

# The ATMODAT Standard enhances FAIRness of Atmospheric Model data

Daniel Heydebreck<sup>1</sup>, Amandine Kaiser<sup>1</sup>, Anette Ganske<sup>2</sup>, Angelina Kraft<sup>2</sup>, Heinke Schlutzen<sup>3</sup>, and Vivien Voss<sup>3</sup>

<sup>1</sup>German Climate Computing Centre (DKRZ)

<sup>2</sup>Technische Informationsbibliothek (TIB)

<sup>3</sup>University of Hamburg (UHH)

November 28, 2022

## Abstract

Within the AtMoDat project (Atmospheric Model Data, [www.atmodat.de](http://www.atmodat.de)), a standard has been developed which is meant for improving the FAIRness of atmospheric model data published in repositories. Atmospheric model data form the basis to understand and predict natural events, including atmospheric circulation, local air quality patterns, and the planetary energy budget. Such data should be made available for evaluation and reuse by scientists, the public sector, and relevant stakeholders. Atmospheric modeling is ahead of other fields in many regards towards FAIR (Findable, Accessible, Interoperable, Reusable, see e.g. Wilkinson et al. (2016, doi:10.1101/418376)) data: many models write their output directly into netCDF or file formats that can be converted into netCDF. NetCDF is a non-proprietary, binary, and self-describing format, ensuring interoperability and facilitating reusability. Nevertheless, consistent human- and machine-readable standards for discipline-specific metadata are also necessary. While standardisation of file structure and metadata (e.g. the Climate and Forecast Conventions) is well established for some subdomains of the earth system modeling community (e.g. the Coupled Model Intercomparison Project, Jukes et al. (2020, <https://doi.org/10.5194/gmd-13-201-2020>)), other subdomains are still lacking such standardisation. For example, standardisation is not well advanced for obstacle-resolving atmospheric models (e.g. for urban-scale modeling). The ATMODAT standard, which will be presented here, includes concrete recommendations related to the maturity, publication, and enhanced FAIRness of atmospheric model data. The suggestions include requirements for rich metadata with controlled vocabularies, structured landing pages, file formats (netCDF), and the structure within files. Human- and machine-readable landing pages are a core element of this standard and should hold and present discipline-specific metadata on simulation and variable level.

# The ATMODAT Standard enhances FAIRness of Atmospheric Model data

Daniel Heydebreck [1], Amandine Kaiser [1], Anette Ganske [2], Angelina Kraft [2], Heinke Schlunzen [3], Vivien Voss [3]

[1] German Climate Computing Centre (DKRZ), [2] Technische Informationsbibliothek (TIB), [3] University of Hamburg (UHH)



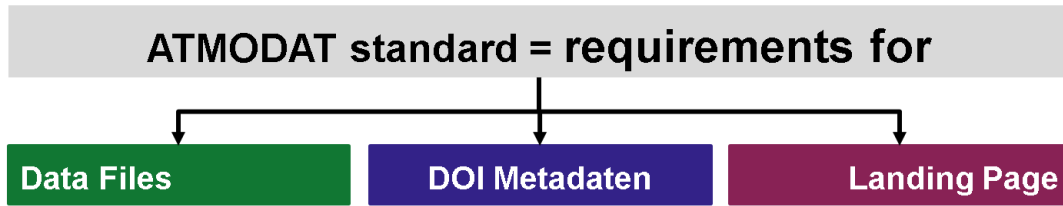
## INTRODUCTION

Atmospheric model data form the basis to understand and predict natural atmospheric events, including atmospheric circulation, local air quality pattern, and the planetary energy budget. High-quality data of this type should be made available for evaluation and reuse by scientists, the public sector, and relevant stakeholders.

A standard has been developed within the **AtMoDat** project (Atmospheric Model Data), which is meant for improving the **FAIR**ness of atmospheric model data published in repositories. Additionally, the standard requires open data (**open licenses**).

A precondition for this standard is the publication of data with a **DataCite** DOI. Nevertheless, these recommendations are also applicable to other PIDs, if they are associated with metadata.

## OVERVIEW STANDARD



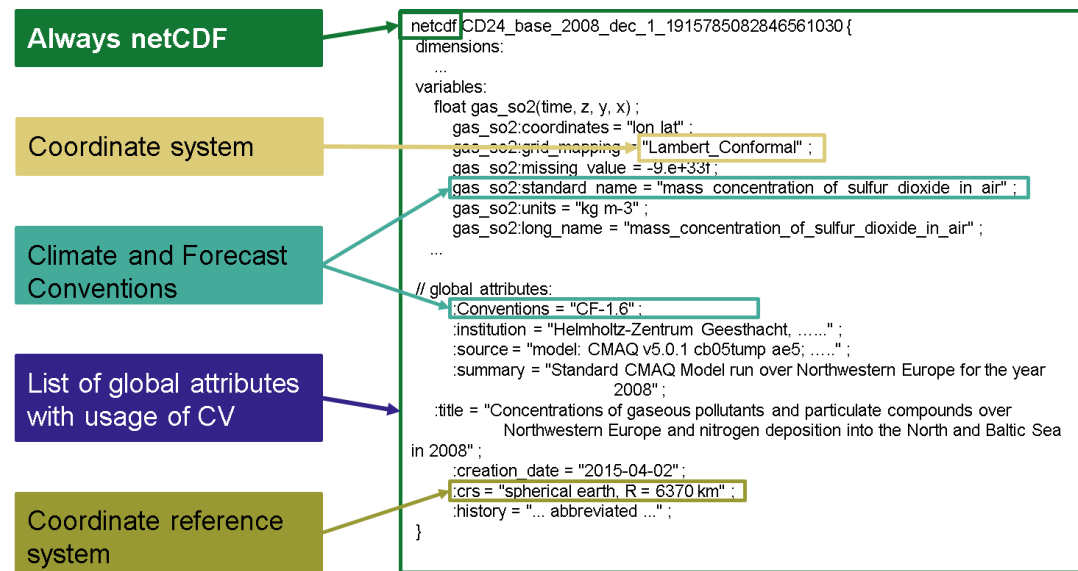
The ATMODAT standard is easy to implement and provides checklists for data curators and data producers. A dataset that complies with this standard will follow the FAIR principles, and its metadata will be of high quality.



If all requirements are fulfilled, the EASYDAB label can be assigned to this dataset or dataset collection. Thereby, indicating the high standardization and formalization.

# METADATA STANDARDS FOR FILES AND DOI

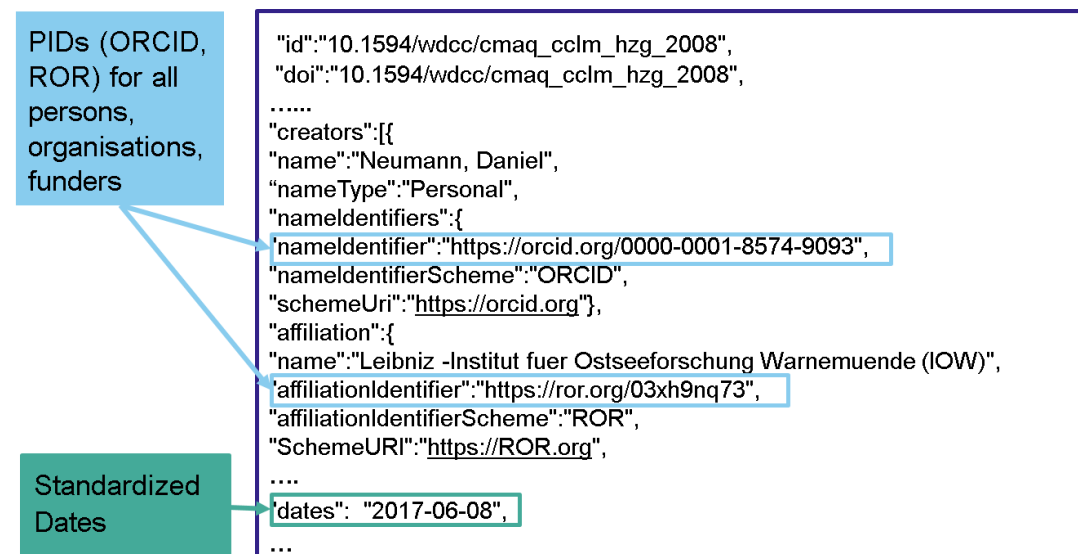
## Data Files



NetCDF Climate and Forecast Metadata Conventions: **CF-Conventions**

## Metadata for DataCite DOIs

Use all available **metadata properties**



Always describe all performed maturity (quality) tests, see our **maturity poster**.

Recommended PIDs (Persistent Identifiers):

Persons: **ORCID**

Documents: **DOI**

Institutions: **ROR**

Funders: **Funder registry**

Standardized Dates: **ISO8601**

[https://en.wikipedia.org/wiki/ISO\\_8601](https://en.wikipedia.org/wiki/ISO_8601)

# LANDING PAGE

## Landing Page (human-readable)

- must be publicly and permanently available
- includes a complete citation of the dataset
- contains the DOI itself

Sub-pages with details of datasets or variables, if one DOI is assigned to a dataset-collection

Branding Logo

Use of controlled vocabulary

Contact Person with ORCID

Information, where and how data can be downloaded

Spatial and temporal information for the data

All DOI-Metadata properties are listed!

## Landing Page (machine-readable)

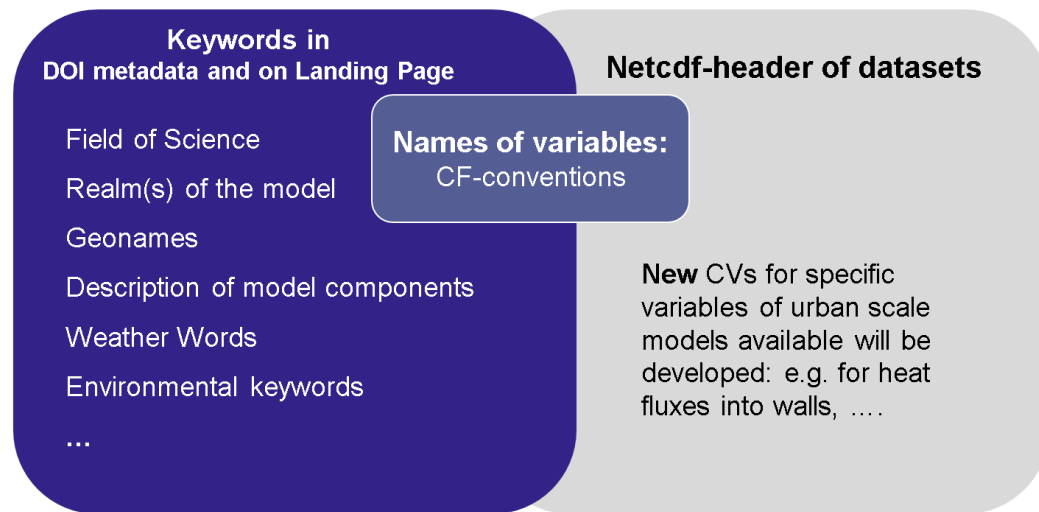
Machine interpretable language

PIDs (ORCID, ROR) for all persons, organisations, funders

All DOI-Metadata properties are listed!

# CONTROLLED VOCABULARIES (CVS)

## Controlled Vocabulary (CV)



Recommendations for new CVs for urban scale models are welcome → [user survey on a micro-scale model data standard](#)

### Controlled Vocabularies:

Field of Science: <https://www.oecd.org/science/inno/38235147.pdf>

Realm(s) of the model: [https://github.com/WCRP-CMIP/CMIP6\\_CVs/blob/master/CMIP6\\_realm.json](https://github.com/WCRP-CMIP/CMIP6_CVs/blob/master/CMIP6_realm.json)

Geonames: <https://www.geonames.org/>

Description of model components: <https://specializations.es-doc.org/static/index.html?target=cmip6.atmos&client=esdoc-url-rewrite>

Weather Words: <http://www.bom.gov.au/info/wwords/>

Environmental keywords: <https://www.eionet.europa.eu/gemet/en/themes/>

CF-conventions - standard names: <http://cfconventions.org/Data/cf-standard-names/72/build/cf-standard-name-table.html>



# SUMMARY

The ATMODAT standard contains concrete recommendations related to the publication and enhanced FAIRness of atmospheric model data. It gives instructions for rich metadata with controlled vocabularies, structured landing pages, file formats (netCDF), and the structure within files. Human- and machine-readable landing pages should hold and present discipline-specific metadata on simulation and variable level.

**Contact:** [info@atmodat.de](mailto:info@atmodat.de)

**Visit us:** <https://www.atmodat.de>

**Acknowledgments** AtMoDat is funded by the German Federal Ministry for Education and Research (BMBF) (FKZ: 16QK02A-D).



## AUTHOR INFORMATION

Daniel Heydebreck: heydebreck@dkrz.de; <https://orcid.org/0000-0001-8574-9093>

Amandine Kaiser: kaiser@dkrz.de; <https://orcid.org/0000-0002-2756-9964>

Anette Ganske: anette.ganske@tib.eu; <https://orcid.org/0000-0003-1043-4964>

Angelina Kraft: angelina.kraft@tib.eu; <https://orcid.org/0000-0002-6454-335X>

Heinke Schluenzen: heinke.schluenzen@uni-hamburg.de; <https://orcid.org/0000-0002-5711-547X>

Vivien Voss: vivien.voss@uni-hamburg.de; <https://orcid.org/0000-0001-7329-8192>

# ABSTRACT

Within the AtMoDat project (**Atmospheric Model Data**, [www.atmodat.de](http://www.atmodat.de)), a standard has been developed which is meant for improving the FAIRness of atmospheric model data published in repositories. Atmospheric model data form the basis to understand and predict natural events, including atmospheric circulation, local air quality pattern and the planetary energy budget. Such data should be made available for evaluation and reuse by scientists, the public sector, and relevant stakeholders.

Atmospheric modelling is ahead of other fields in many regards towards FAIR (Findable, Accessible, Interoperable, Reusable, see e.g. Wilkinson et al., 2016, doi:10.1101/418376) data: many models write their output directly into netCDF or file formats that can be converted into netCDF. NetCDF is a non-proprietary, binary and self-describing format, ensuring interoperability and facilitating reusability. Nevertheless, consistent human- and machine-readable standards for discipline-specific metadata are also necessary. While standardisation of file structure and metadata (e.g. the Climate and Forecast Conventions) is well established for some subdomains of the earth system modelling community (e.g. the Coupled Model Intercomparison Project, Juckes et al, 2020, doi:10.5194/gmd-13-201-2020), other subdomains are still lacking such standardisation. For example, standardisation is not well advanced for obstacle-resolving atmospheric models (e.g. for urban-scale modeling).

The ATMODAT standard, which will be presented here, includes concrete recommendations related to the maturity, publication and enhanced FAIRness of atmospheric model data. The suggestions include requirements for rich metadata with controlled vocabularies, structured landing pages, file formats (netCDF) and the structure within files. Human- and machine readable landing pages are a core element of this standard, and should hold and present discipline-specific metadata on simulation and variable level.