EGU2020-8463
https://doi.org/10.5194/egusphere-egu2020-8463
EGU General Assembly 2020

AtMoDat: Improving the reusability of ATmospheric MOdel DATa with DataCite DOIs paving the path towards FAIR data

Daniel Neumann[1], Anette Ganske[2], Vivien Voss[3], Angelina Kraft[2], Heinke Höck[1], Karsten Peters[1], Johannes Quaas[4], Heinke Schluenzen[3], and Hannes Thiemann[1]

[1]Deutsches Klimarechenzentrum (DKRZ), Data Management, Bundesstr. 45a, Hamburg, Germany (daniel.neumann@dkrz.de)

[2]Technische Informationsbibliothek (TIB), Welfengarten 1 B, 30167 Hannover, Germany

[3]University of Hamburg, Meteorological Institute, Bundesstr. 55, 20146 Hamburg, Germany

[4]University of Leipzig, Leipzig Institute for Meteorology, Stephanstr. 3, 04103 Leipzig, Germany

The generation of high quality research data is expensive. The FAIR principles were established to foster the reuse of such data for the benefit of the scientific community and beyond. Publishing research data with metadata and DataCite DOIs in public repositories makes them findable and accessible (FA of FAIR). However, DOIs and basic metadata do not guarantee the data are actually reusable without discipline-specific knowledge: if data are saved in proprietary or undocumented file formats, if detailed discipline-specific metadata are missing and if quality information on the data and metadata are not provided. In this contribution, we present ongoing work in the AtMoDat project, -a consortium of atmospheric scientists and infrastructure providers, which aims on improving the reusability of atmospheric model data.

Consistent standards are necessary to simplify the reuse of research data. Although standardization of file structure and metadata is well established for some subdomains of the earth system modeling community – e.g. CMIP –, several other subdomains are lacking such standardization. Hence, scientists from the Universities of Hamburg and Leipzig and infrastructure operators cooperate in the AtMoDat project in order to advance standardization for model output files in specific subdomains of the atmospheric modeling community. Starting from the demanding CMIP6 standard, the aim is to establish an easy-to-use standard that is at least compliant with the Climate and Forecast (CF) conventions. In parallel, an existing netCDF file convention checker is extended to check for the new standards. This enhanced checker is designed to support the creation of compliant files and thus lower the hurdle for data producers to comply with the new standard. The transfer of this approach to further sub-disciplines of the earth system modeling community will be supported by a best-practice guide and other documentation. A showcase of a standard for the urban atmospheric modeling community will be presented in this session. The standard is based on CF Conventions and adapts several global attributes and controlled vocabularies from the well-established CMIP6 standard.

Additionally, the AtMoDat project aims on introducing a generic quality indicator into the DataCite

metadata schema to foster further reuse of data. This quality indicator should require a discipline-specific implementation of a quality standard linked to the indicator. We will present the concept of the generic quality indicator in general and in the context of urban atmospheric modeling data.