintaining a United Kingdom focus, this project is not operating in isolation as its members are working alongside international groups such as the CODATA-ICSTI Task Group on Data Citations, the DataCite Working Group on Criteria for Datacentres, and the joint Scientific Commission for Oceanography / International Oceanographic Data and Information Exchange / Marine Biological Laboratory, Woods Hole Oceanographic Institution Library working group on data publication.