



Data publication – policies and procedures from the PREPARDE project

Sarah Callaghan (1), Fiona Murphy (2), Jonathan Tedds (3), John Kunze (4), Rebecca Lawrence (5), Matthew S. Mayernik, (6), Angus Whyte (7), and Timothy Roberts (2)

(1) British Atmospheric Data Centre, United Kingdom (sarah.callaghan@stfc.ac.uk), (2) Wiley-Blackwell, (3) University of Leicester, (4) California Digital Library, (5) Faculty of 1000 Ltd, (6) NCAR, (7) Digital Curation Centre

Data are widely acknowledged as a first class scientific output. Increases in researchers' abilities to create data need to be matched by corresponding infrastructures for them to manage and share their data. At the same time, the quality and persistence of the datasets need to be ensured, providing the dataset creators with the recognition they deserve for their efforts.

Formal publication of data takes advantage of the processes and procedures already in place to publish academic articles about scientific results, enabling data to be reviewed and more broadly disseminated. Data are vastly more varied in format than papers, and so the policies required to manage and publish data must take into account the complexities associated with different data types, scientific fields, licensing rules etc.

The Peer REview for Publication & Accreditation of Research Data in the Earth sciences (PREPARDE) project is JISC- and NERC-funded, and aims to investigate the policies and procedures required for the formal publication of research data. The project is investigating the whole workflow of data publication, from ingestion into a data repository, through to formal publication in a data journal. To limit the scope of the project, the focus is primarily on the policies required for the Royal Meteorological Society and Wiley's Geoscience Data Journal, though members of the project team include representatives from the life sciences (F1000Research), and will generalise the policies to other disciplines.

PREPARDE addresses key issues arising in the data publication paradigm, such as: what criteria are needed for a repository to be considered objectively trustworthy; how does one peer-review a dataset; and how can datasets and journal publications be effectively cross-linked for the benefit of the wider research community and the completeness of the scientific record? To answer these questions, the project is hosting workshops addressing these issues, with interactions from key stakeholders, including data and repository managers, researchers, funders and publishers. The results of these workshops will be presented and further comment and interaction sought from interested parties.