Meteorol. Z. (Contrib. Atm. Sci.), Vol. **31**, No. 6, 493–504 (published online August 2, 2022)

DKT12

Publication of Atmospheric Model Data using the ATMODAT Standard

Anette Ganske[1*], Angelika Heil[2], Andrea Lammert[2], Jan Kretzschmar[3] and Johannes Quaas[3]

[1]TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
[2]German Climate Computing Center (DKRZ), Hamburg, Germany
[3]Leipzig Institute for Meteorology, Universität Leipzig, Leipzig, Germany

(Manuscript received October 28, 2021; in revised form May 6, 2022; accepted May 17, 2022)

Abstract

Scientific data should be published in a way so that other scientists can benefit from these data, enabling further research. The FAIR Data Principles are defining the basic prerequisite for a good data publication: data should be Findable, Accessible, Interoperable, and Reusable. Increasingly, research communities are developing discipline-specific data publication standards under consideration of the FAIR Data Principles. A very comprehensive yet strict data standard has been developed for the climate model output within the Climate Model Intercomparison Project (CMIP), which largely builds upon the Climate and Forecast Metadata Conventions (CF conventions). There are, however, many areas of atmospheric modelling where data cannot be standardised according to the CMIP data standard because, e.g., the data contain specific variables which are not covered by the CMIP standard. Furthermore, fulfilling the strict CMIP data standard for smaller Model Intercomparison Projects (MIPs) requires much effort (in time and manpower) and hence the outcome of these MIPs often remains non-standardised. For innovative model diagnostics, preexisting standards are also not flexible enough. For that reason, the ATMODAT standard, a quality guideline for atmospheric model data, was created. The ATMODAT standard defines a set of requirements that aim at ensuring the high reusability of atmospheric model data publications. The requirements include the use of the netCDF file format, the application of the CF conventions, rich and standardised file metadata, and the publication of the data with a DataCite DOI. Additionally, a tool for checking the conformity of data and metadata to this standard, the *atmodat data checker*, was developed and is available on GitHub under an open licence. By using the more flexible ATMODAT standard, the publication of standardised datasets is simplified for smaller MIPs. This standardisation process is presented as an example using the data of an aerosol-climate model from the AeroCOM MIP. Furthermore, the landing pages of ATMODAT-compliant data publications can be highlighted with the EASYDAB logo. EASYDAB (Earth System Data Branding) is a newly developed quality label for carefully curated and highly standardised data publications. The ATMODAT data standardisation can easily be transferred to data from other disciplines and contribute to their improved reusability.

Keywords: AtMoDat, Data Publication, AeroCOM1, ECHAM6-HAM2, FAIR, Standardised Data, ATMODAT Standard, EASYDAB, Quality Guideline

1 Introduction

Research data publishing is a means through which research data are disseminated to a broad audience. Increasingly, funding agencies and journal publishers are making it a requirement that researchers publish the data which they created during a project or which they used for a scientific publication. However, making data accessible via a repository does not necessarily mean that they are easily reusable by others. To achieve this, the data publication must meet certain founding principles for good data management and stewardship, such as formulated in the FAIR Data Principles (Wilkinson et al., 2016). The FAIR principles advise data producers and publishers to improve the Findability, Accessibility, Interoperability, and Reusability of digital assets. Among others, they specify that data shall be described with sufficient human and machine-readable metadata and that data must be referenced by a persistent identifier (PID), such as a digital object identifier (DOI[1]).

The FAIR principles have been formulated on a general level. The interpretation of the FAIR principles and their implementation by individual research communities has led to the establishment of community-specific data standards that are often instrumental in determining which metadata elements are needed and how they should be described. The OGC Geopackage Encoding Standard (Open Geospatial Consortium, 2017) for the exchange of earth images and raster maps and the Technical Guidance for data in the INSPIRE Infrastructure (INSPIRE, 2007) are examples for such community-specific standards, as well as the standard for the results of the programme "Urban Climate Under

*Corresponding author: Anette Ganske, Technische Informationsbibliothek (TIB), Welfengarten 1B, 30167 Hannover, Germany, e-mail: anette.ganske@tib.eu

[1]https://www.doi.org/

Change [UC]^2" (Scherer et al., 2019; Scherer et al., 2020), which includes specifications for the data format, prescribes a special set of mandatory metadata elements and provides lists with controlled vocabulary for variables names, institutions, and locations in the field of urban climate research. Another research area, where specific data standards have been developed from early on, is global climate modelling. Within the framework of the Coupled Model Intercomparison Project (CMIP, Eyring et al., 2016), climate research institutions worldwide are performing coordinated model simulations and are exchanging, intercomparing, and analysing the model results. The assessments serve as a basis for the Intergovernmental Panel on Climate Change (IPCC[2]) reports which summarise the state of knowledge on anthropogenic climate change and which build the scientific foundations for global climate policy decisions. To facilitate data exchange and evaluation, the CMIP community has agreed upon a very comprehensive data standard that includes a common naming system for files, directories, variables, and metadata (Eyring et al., 2016; Juckes et al., 2020). The CMIP standard also specifies file format, data structure, and file content requirements. Among others, it defines a number of required and recommended "global attributes" which provide metadata that relate to the entire dataset. There are also mandatory metadata for describing the coordinates and the variables.

The network Common Data Format (netCDF) is defined as the standard file format for CMIP model output. NetCDF is an open, binary, platform-independent format that is particularly suitable for storing multidimensional gridded data along with grid definitions. By allowing for simultaneous storage of the actual data with their respective metadata, netCDF enables the creation of self-describing data files.

For model output generated within CMIP, the metadata for the variables in the respective dataset have to be compliant with the netCDF Climate and Forecast (CF) Metadata Conventions[3] (hereinafter referred to as CF conventions, Eaton et al., 2021). The CF conventions are an internationally agreed standard that defines what metadata are required to unambiguously describe variables in a dataset related to climate and weather forecasts.

The overall aim of the AtMoDat project[4] (Atmospheric Model Data: Data Quality, Curation Criteria and DOI Branding) was the enhancement of the FAIRness of published atmospheric model data. The assignment of a DataCite DOI, which fulfills an important aspect of the FAIR principles[5], is a prerequisite of the ATMODAT standard (Ganske et al., 2021). Therefore, this standard includes requirements for the metadata for the DOI and the corresponding landing pages. As it is also usual that data collections are published with a single DOI for the whole collection comprising several data files, these metadata can differ from the metadata describing the data in the individual files. Like the above-mentioned standards (Scherer et al., 2020; Open Geospatial Consortium, 2017; Breitbach et al., 2021) the ATMODAT standard also prescribes and recommends metadata elements for the individual data sets.

Regarding the R in FAIR the CMIP project is a role model as the CMIP data standard means highly standardized global climate model data, which are fit for reuse. Therefore a set of tools, software, and workflows is required which is based on CMIP controlled vocabularies (CV) – specific for the members of the CMIP community and the CMIP research questions. This high level of standardization within the CMIP project comes at the expense of high obstacles when applying it to datasets outside the intended scope of global climate modeling. For example, CMIP only allows for a predefined set of variables to be published under their standard and has furthermore relatively stringent requirements with regards to file metadata. Within the AtMoDat Project it was investigated to which extent the CMIP data standard can be transferred to or adapted for other areas of meteorology and climate research employing atmospheric models (Ganske et al., 2020). One example is the field of smaller "grassroots" model intercomparison projects (MIPs) in support of CMIP, such as the AeroCOM[6] MIPs. A key activity of AeroCOM is to assess how aerosol-climate effects are represented in different climate models and to evaluate how modelled aerosols compare to observations (Kinne et al., 2006). In this framework, new and not yet standardised model diagnostics are implemented and written to the model output files. Frequently, such output cannot be made fully compatible with the CMIP data standard.

A complete adaption of the CMIP data standard to data from urban climate research or smaller MIPs was not considered feasible; a leaner data standard was therefore developed. This reduced standard, the ATMODAT standard, is introduced in Section 2. Additionally, the Earth System Data Branding[7] (EASYDAB) was developed which certifies standardised datasets. EASYDAB is introduced in Section 3. The application of the ATMODAT standard to selected datasets from an AeroCOM MIP is described in Section 4. Finally, we provide a conclusion and outlook.

2 The ATMODAT standard

The ATMODAT standard (Ganske et al., 2021) is a quality guideline for the FAIR publication of atmospheric model data with open[8] licences. Its prerequisite is the publication of data sets with a DataCite DOI

[2] https://www.ipcc.ch/
[3] http://cfconventions.org
[4] https://www.atmodat.de/
[5] principle F1: (Meta) data are assigned globally unique and persistent identifiers
[6] https://aerocom.met.no
[7] https://www.easydab.de
[8] We use the definition of https://opendefinition.org/

in a repository. In addition, as recommended by DataCite[9], the DOI must always resolve to a web page, the so-called landing page. The DOI can be assigned to individual datasets, i.e. to one file of scientific data and its metadata. Alternatively, it can be assigned jointly to an entire dataset collection consisting of several datasets whose data were calculated, for example, within the same model simulation.

As some elements contained in the FAIR Data Principles are not clearly interpretable, complete compliance of published data to the FAIR principles is hard to achieve (see Dunning et al., 2017). To seek adherence of the published data to the FAIR Data Principles, the ATMODAT standard specifies requirements for the file format which is netCDF, the metadata, and the landing page.

2.1 Requirements for the metadata

Metadata must describe the data such that potential data users are able to decide whether these data are useful for their application. For this purpose, it must be sufficiently documented how the raw data were generated and how they were post-processed.

For data publication with a DataCite DOI, metadata are needed in the files, for the DOI, and on the landing page. All necessary information can either be written directly into the metadata or be referenced with links to external documents - preferably via persistent identifiers (PIDs). Such an external document could be, for example, the documentation of the numerical model used to generate the data. These external documents, or at least their metadata, should be machine-readable. As DataCite provides all DOI metadata both in machine-readable and machine-interpretable forms, it is ensured that automated lists can be created from these metadata, e.g. for retrieving all data publications of a research institute on a specific topic.

The ATMODAT standard specifies the following principles to enable machine readability and interpretability and thus FAIRness of all metadata:

- The language of the metadata is English.
- Whenever possible, all information on persons and institutions should be supplemented with a PID, i.e. an ORCID for a person or a ROR for an institution. Suggestions for suitable PIDs can be found in Madden et al. (2020).
- All links to documents should be persistently specified, e.g. with a DOI.
- All temporal information must be provided in a standardised form, e.g. according to ISO 8601-1 (2019) or ISO 19108 (2002).
- The spatial reference system must be specified, e.g. WGS84[10].
- The use of controlled vocabulary (CV) is recommended. Geographical names should be looked up in geonames[11], research areas in the Australian and New Zealand Standard Research Classification[12]. CF standard names[13] should be used as variable attributes whenever applicable. Other keywords can possibly be found in one of the recommended vocabularies, which are listed in the supplement. Nevertheless, a corresponding word from a CV cannot be found for all keywords, and sometimes it is cumbersome to search for a matching vocabulary. Some repositories (e.g. Pangaea[14]) use Terminology Services such as the one of GFBio[15], to offer guidance for choosing suitable terms. Unfortunately, there exists no Terminology Service for the entire Earth System Sciences.

2.1.1 File metadata

The CF conventions specify how to standardise variable metadata in order to make them automatically processable. The ATMODAT standard requires adherence to the CF conventions (version 1.4 or higher). A central element of the CF conventions are *standard_names*, which are unique identifiers for variables. Available *standard_names* are listed in the CF standard name table[16]. For each *standard_name*, the table contains a precise description and a default unit recommendation. However, *standard_names* are not defined for all variables of atmospheric models. In this case, a user-defined *long_name* must be assigned to the variable which shortly describes the respective variable.

In addition to the compliance with the CF conventions, the ATMODAT standard requires that netCDF files contain a description of the time, coordinate and vertical axes, as well as certain global attributes, e.g. title, author, conventions, contact. Table 1 specifies the most important attributes, the full list with details can be found in (Ganske et al., 2021). Even though the ATMODAT standard demands for using as many of these attributes as possible, not all attributes are applicable for all types of atmospheric model results. Therefore, if a recommended attribute is not used because it is not applicable, the metadata is still in accordance with the ATMODAT standard.

2.1.2 DOI metadata

DOI metadata that describe the published data are submitted to DataCite for the DOI. The DOI metadata are

[9] https://datacite.org/
[10] https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs84
[11] https://www.geonames.org
[12] https://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA2574180004463E?opendocument
[13] http://cfconventions.org/Data/cf-standard-names/72/build/cf-standard-name-table.html
[14] https://pangaea.de
[15] https://terminologies.gfbio.org/
[16] https://cfconventions.org/Data/cf-standard-names/78/build/cf-standard-name-table.html

Table 1: Attributes for the Data Files and their requirements for the ATMODAT Standard (M = mandatory, R = recommended). Optional arguments are not included in this table

Attribute	Description	Requirement
contact	A contact person and contact details. Does not need to be the creator. Reasonable to provide contact details that are valid in long term.	R
Conventions	Conventions that were followed, e.g. "CF-1.8 ATMODAT-3.0"	M
creation_date	Date of dataset's creation, e.g. 2022-01-01	R
creator	Person(s), who created the dataset.	R
crs	Coordinate reference system	R
frequency	Sampling frequency	R
geospatial_lat_resolution	If data was calculated on a regular grid: latitudal grid points' distance	R
geospatial_lon_resolution	If data was calculated on a regular grid: longitudal grid points' distance	R
geospatial_vertical_resolution	If data was calculated on a regular grid with vertical extension: grid points' distance in vertical direction	R
history	List of modifications to the original data.	R
institution	The name of the institution principally responsible for originating this data.	M
institution_id	Abbreviation of institution name and preferably the ROR of the institution	R
keywords	Keywords describing the data, preferably taken from a controlled vocabulary.	R
license	Must be an open license, e.g. CC-By 4.0	R
nominal_resolution	Approximate horizontal resolution, if data has not been calculated on a regular grid.	R
realm	Compartment(s) of the model, e.g. atmosphere, ocean, ...	R
source	Source of data: model name and version	M
standard_name_vocabulary	The name and version of the controlled vocabulary from which variable standard names are taken, e.g. "CF Standard Name Table v27"	R
summary	Short description of the dataset.	R
title	A short phrase or sentence describing the dataset.	R

entered into the metadata fields of the DataCite metadata schema (DataCite Metadata Working Group, 2019). If a joint DOI has been assigned for several datasets (dataset collection), the DOI metadata describe the dataset collection.

In addition to the general principles for metadata, the ATMODAT standard formulates additional requirements:

- The DataCite metadata field *Creator* must name all persons involved in the production of the dataset. As well, at least the contact person and preferably other contributors to the production of the dataset, e.g. the data curator or the data manager, must be mentioned in the field *Contributor*. This enables the automatic creation of summaries of the publication of a single researcher, of all researchers in an institution or of all publications within a project. For the same reasons, it is strongly recommended to indicate the funding under *FundingReference* (if applicable).
- All time information about the creation or publication of the dataset or dataset collection must be captured in the DOI's metadata.
- The DataCite metadata field *Subject* must include "AtMoDat", the research area, and more keywords describing the dataset. Additionally, it must contain information about the realm(s) (compartment) of the model used to produce the data (e.g. atmosphere, ocean, ...), preferably taken from the CMIP6 list[17].
- Related publications and datasets should be cited with the help of the DataCite metadata field *RelatedIdentifier*, such as the model documentation, all publications for which the data was used, the data that was used for the boundary conditions, datasets that were calculated with the same model, ...
- The DataCite metadata field *Description* should contain the abstract, which briefly describes the data. Additionally, all important information about the

[17] https://github.com/WCRP-CMIP/CMIP6_CVs/blob/master/CMIP6_realm.json

Table 2: Elements which should be included in the DataCite metadata field *Description* and their requirements for the ATMODAT standard (M = mandatory, R = recommended)

Element	Description	Requirement
Abstract	Short description of the dataset.	M
Additional simulation time information	E.g. temporal aggregation for output	R
Calender used	Gregorian or other calender	R
Description of grid	E.g. rectangular grid or triangular grid.	R
Model's name	Name of the model that was used to calculate the data.	M
Model's version	Version of the model that was used to calculate the data.	R
Horizontal resolution of the model	If data was calculated on a regular grid: grid points' distances	R
Geographic reference system used by the model	E.g. WGS84	R
Vertical coordinate of model	If data was calculated on a regular grid with vertical extension: gridpoints' distance in vertical direction	R
Spatial coverage of model	Applicable, if models spatial coverage differs from the spatial coverage of the archived data.	R
Basic approximations	Approximations which are used in equations solved by the model's code, e.g. the hydrostatic approximation.	R
Possible usage of the data	Motivation to calculate the data; for what can the data be used; degree of accurracy; any known problems with the data	R

data, which cannot be written in other DataCite metadata fields, should be added to the description, see Table 2.

A list of all available DataCite metadata fields can be found in Table 3. It is strongly recommended to use all metadata fields from the DataCite metadata schema that are applicable to describe the data. This approach enforces the self-explanatory description of the data. Further information about the DOI metadata fields can be found in the ATMODAT standard (Ganske et al., 2021) and in the DataCite Metadata Schema (DataCite Metadata Working Group, 2019).

2.2 Requirements for the landing page

If datasets provided with a DOI should comply with the FAIR principles, certain requirements for the design of a landing page have to be met. Both, the landing page and the subordinate web pages must be publicly and permanently available. Nevertheless, the layout of the landing page may change and metadata may be added.

DataCite DOIs are always connected to an HTML landing page, which is human- and machine-readable. Resolving a DOI (together with the URL of the DOI-Server[18]) in a conventional web browser will redirect to this landing page, which must contain information on how to access the data. It is a requirement of a DataCite DOI that the landing page must always be accessible, even if the dataset is no longer available. In this case, the so-called Tombstone Page contains the metadata of the dataset and the information, that the dataset is no longer available.

The ATMODAT standard requires that all metadata fields, specified for the DOI, are listed on the human-readable part of the landing page. The names of the individual metadata fields do not have to be used, e.g. *Creator* might be called Author or Provider by the publishing repository. Additional metadata fields, which are often important to describe atmospheric model data, but for which no DataCite metadata field exists (see Table 4), must also be written on the landing page. In addition, for dataset collections or datasets with several variables, the descriptions of the individual datasets or variables should also appear on the landing page (see Table 5). If the datasets or dataset collections have been checked for compliance with a quality guideline before publication, the result of these checks should also be noted on the landing page.

The machine-readable part of the landing page (the page source) should at least also contain all metadata for the DataCite DOI. It should be provided compliant to schema.org[19] in order to enable search engines, such as Google Dataset Search or Bing, to extract relevant information for their search algorithms from the individual web pages. Alternatively, an equivalent structure based on the W3C DCAT format (Data Catalog Vocabulary Albertoni et al., 2020) may be used, see e.g. the data description from Google[20].

[18] https://doi.org/

[19] https://schema.org/

[20] https://developers.google.com/search/docs/data-types/dataset

Table 3: DataCite metadata fields and their requirements for the ATMODAT standard (M = mandatory, R = recommended)

DataCite metadata field	Description	Requirement
Identifier	DOI of the dataset	M
AlternateIdentifier (with type sub-property)	Applicable if the dataset/dataset collection was assigned another PID.	R
Contributor(s)	Each Contributor's family name, given name	M
+ sub-properties	Each Contributor's name identifier (e.g. ORCID) and affiliation.	R
Creator(s)	Each Creator's family name, given name	M
+ sub-properties	Each Creator's name identifier (e.g. ORCID) and affiliation	R
Date of creation or update	Date of creation or update of the dataset/dataset-collection.	M
+ other date-types	Available: end of an embargo; Collected: date, when e.g. data were measured, Valid: Time range of the model simulation.	R
Description	Short description of the dataset.	M
File Format	Always netCDF	M
Funding Reference	Funder's name, identifier, and award related information	R
GeoLocation	Location of the data (e.g. model area, coordinates of single points).	R
Language	Always english.	M
Publisher	Name and PID (ROR) of the publisher	M
PublicationYear	Year of dataset's publication	M
RelatedIdentifier	Related information about documentation, boundary conditions, publications citing the dataset, other related datasets, and maturity tests. Use PIDs, if applicable.	R
ResourceType	For ATMODAT always "Dataset"	M
Rights	An open licence that is at least available in English language such as CC-BY 4.0	M
Size	Size of the dataset or dataset collection.	R
Subject	"AtMoDat", research area, model's realm, and additional keywords describing the data, preferably taken from a controlled vocabulary (CV).	M
+ sub-properties	If keywords are chosen from a CV, then Name and URI of the CV.	R
Title	A short phrase or sentence describing the dataset	M
Version	Version of the dataset/dataset collection.	R

Table 4: Additional metadata fields, which should/must be written on the Landing Page and their requirements (M = mandatory, R = recommended)

Metadata field	Requirement
Model	M
Model version	R
Grid	R
Projection	R
Vertical Coordinate	R
Temporal Coverage	R
Spatial Coverage	R
Basic Approximations	R

Table 5: If the DOI is assigned to a Dataset collection, these metadata fields should be additionally displayed on the landing page, if applicable (R = recommended)

Metadata field	Requirement
Variable/ Dataset Name	R
Temporal Aggregation	R
Spatial Aggregation	R
Dimension	R
Size	R

2.3 Ensuring compliance with requirements

The ATMODAT standard is aimed at repositories (in the area of data curation), but also at the scientific community (in the area of data production). It contains checklists that support data producers and data curators in controlling whether data are compliant with the ATMODAT standard. The *atmodat data checker* was developed to facilitate and, above all, automate this checking process, even for large amounts of data. The checker is a modular Python program package, which was published un-

der an open licence on Github[21]. Thus, the *atmodat data checker* is helpful for the production and publication of standardised data. An example of its use is provided in Chapter 4. A corresponding checker for checking DOI metadata is in preparation.

3 EASYDAB

The Earth System Data Branding (EASYDAB) was developed within the AtMoDat project to highlight specific Earth System Science datasets that have been published with a DataCite DOI: The EASYDAB logo on the landing page indicates that the datasets have an open licence and accordance with the FAIR Data Principles has been sought. The FAIRness of the datasets could be described e.g. as the result of one of the existing FAIR-tests, like F-UJI FAIRness Test[22] or ARDC FAIR Self Assessment tool[23]. Given the fact that the FAIR principles are not indicating strict rules or one specific test, EASYDAB recommends following the FAIR principles as much as possible for the responsible repository. Furthermore, the EASYDAB logo indicates that the datasets have been checked by the responsible repository for compliance with a quality guideline (Ganske et al., 2022). For example, datasets that comply with the requirements of the ATMODAT standard can be published with EASYDAB. Alternatively, repositories may use their own quality guidelines for EASYDAB publications, provided that they achieve a comparably high quality of published data as with the ATMODAT standard.

Many quality guidelines have been written for the exchange of standardised data in data portals, e.g. the guideline for data from the [UC]^2 research programme (Scherer et al., 2020), the technical guidance for data in the INSPIRE Infrastructure (INSPIRE, 2007), and the OGC Geopackage Encoding Standard (Open Geospatial Consortium, 2017) for the exchange of earth images and raster maps. These guidelines do not only prescribe and recommend the needed metadata, but some also specify file types or how metadata should be published. The CF conventions (Eaton et al., 2021) and the Attribute Convention for Data Discovery (ACCD[24]) both recommend how to describe metadata in netCDF datasets. Other standards like ISO 19115-1 (2014) and ISO 19115-2 (2019) prescribe the semantics of metadata fields and are widely used in other standards (e.g. by the INSPIRE guidance). The Data Stewardship Maturity Matrix for the NOAA OneStop Project (Peng et al., 2019) and the Quality Maturity Matrix (Höck et al., 2020) of the German Climate Computing Center (DKRZ) are examples, of how the curation of the datasets is standardised and documented so that the FAIRness of the data sets increases with higher levels of maturity. Nevertheless, in contrast to the ATMODAT standard, most of these guidelines and curation processes do not presume the assignment of a PID to the data, open[25] file formats and licences. Therefore, it depends on the repository whether it does not only curate its data following a specific quality guideline but also assigns DataCite DOIs and follows additional internal regulations, e.g. about the structure and content of the landing pages. Only in this case, a repository can highlight its landing pages with the EASYDAB logo.

The intentions of these quality guidelines are the standardisation of datasets, the enhancement of the quality of metadata, and the increase of FAIRness. Nevertheless, the quality of data is more than that and the quality of data and metadata should be assessed and documented at each production step of the data production cycle (Peng et al., 2021b). Even though a precondition for an EASYDAB branding is only high quality metadata, it would be a surplus if the curation process for a dataset would have included checks for all the four different aspects of quality (Scientific, Product, Stewardship, and Service), as defined in Ramapriyan et al. (2017).

The EASYDAB logo is protected. In order to use it, repositories must conclude a contract with the German National Library of Science and Technology (TIB[26]). In this contract, the respective repositories undertake to display the EASYDAB logo only on the landing pages of the datasets/dataset collections which comply with the EASYDAB guidelines.

According to (Peng et al., 2021a) is "Knowledge about the quality of data and metadata important to support informed decisions on the (re)use of individual datasets". Therefore, datasets with enhanced quality of (meta)data should be labelled so that they can stand out of the crowd of all datasets. This is the purpose of EASYDAB. Using the EASYDAB logo, repositories can indicate that they carefully curate datasets with FAIRness principles in mind and make them reusable. It is easier for users to find and evaluate well-described datasets and to use data that is relevant to them. In collaboration with the World Data Center for Climate (WDCC[27]), the first dataset (Löwe, 2022) was published with EASYDAB.

4 Application of the ATMODAT standard

After having outlined the requirements necessary to publish a dataset in accordance with the ATMODAT standard, we will provide an example of the publication process. The example data originate from simulations with the global aerosol-climate model ECHAM6-HAM2 (Zhang et al., 2012; Tegen et al., 2019) and

[21] https://github.com/AtMoDat/atmodat_data_checker; Version 1.2 published as (Kretzschmar et al., 2022)
[22] https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool
[23] https://ardc.edu.au/fair-data/fair-self-assessment-tool/
[24] https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3
[25] we use the definition of open from https://opendefinition.org/
[26] see www.easydab.de
[27] https://cera-www.dkrz.de/WDCC/ui/cerasearch/

were created within an AeroCOM MIP. These simulations aim at providing a process-based observational constraint on the cloud lifetime effect (Albrecht, 1989; Gryspeerdt et al., 2019; Bellouin et al., 2020) by investigating the parameterised precipitation stemming from liquid-only clouds ("warm rain") in global climate models. Precipitation from liquid-phase clouds is particularly important in the subtropics and tropics, whereas in the extra-tropics, rain usually is formed via the ice phase (Mülmenstädt et al., 2015). The formation of rain in liquid-phase clouds is mediated by the presence of aerosols in the atmosphere. An increase in aerosol concentration decreases the efficiency of cloud condensate removal by precipitation, which increases cloud lifetime.

To provide a process-based observational constraint on cloud lifetime influenced by "warm rain" occurrence, the ECHAM6-HAM2 simulations were compared to space-borne cloud radar observations (Mülmenstädt et al., 2020, 2021). To ensure an apples-to-apples comparison, the Cloud Feedback Model Intercomparison Project Observations Simulator Package (COSP; Bodas-Salcedo et al., 2011; Kay and National Center for Atmospheric Research Staff (Eds.), 2019; Kretzschmar et al., 2019) was implemented into the model that provided the same variables as available from the space-borne cloud radar. The employed satellite simulator subdivides a model gridbox into so-called subcolumns, making each subcolumn comparable in size to a satellite pixel (Pincus and Klein, 2000). While comparing model output with satellite observations, it is essential to represent the subgrid-scale variability of cloud properties (i.e., cloud cover and hydrometeors), as the model grid size is much larger than that of the satellite data.

A subsequent analysis of the model simulations not only requires the standard model output like temperature and mass concentrations of cloud condensate and hydrometeors but also the output from the radar simulator on these subcolumns. A particular feature of the subcolumn output is that it is 5-dimensional (with time, model level, subcolumn, latitude, and longitude as the respective dimensions). In particular, the sampling of the subgrid-scale variability via the subcolumns is not a standard output of an atmospheric model (Figure 1), and standard tools as well as postprocessing software are not able to handle more than the typical four dimensions (three-dimensional in space, plus time). This is thus an example where CMIP standardisation fails. Here, the flexibility of the ATMODAT standard comes into play, as it is compatible with such non-standard output, while at the same time ensuring FAIRness of the data. To avoid overly large netCDF files from the 4- and 5-dimensional variables, each variable was extracted from the raw model output and was furthermore split up into data files with monthly slices.

The ATMODAT standard defines mandatory global attributes for each data file and further demands compliance with the CF conventions. Such basic conformity

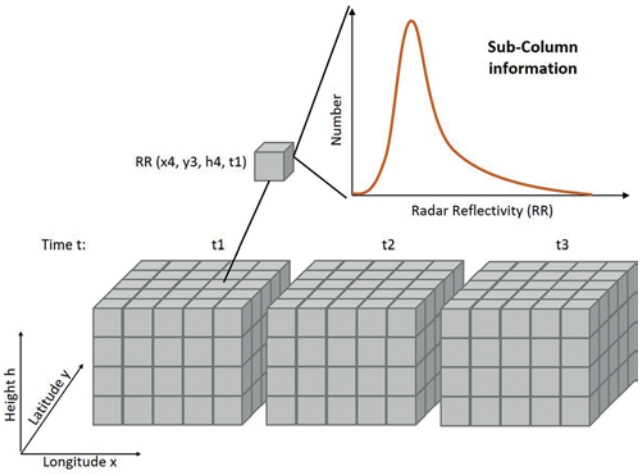

Figure 1: Schematic illustration of 5-dimensional output from radar simulator for ECHAM6-HAM2 global simulations. The regular 4-dimensional model output has the dimensions longitude, latitude, height, and time (boxes in the front). Additional to the specific value of the variable RR(x,y,h,t) each gridbox includes as fifth dimension subcolumn information resulting from the COSP satellite observation simulations (shown above as distribution function for RR).

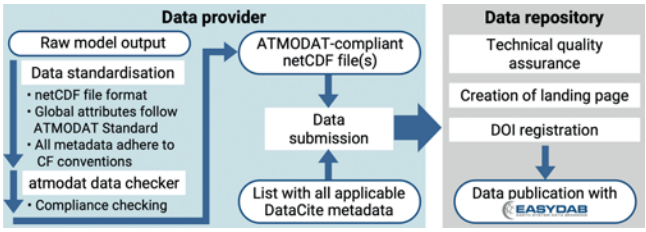

Figure 2: Workflow diagram illustrating the steps that lead from the raw model data to an EASYSAB data publication.

can easily be achieved by smaller MIPs or even single researchers. Following the ATMODAT standard allows them to publish their datasets in an easier way compared to the CMIP standard, which has much more stringent requirements and lacks the flexibility needed for publishing FAIR model data outside the scope of global climate modelling. In the following, we will discuss the workflow to achieve compliance with the ATMODAT standard.

As stated above, the ATMODAT standard requires a data file to follow at least CF conventions version 1.4 or greater. For this dataset, we chose to directly go to the latest release, which was version 1.8 at the time of data preparation. Before submitting the datasets to the respective repositories, the data provider has to ensure that their datasets comply with the ATMODAT standard using the *atmodat data checker* (Figure 2). Possible inconsistencies with the ATMODAT standard are reported by this checker in a human and machine-readable format. While the former is an aid for the data provider to bring their datasets into accordance with the ATMODAT standard, the latter allows repositories for an automated evaluation of the quality of the submitted datasets. In

a)

```
=== Short summary ===

ATMODAT Standard Compliance Checker Version: 1.2.0
Checking against: ATMODAT Standard 3.0, CF Version 1.8

Number of checked files: 1
Mandatory checks passed: 2/4 (1 missing, 1 error(s))
Recommended checks passed: 2/20 (17 missing, 1 error(s))
Optional checks passed: 0/9 (9 missing, 0 error(s))
CF checker errors: 2
CF checker warnings: 4
```

b)

```
=== Short summary ===

ATMODAT Standard Compliance Checker Version: 1.2.0
Checking against: ATMODAT Standard 3.0, CF Version 1.8

Number of checked files: 1
Mandatory checks passed: 4/4 (0 missing, 0 error(s))
Recommended checks passed: 20/20 (0 missing, 0 error(s))
Optional checks passed: 5/9 (4 missing, 0 error(s))
CF checker errors: 0
CF checker warnings: 0
```

Figure 3: Exemplary short summary of the *atmodat data checker* for a single file before (a) and after (b) modification for compliance with the ATMODAT standard.

Figure 3a, an example of the so-called short summary output by the *atmodat data checker* for an exemplary file of the dataset is shown. Besides the results of the checks on the global attributes, additional information on the CF compliance of the file is provided.

From this short summary output, we find that this dataset is not compliant with the CF Conventions version 1.8 in its present state, as the CF checker reports some errors. Furthermore, another prerequisite of the ATMODAT standard is not fulfilled at the moment as mandatory attributes, required by the ATMODAT standard, are missing and/or contain errors. Such errors indicate that a global attribute is not in the right format or that the content of an attribute does not stem from a predefined controlled vocabulary. The major issues regarding the CF checker errors are related to the description of the vertical coordinate variables used to store data on model levels. These vertical coordinates were updated using the Climate Data Operators[28], developed by the Max Planck Institute for Meteorology in Hamburg. A further issue with the dataset is that some variables do not contain sufficient variable attributes by default, as either a *standard_name* or a *long_name* is missing. Simple python scripts have been developed by the AtMoDat team and are distributed with the *atmodat data checker* which were used to fill missing variable attributes. Furthermore, these scripts were used to fill missing global attributes in the netCDF files. Apart from the mandatory attributes required by the ATMODAT standard, we provided recommended and optional global attributes whenever possible. As mentioned in Section 2, not all optional global attributes can be fulfilled by each dataset. The whole dataset preparation is an iterative process and multiple runs of the *atmodat data checker* might be necessary to achieve compliance with the ATMODAT standard. A more detailed output for the respective checks is provided as an output of the checker (not shown), to support the preparation of data files for compliance with the ATMODAT standard. The *atmodat data checker* can also be used to check all netCDF files in a directory at once.

Figure 3b shows the output of the *atmodat data checker* after modifying it for compliance the ATMODAT standard. We have used the python scripts distributed with the *atmodat data checker* to fill missing attributes or revise wrongly filled attributes. After that step, the exemplary data file fulfills all mandatory and recommended global attributes and is compliant with the CF Conventions. In this particular example, there are still some optional attributes indicated as missing which is due to the fact that those attributes are not applicable for this dataset. At the time of publication, the respective datasets were in the final stages of the publication process and are expected to be available in near future.

A DataCite DOI is a prerequisite for a dataset that should be published in accordance with the ATMODAT standard. In the specific example, the WDCC was chosen as the repository. WDCC has a long history in the preservation and publication of climate data and their detailed description for reuse. In advance of a data publication, metadata have to be prepared by the data provider with advice from the data curator (Figure 2). The comprehensive technical quality assurance of the data by the WDCC curator regarding the ATMODAT standard will clearly be simplified by the use of the *atmodat data checker*. The metadata ingests, quality assurance, and data ingests will be followed by the publication and an additional check regarding the ATMODAT standard's requirements for the landing page. Subsequently, the DataCite DOI for the dataset will be allocated. Data published this way are fulfilling all requirements of EASYDAB and will be labelled with the EASYDAB logo.

5 Conclusion and outlook

Atmospheric model data are an indispensable basis for the investigation of the Earth's atmosphere. Climate model data provide information, for example, on how atmospheric processes change on a global or regional scale in a changing climate.

Within CMIP, extensive requirements were established to standardise the data of global and regional climate model simulations, which enables automatic processing. Within the AtMoDat project, these standardisations are adapted and generalised to data of various types

[28] https://code.mpimet.mpg.de/projects/cdo/

of atmospheric models, e.g. high-resolution models or aerosol-climate models. As a result, the ATMODAT standard was developed which provides precise recommendations to achieve enhanced reusability of atmospheric model data that are published in repositories.

EASYDAB, the Earth System Data Branding, was created to enable high-quality datasets such as those conforming to the ATMODAT standard to be clearly distinguishable in repositories. The EASYDAB logo helps users to quickly identify and select these datasets. Easy access to standardised, quality-checked research data not only enables more efficient reuse. It also leads to increased user confidence in the repository that provides this data, and so, the EASYDAB branding also reinforces the value of the DataCite DOI. The longer-term goal is that quality-assured data publications, due to their high benefits, are increasingly appreciated and demanded by users as well as funding organisations.

An application of the ATMODAT standard is demonstrated for datasets from the AeroCOM model intercomparison project. First ATMODAT standard-compliant datasets, including DataCite DOIs, are being published in the World Data Center for Climate (WDCC). An example is the recent data publication of a weather type catalogue for the North Sea (Löwe, 2022). For the publication, the original ASCII time series files were converted to CF-compliant netCDF files that contain rich metadata to render each file stand-alone self-describing. For filling the metadata, the instructions and recommendations laid down in the ATMODAT standard were followed and their proper implementation was verified with the *atmodat data checker*. Prior to submitting the files to the WDCC, all required metadata for the landing page and the DOI are compiled in compliance with the ATMODAT standard requirements and provided with the data files in the final submission. Prior to publishing the data, the WDCC curators re-examine all metadata for compliance with the ATMODAT standard. If certain recommended elements of the ATMODAT standard are not fulfilled, the curator consults with the data provider and requests a short note with the reasonings. The note together with the short summary output of the *atmodat data checker* are then published on the landing page of the data publication.

The standardisation procedures developed in AtMoDat project can easily be transferred to other disciplines and can, if necessary, be extended. The developed *atmodat standard checker* supports both, the data provider and the repository, in verifying compliance with the ATMODAT standard. Due to its simple framework, this tool can furthermore be adapted to check compliance of data sets with other discipline-specific standards.

In summary, the ATMODAT standard and EASYDAB both contribute towards enhancing the data standardisation in various research areas of the Earth System Sciences. They represent an essential contribution to improving the overall FAIRness of research data and facilitate the intra- and interdisciplinary data exchange.

Acknowledgments

The AtMoDat project is funded within the framework of "Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten" by the German Federal Ministry of Education and Research (BMBF; FKZ: 16QK02A, 16QK02B, 16QK02C, 16QK02D). We thank the DataCite team, the CF conventions community and the other AtMoDat partners for fruitful discussions. The publication of this article was funded by the Open Access Fund of Technische Informationsbibliothek (TIB).

Supplement

List of Controlled Vocabularies

Controlled vocabulary for fields of science, places, and other keywords can be found in

1. **Field of science:** https://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/6BB427AB9696C225CA2574180004463E?opendocument
2. **Realm of the model:** https://github.com/WCRP-CMIP/CMIP6_CVs/blob/master/CMIP6_realm.json
3. **Geonames:** https://www.geonames.org/
4. **EU Vocabularies (includes continents, countries and places):** https://op.europa.eu/en/web/eu-vocabularies/home
5. **Geographic identifiers for marine regions:** https://www.marineregions.org/
6. **Names of Variables:** CF-Conventions http://cfconventions.org/Data/cf-standard-names/72/build/cf-standard-name-table.html
7. **Description of Data:** http://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.pdf
8. **United Nations Terminology Database:** https://unterm.un.org/
9. **Global Change Master Directory (GCMD):** https://earthdata.nasa.gov/earth-observation-data/find-data/gcmd/gcmd-keywords
10. **Climate Tagger:** https://www.climatetagger.net/climate-thesaurus/
11. **Marine keywords:** https://vocab.nerc.ac.uk/search_nvs/
12. **Environnmental keywords:** https://www.eionet.europa.eu/gemet/

13. **Glossary of the Australien Bureau of Meteorology:** http://www.bom.gov.au/lam/glossary/
14. **Australien Bureau of Meteorology – Weather Words:** http://www.bom.gov.au/info/wwords/
15. **International Glossary of Hydrology:** https://unesdoc.unesco.org/ark:/48223/pf0000221862
16. **CGI Observation Method vocabulary:** https://vocabs.ardc.edu.au/viewById/89

More lists can be found in BARTOC (Basel Register of Thessauri, Ontologies and Classifications), see http://www.bartoc.org/en.

References

Albertoni, R., D. Browning, S. Cox, G.B. A., A. Perego, P. Winstanley, 2020: Data catalog vocabulary (DCAT). – https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/ (accessed 2022-02-18).

Albrecht, B., 1989: Aerosols, cloud microphysics, and fractional cloudiness. – Science **245**, 1227–1230, DOI: 10.1126/science.245.4923.1227.

Bellouin, N., J. Quaas, E. Gryspeerdt, S. Kinne, P. Stier, D. Watson-Parris, O. Boucher, K. Carslaw, M. Christensen, A.L. Daniau, J.L. Dufresne, G. Feingold, S. Fiedler, P. Forster, A. Gettelman, J.M. Haywood, U. Lohmann, F. Malavelle, T. Mauritsen, D. McCoy, G. Myhre, J. Mülmenstädt, D. Neubauer, A. Possner, M. Rugenstein, Y. Sato, M. Schulz, S.E. Schwartz, O. Sourdeval, T. Storelvmo, V. Toll, D. Winker, B. Stevens, 2020: Bounding global aerosol radiative forcing of climate change. – Rev. Geophys. **58**, e2019RG000660, DOI: 10.1029/2019RG000660.

Bodas-Salcedo, A., M.J. Webb, S. Bony, H. Chepfer, J.L. Dufresne, S.A. Klein, Y. Zhang, R. Marchand, J.M. Haynes, R. Pincus, V.O. John, 2011: COSP: Satellite simulation software for model assessment. – Bull. Amer. Meteor. Soc. **92**, 1023–1043, DOI: 10.1175/2011BAMS2856.1.

Breitbach, G., B. Geyer, U. Kleeberg, E. Meyer, R. Onken, P.S. Sommer, 2021: Binding regulations for storing data as netCDF files. – HCDC, https://gitlab.hzdr.de/hcdc/hereon-netcdf/hereon-netcdf-en (accessed 2022-05-03).

DataCite Metadata Working Group, 2019: DataCite metadata schema documentation for the publication and citation of research data. Version 4.3. – https://schema.datacite.org/meta/kernel-4.3/, DOI: 10.14454/f2wp-s162.

Dunning, A., M. de Smaele, J. Böhmer, 2017: Are the FAIR Data Principles fair?. – IJDC **12**, 177–195, DOI: 10.2218/ijdc.v12i2.567.

Eaton, D., J. Gregory, B. Drach, K. Taylor, S. Hankin, J. Blower, J. Caron, R. Signell, P. Bentley, G. Rappa, H. Höck, A. Pamment, M. Juckes, M. Raspaud, R. Horne, T. Whiteaker, D. Blodgett, C. Zender, D. Lee, D. Hassell, A.D. Snow, T. Kölling, D. Allured, A. Jelenak, A. Meier Soerensen, L. Gaultier, S. Herlédan, 2021: NetCDF Climate and Forecast (CF) Metadata Conventions. – https://cfconventions.org/Data/cf-conventions/cf-conventions-1.9/cf-conventions.html (accessed 2022-1-31).

Eyring, V., S. Bony, G.A. Meehl, C.A. Senior, B. Stevens, R.J. Stouffer, K.E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. – GGeosci. Model Dev. **9**, 1937–1958, DOI: 10.5194/gmd-9-1937-2016.

Ganske, A., D. Heydebreck, H. Höck, A. Kraft, J. Quaas, A. Kaiser, 2020: A short guide to increase fairness of atmospheric model data. – Meteorol. Z. (Contrib. Atm. Sci.) **29**, 483–491, DOI: 10.1127/metz/2020/1042.

Ganske, A., A. Kraft, A. Kaiser, D. Heydebreck, A. Lammert, H. Höck, H. Thiemann, V. Voss, D. Grawe, B. Leitl, K.H. Schlünzen, J. Kretzschmar, J. Quaas, 2021: ATMODAT Standard (v3.0). – World Data Center for Climate (WDCC) at DKRZ, DOI: 10.35095/WDCC/atmodat_standard_en_v3_0.

Ganske, A., A. Kraft, M. Burger, H. Thiemann, A. Lammert, A. Kaiser, A. Heil, H. Hoeck, 2022: EASYDAB Guideline (v1.1). – World Data Center for Climate (WDCC) at DKRZ, DOI: 10.35095/WDCC/EASYDAB_Guideline_v1.1.

Gryspeerdt, E., T. Goren, O. Sourdeval, J. Quaas, J. Mülmenstädt, D.S., C. Unglaub, A. Gettelman, M. Christensen, 2019: Constraining the aerosol influence on cloud liquid water path. – Atmos. Chem. Phys. **19**, 5331–5347, DOI: 10.5194/acp-19-5331-2019.

Höck, H., F. Toussaint, H. Thiemann, 2020: Fitness for Use of Data Objects Described with Quality Maturity Matrix at Different Phases of Data Production.. – Data Sci. J. **19**, 45, DOI: 10.5334/dsj-2020-045.

INSPIRE, 2007: Temporary MIG subgroup for action MIWP-8:Technical Guidance for the implementation of INSPIRE dataset and service metadata based on ISO/TS 19139:2007. – http://inspire.ec.europa.eu/id/document/tg/metadata-iso19139/2.0.

ISO 19108, 2002: Geographic information – Temporal schema. – International Organization for Standardization, https://www.iso.org/.

ISO 19115-1, 2014: Geographic information – Metadata – Part 1: Fundamentals. – International Organization for Standardization, https://www.iso.org/.

ISO 19115-2, 2019: Geographic information – Metadata – Part 2: Extensions for acquisition and processing. – International Organization for Standardization, https://www.iso.org/.

ISO 8601-1, 2019: Date and time – Representations for information interchange – Part 1: Basic rules. – International Organization for Standardization, https://www.iso.org/.

Juckes, M., K.E. Taylor, P.J. Durack, B. Lawrence, M.S. Mizielinski, A. Pamment, J.Y. Peterschmitt, M. Rixen, S. Sénési, 2020: The CMIP6 data request (DREQ, version 01.00.31). – Geosci. Model Develop. **13**, 201–224, DOI: 10.5194/gmd-13-201-2020.

Kay, J., National Center for Atmospheric Research Staff (Eds.), 2019: The Climate Data Guide: COSP: Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package. – CLIMATE DATA GUIDE, https://climatedataguide.ucar.edu/climate-data/cosp-cloud-feedback-model-intercomparison-project-cfmip-observation-simulator-package.

Kinne, S., M. Schulz, C. Textor, S. Guibert, Y. Balkanski, S.E. Bauer, T. Berntsen, T.F. Berglen, O. Boucher, M. Chin, W. Collins, F. Dentener, T. Diehl, R. Easter, J. Feichter, D. Fillmore, S. Ghan, P. Ginoux, S. Gong, A. Grini, J. Hendricks, M. Herzog, L. Horowitz, I. Isaksen, T. Iversen, A. Kirkevåg, S. Kloster, D. Koch, J.E. Kristjansson, M. Krol, A. Lauer, J.F. Lamarque, G. Lesins, X. Liu, U. Lohmann, V. Montanaro, G. Myhre, J. Penner, G. Pitari, S. Reddy, O. Seland, P. Stier, T. Takemura, X. Tie, 2006: An aerocom initial assessment – optical properties in aerosol component modules of global models. – Atmos. Chem. Phys. **6**, 1815–1834, DOI: 10.5194/acp-6-1815-2006.

Kretzschmar, J., M. Salzmann, J. Mülmenstädt, J. Quaas, 2019: Arctic clouds in ECHAM6 and their sensitivity to cloud

microphysics and surface fluxes. – Atmos. Chem. Phys. **19**, 10571–10589, DOI: 10.5194/acp-19-10571-2019.

Kretzschmar, J., A. Heil, A. Kaiser, 2022: Atmodat/atmodat_data_checker: Version 1.3. – Zenodo, DOI: 10.5281/zenodo.6118655.

Löwe, P., 2022: Lamb weather types (reduced set) and gale days over the north sea since 1948 based on ncep/ncar reanalysis 1 daily mean sea level pressure fields. – WDC CLIMATE, DOI: 10.26050/WDCC/LambWTyRSetAndGaleDaysOverTheNo.

Madden, F., R. van Horik, S. van de Sandt, A. Lavasa, H. Cousijn, 2020: Guides to choosing persistent identifiers – version 2. – Zenodo, DOI: 10.5281/zenodo.3956569.

Mülmenstädt, J., O. Sourdeval, J. Delanoë, J. Quaas, 2015: Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from a-train satellite retrievals. – Geophys. Res. Lett. **42**, 6502–6509, DOI: 10.1002/2015GL064604.

Mülmenstädt, J., C. Nam, M. Salzmann, J. Kretzschmar, T.S. L'Ecuyer, U. Lohmann, P.L. Ma, G. Myhre, D. Neubauer, P. Stier, K. Suzuki, M. Wang, J. Quaas, 2020: Reducing the aerosol forcing uncertainty using observational constraints on warm rain processes. – Sci. Advan.**6**, published online, DOI: 10.1126/sciadv.aaz6433.

Mülmenstädt, J., M. Salzmann, J.E. Kay, M.D. Zelinka, P.L. Ma, C. Nam, J. Kretzschmar, S. Hörnig, J. Quaas, 2021: An underestimated negative cloud feedback from cloud lifetime changes. – Nature Climate Change **11**, 508–513, DOI: 10.1038/s41558-021-01038-1.

Open Geospatial Consortium, 2017: OGC GeoPackage Encoding Standard. – Open Geospatial Consortium, http://www.opengis.net/doc/IS/geopackage/1.2 Version 1.2.

Peng, G., A. Milan, N. Ritchey, R. Partee II, S. Zinn, E. McQuinn, K. Casey, P. Lemieux III, R. Ionin, P. Jones, A. Jakositz, D. Collins, 2019: Practical Application of a Data Stewardship Maturity Matrix for the NOAA OneStop Project. – Data Sci. J. 1–18, DOI: 10.5334/dsj-2019-041.

Peng, G., R. Downs, C. Lacagnina, H. Ramapriyan, I. Ivánová, D. Moroni, Y. Wei, G. Larnicol, L. Wyborn, M. Goldberg, J. Schulz, I. Bastrakova, A. Ganske, L. Bastin, S. Khalsa, M. Wu, C.L. Shie, N. Ritchey, D. Jones, T. Habermann, C. Lief, I. Maggio, M. Albani, S. Stall, L. Zhou, M. Drévillon, S. Champion, C. Hou, F. Doblas-Reyes, K. Lehnert, E. Robinson, K. Bugbee, 2021a: Call to action for global access to and harmonization of quality information of individual earth science datasets. – Data Sci. J. **20**, 19, DOI: 10.5334/dsj-2021-019.

Peng, G., C. Lacagnina, R. Downs, A. Ganske, H. Ramapriyan, I. Ivánová, L. Wyborn, D. Jones, L. Bastin, C.L. Shie, D. Moroni, 2021b: Global community guidelines for documenting, sharing, and reusing quality information of individual digital datasets. – Data Sci. J. **21**, published online, DOI: 10.5334/dsj-2022-008.

Pincus, R., S.A. Klein, 2000: Unresolved spatial variability and microphysical process rates in large-scale models. – J. Geophys. Res. Atmos. **105**, 27059–27065, DOI: 10.1029/2000JD900504.

Ramapriyan, H., G. Peng, D. Moroni, C. Shie, 2017: Ensuring and improving information quality for earth science data and products… – D-Lib Magazine **23**, published online, DOI: 10.1045/july2017-ramapriyan.

Scherer, D., F. Antretter, S. Bender, J. Cortekar, S. Emeis, U. Fehrenbach, G. Gross, G. Halbig, J. Hasse, B. Maronga, S. Raasch, K. Scherber, 2019: Urban Climate Under Change $[UC]^2$ – A National Research Programme for Developing a Building-Resolving Atmospheric Model for Entire City Regions. – Metorol. Z. **28**, 95–104, DOI: 10.1127/metz/2019/0913.

Scherer, D., U. Fehrenbach, T. Grassmann, A. Holtmann, F. Meier, K. Scherber, D. Pavlik, T. Höhne, F. Kanani-Sühring, B. Maronga, F. Ament, S. Banzhaf, I. Langer, G. Halbig, M. Kohler, R. Queck, S. Stratbücker, M. Winkler, R. Wegener, M. Zeeman, 2020: $[UC]^2$-Datenstandard "Stadtklima im Wandel". – http://www.uc2-program.org/uc2_datenstandard.pdf Version 1.4.1.

Tegen, I., D. Neubauer, S. Ferrachat, C. Siegenthaler-Le Drian, I. Bey, N. Schutgens, P. Stier, D. Watson-Parris, T. Stanelle, H. Schmidt, S. Rast, H. Kokkola, M. Schultz, S. Schroeder, N. Daskalakis, S. Barthel, B. Heinold, U. Lohmann, 2019: The global aerosol–climate model ECHAM6.3–HAM2.3 – Part 1: Aerosol evaluation. – Geosci. Model Develop. **12**, 1643–1677, DOI: 10.5194/gmd-12-1643-2019.

Wilkinson, M.D., M. Dumontier, I. Aalbersberg, J. Jsbrand, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. t'Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, 2016: The FAIR guiding principles for scientific data management and stewardship. – Scientific data **3**, 160018, DOI: 10.1038/sdata.2016.18.

Zhang, K., D. O'Donnell, J. Kazil, P. Stier, S. Kinne, U. Lohmann, S. Ferrachat, B. Croft, J. Quaas, H. Wan, S. Rast, J. Feichter, 2012: The global aerosol-climate model ECHAM-HAM, version 2: sensitivity to improvements in process representations. – Atmos. Chem. Phys. **12**, 8911–8949, DOI: 10.5194/acp-12-8911-2012.