10 years of data publication with Digital Object Identifiers (DOI) at the World Data Center for Climate (WDCC)

Martina Stockhause, Michael Lautenschlager, Heinke Höck, and Frank Toussaint
WDC Climate / DKRZ, Hamburg, Germany (stockhause@dkrz.de)

The World Data Center for Climate (WDCC) has been Data DOI publication agency for 10 years. Before issuing the first Data DOI, five years were spent on the development of a concept out of the idea to make scientific data citable. Apart from the unique and persistent identification of data objects, questions e.g. about the organizational structure including the roles of libraries and data centers, cost models, data quality, and the granularity of data publications were targeted in this pre-phase.

The first Data DOI publication system was developed within the German project 'std_doi' (Scientific and Technical Data DOI) funded by the German Research Foundation (DFG). The German National Library of Science and Technology (TIB) acted as DOI Registration Agency and the data centers PANGAEA, GFZ and WDCC as publication agencies and long-term data archives. The first publication of scientific data came from WDCC in March 2004 (doi:10.1594/WDCC/EH4_OPYC_SRES_A2). About five years after the first data publications, in December 2009, the international organization DataCite was formed, leading to several organizational as well as technical changes affecting WDCC's scientific data publication process. Recently, WDCC became the data publisher for the decentrally disseminated data of the project CMIP5 (Climate Model Intercomparison Project Phase 5) and participated in the development of the data publication system 'atarrabi', requiring several additions and changes in WDCC's established data publication procedure.

The challenges for the future of WDCC's scientific data publication lie in the connections of data to other DOI objects like scientific publications or to other persistent identifiers such as e.g. ORCID for persons or EPIC Handles for changeable data objects in the scientific workflow. Issues like data quality and data granularity have not been solved for the time being but are still under investigation.