



Data Publishing - View from the Front

David Carlson (1) and Hans Pfeiffenberger (2)

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

As data publishing journals - Earth System Science Data (ESSD, Copernicus, since 2009), Geophysical Data Journal (GDJ, Wiley, recent) and Scientific Data (SD, Nature Publishing Group, anticipated from May 2014) - expose data sets, implement data description and data review practices, and develop partnerships with data centres and data providers, we anticipate substantial benefits for the broad earth system and environmental research communities but also substantial challenges for all parties. A primary advantage emerges from open access to convergent data: subsurface hydrographic data near Antarctica, for example, now available for combination and comparison with nearby atmospheric data (both documented in ESSD), basin-scale precipitation data (accessed through GDJ) for comparison and interpolation with long-term global precipitation records (accessed from ESSD), or, imagining not too far into the future, stomach content and abundance data for European fish (from ESSD) linked to genetic or nutritional data (from SD). In addition to increased opportunity for discovery and collaboration, we also notice parallel developments of new tools for (published) data visualization and display and increasing acceptance of data publication as a useful and anticipated dissemination step included in project- and institution-based data management plans. All parties - providers, publishers and users - will benefit as various indexing services (SCI, SCOPUS, DCI etc.) acknowledge the creative, intellectual and meritorious efforts of data preparation and data provision. The challenges facing data publication, in most cases very familiar to the data community but made more acute by the advances in data publishing, include diverging metadata standards (among biomedical, green ocean modeling and meteorological communities, for example), adhering to standards and practices for permanent identification while also accommodating 'living' data, and maintaining prompt but rigorous review and evaluation processes in the face of unfamiliarity and overwhelming workloads.