Data Publications – persistent and citable products of research

Kirsten Elger[1], Kerstin Lehnert[2], & Damian Ulbricht[1]

[1] *GFZ German Research Centre for Geosciences, Germany*
[2] *Interdisciplinary Earth Data Alliance, Lamont-Doherty Earth Observatory, Columbia University, USA*

During the past decade, the relevance of research data has been rising significantly and the free and open access not only to scientific results, but to research datasets has been identified as a key issue by the scientific community, funding agencies and the public. As a consequence, there is a dynamic co-evolution of national and international guidelines on management of and open access to research data (e.g. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, G8 Open Data Charter, EU Horizon 2020 Guidelines, etc.) and the development of concepts to make data persistently accessible and citable products of research.

Especially in permafrost and climate research, long-term observatories and world-wide monitoring programmes are essential to understand the impact of, e.g., permafrost thaw on the Earth climate system and consequently of climate change. Many datasets are online available, via data portals or databases like, e.g. at the National Snow and Ice Datacentre, the GTN-P Database, NORPERM, Fluxnet, etc., but often without the possibility to give the data providers recognition and acknowledgement for contributing data to a global network.

During the past years, the publication of research datasets with assigned digital object identifier (DOI) has emerged as best practice for citable and persistent open access research data together with the deserved recognition of the data providers. An important step for the international acceptance and recognition of DOI-referenced data publication is the'Statement of Commitment'of the Coalition for Data Publication in the Earth and Space Sciences (COPDESS) that aims to promote joint policies and procedures for the publication and citation of data across Earth Science journals. Key commitments, signed by many publishers and data centres, are the acceptance of data citations within reference lists of research articles, the improvement of cross-references between journal articles and published datasets, and the strong recommendation to, whenever possible, store datasets in appropriate, theme-specific data repositories, in which data access and long-term preservation are guaranteed and datasets are accompanied by sufficient metadata to enable data reuse and discovery.

A comprehensive data description is essential for data reuse. Data publication is mostly known as supplementary material to scientific articles. Very often, however, the data description in the journal article is not sufficient for data reuse leading to many published dataset not being used as much as they could be. To improve this, many disciplines have developed and are developing "Data Journals" that aim to publish scientific articles with the detailed description of datasets, data portals or data collections that will be published along with the datasets. There are different formats for articles in data journals, but all have in common that they lack of any scientific interpretation of the described datasets. For standalone publication of datasets where neither format is an option, accompanying data reports are a convenient and flexible tool for enhanced data description.