



Data publication projects at the British Oceanographic Data Centre

Adam Leadbetter, Roy Lowry, and Gwenaëlle Moncoiffé

British Oceanographic Data Centre, Liverpool, United Kingdom (alead@bodc.ac.uk)

The British Oceanographic Data Centre (BODC) is part of the IOC International Oceanographic Data and Information Exchange (IODE) network. These data centres significantly change the data submitted to them, adding value through the metadata generation, quality control of the data and ingestion to a schema common to that data centre. In this data management paradigm, the current best available version of the data is served upon request, with no snapshots being preserved and no formal versioning process during the work up of the data. During the 1990s, BODC deviated from this paradigm by publishing dataset snapshots on CD-ROM from the dynamic system as part of its project data management service. However, this practice ceased in 2001 as CD-ROMs were deemed to be anachronistic, and the focus switched to delivering "best currently available" data via the internet.

Since that time, the digital library community has gained momentum based on a different data publication paradigm. This is based upon a fixed, unchanging dataset that is accessible from a persistent identifier (for example a digital object identifier [DOI]) for citation in the scientific literature and is usable on decadal timescales (as per the Open Archival Information System). For an IODE data centre to engage in the digital library publishing paradigm requires a mapping of the data centre's standard procedures onto the new data publishing paradigms.

BODC has engaged in two data publication pilot projects: the first is to bring the CD-ROM model up-to-date by publishing time frozen snapshots of project datasets online with DOIs (as provided by the British Library, under the auspices of the DataCite project). However, the ingestion and compilation of a complete project dataset at BODC can be a lengthy process and the scientific publishing industry is now pressing for DOIs resolving to data to accompany research papers. Therefore the second data publication pilot project at BODC aims to publish data on submission to the data centre if the dataset meets the criteria laid out in a dataset quality assessment checklist. For datasets meeting these criteria, a DOI can be assigned which allows the originating scientists to cite their data in publications immediately, removing the time lag associated with assigning a DOI to an entire project's data.