



ESSD: Real World Issues and Challenges of High-Quality Data Publication

Hans Pfeiffenberger (1) and David Carlson (2)

(1) Alfred Wegener Institut, Bremerhaven, Germany (hans.pfeiffenberger@awi.de), (2) CarlsonWorks, Boulder, USA (ipy.djc@gmail.com)

The Copernicus data publication journal Earth System Science Data (ESSD) represents an important and unique (and by no means final!) step forward in the larger world of data publication. Working with authors, reviewers, editors and data centres, ESSD has successfully produced many high-quality data publications across a wide variety of scientific disciplines, for individual data sets, multiple data sets as the product of scientific consortia and in special issues coordinated with other science journals. The ESSD success also exposes issues and challenges for present and future data publication, particularly around the topic and implementation of persistent identifiers.

- As ESSD encourages redundant data sets across multiple data centres for access and archive purposes, how will DOIs be employed to accurately point to those distributed or replicated data? How can authenticity and integrity be verified?
- How can or should object identifiers be employed in pointing from raw to quality-controlled and finally derived data processing levels; how can we designate or distinguish among these, particularly as those terms vary substantially among, for example, geophysical and ecological communities? Likewise, how to distinguish an auto-generated data product (e.g. a species identification from GBIF) from a high-effort expertly reviewed data product (e.g. an ESSD publication)?
- For a growing number of ESSD data publications with expected annual or periodic revisions and updates, how should data journals' and the repositories' use of persistent identifiers best record the subsequent versions, extensions or corrections?
- As published data sets become a valued part of high profile science, with attendant deadlines, announcements and publicity, do the various DOI policies and minting practices among cooperating publishers, data centres and journals represent a help or a hindrance?

These questions evolve directly from increasing interest in and activity by ESSD and, as such, represent 'new' and urgent challenges for data publication and data journals; we will describe real examples that drive each question. We urge an open and prompt discussion, perhaps under the auspices of a Research Data Alliance case study, among interested data centres, data managers, DataCite, and data and other journals. The expansion, and quality, of published data requires the effective resolution of these issues and the continued skilful, responsible and efficient use of DOIs or other identifiers.