Data publication activities in the Natural Environment Research Council

Adam Leadbetter, Sarah Callaghan, Roy Lowry, Gwenaelle Moncoiffe, Steve Donegan, Sam Pepler, Nathan Cunningham, Peter Kirsch, Linda Ault, Patrick Bell, Rod Bowie, Kate Harrison, Ben Smith Haddon, Anita Weatherby, Dan Wright and Mark Thorley

Natural Environment Research Council. United Kingdom

alead@bodc.ac.uk

Introduction

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).

Data serving vs. data publication

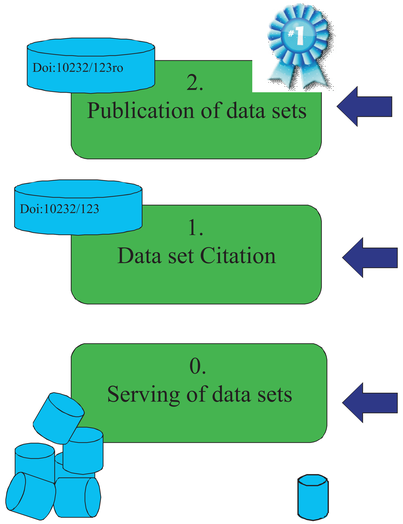
The NERC EDCs have made a distinction between the serving of data and its publication. Data serving is defined in this case as the day-to-day data management tasks of:

- acquiring data and metadata from the originating scientists;
- metadata and data file 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
- requires collaboration with journal publishers who have access to a well established peer-review system

The first of these can be managed in-house by the EDCs, while the remainder require collaboration with 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.



This involves the peer-review of data sets, and gives "stamp of approval" associated with traditional journal publications. Can't be done without effective linking/citing of the data sets.

This is our first step for this project – formulate and formalise a way of citing data sets. Will provide benefits to our users – and a carrot to get them to provide data to us!

This is what data centres do as our day job – take in data supplied by scientists and make it available to other interested parties.

A heirarchy of data serving, citation and full publication of datasets. We consider serving datasets to be publication level 0, citation as level 1 and data papers as level 2

Community benefits

Data publication benefits a large number of groups in the academic community, for example:

- Data scientists who can receive academic credit for the compilation of datasets
- Funding organisations who can ensure greater awareness and reuse of datasets they funded.
- The wider research community, who can see the completeness and trustworthiness of a dataset.

Additionally, data publication can help to improve public accountability of the scientific process.

Results

The project has successfully assigned DOIs to around twenty legacy datasets held by EDCs through the British Library acting on behalf of the DataCite network. Each DOI assigned resolves to an HTML landing page expanding the metadata available for the dataset.

The project is in the process of developing guidelines developing guidelines for dataset submission to an EDC by scientists wishing to receive a DOI for their data.

International links

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[1] working group on data publication.

Acknowledgments

Thanks to the British Library for providing the DOI minting account and DOI allowances to allow this project to cite datasets held in the NERC repositories.

Related work reference

[1] L. Raymond *et al,* 2012. Data Publication: A Partnership between Scientists, Data Managers and Librarians. Geophysical Research Abstracts 14: EGU2012-1440

NATURAL ENVIRONMENT RESEARCH COUNCIL

BODC

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

BRITISH LIBRARY