



Data citation and publication by the UK's Natural Environment Research Council's data centres

Sarah Callaghan (1), Nathan Cunningham (2), Mark Thorley (3), Roy Lowry (4), Gwen Moncoiffe (5), and the NERC Data Citation and Publication Project Team

(1) British Atmospheric Data Centre, STFC, UK (sarah.callaghan@stfc.ac.uk), (2) Polar Data Centre, British Antarctic Survey, UK (njcu@bas.ac.uk), (3) NERC, UK (mrt@nerc.ac.uk), (4) British Oceanographic Data Centre, UK (rkl@bodc.ac.uk), (5) British Oceanographic Data Centre, UK (gmon@bodc.ac.uk)

The validation of scientific research is carried out through a process of peer-review, usually as part of the publication of scientific results in a recognised academic journal. The peer-review process ensures the quality of the work reported in the journal article, while the publication process produces an article which is fixed and citable, and provides its author(s) with academic credit.

In the past, the review process has mainly focussed on the analysis, interpretation and conclusions drawn from a given dataset, as these are the information that can be easily published in hard copy text format with the aid of diagrams. The peer-review process is generally not applied to the underlying data itself, as datasets are usually stored in digital media, in a variety of (often proprietary or non-standard) formats. Yet for the conclusions drawn in the paper to stand, the data supporting them must be of good quality.

A process of data publication, involving peer-review of datasets, would be of benefit to many sectors of the academic community. Data publication and citation will allow scientists to receive academic credit for their work in ensuring that their data and metadata is complete, valid and stored in an accredited data repository. In traditional academic publishing, the object to be cited (i.e. the article) is written and peer-reviewed before it becomes citable. In the case of data publication, it makes more sense to allow citation of the dataset before full peer-review. If scientific peer-review of a dataset is considered the “gold standard”, citation therefore becomes a “silver standard”, confirming that the dataset is complete, unlikely to change, in an appropriate format and has sufficient metadata (at least as far as the host repository are concerned).

The UK's Natural Environment Research Council (NERC) funds six data centres which between them have responsibility for the long-term management of NERC's environmental data holdings. The NERC Science Information Strategy (SIS) has been created to provide the framework for NERC to work more closely and effectively with its scientific communities in delivering data and information management services.

The NERC SIS data citation and publication project is a cross data centre project which aims:

- To increase NERC's influence on work to provide and cite data outputs from scientific work in similar ways to scientific papers.
- To implement publication and citation of datasets held within the NERC data centres.
- To form partnerships with other organisations with the same goal of data publication to exploit common activities and achieve a wider community buy-in. To this end, project team members are involved with both the SCOR/IODE/MBL WHOI Library Data Publication Working Group and the CODATA-ICSTI Task Group on Data Citation Standards and Practises.

The wider win is to the research integrity of NERC's data. Data citation and publication will ensure that data is available, peer-reviewed, citable, easily discoverable and reusable; and it facilitates data transparency and scrutiny. It will be used by researchers to increase their academic status, thereby providing encouragement for them to archive and document their data appropriately, for the benefit of the wider research community.