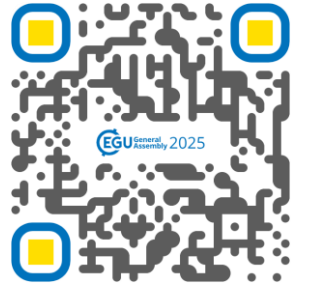


Harmonizing NetCDF Metadata Workflows

A Collaborative Initiative for Enhanced
Data Integration and Reusability



Abstract: EGU25-6484

Björn L. Saß et al.

Data manager & scientific programmer
Helmholtz-Zentrum Hereon

Geesthacht, Germany

DataHub
Research Field Earth & Environment

HCDC
Helmholtz Coastal Data Center



Prior info

- We don't want to create a new schema or CV!
- We aim to combine existing conventions and schemas
- For this talk:

Metadata = global attributes = attributes which describe the data set as a whole, on file level

- NetCDF: a self-describing container for multidimensional data

1. Why we want to harmonize Metadata workflows

Our scientists produce cool data,

1. Why we want to harmonize Metadata workflows

Our scientists produce cool data, without properly describing them,

```
// global attributes:
:ContactPerson_name = "Alexander Boehmlaender";

:ContactPerson_affiliation = Institute of Meteorology and Climate Research Atmosphere

:Conventions = "CF-1.12, ACDD-1.3, DataCite 4.6";
:creation_date = "2024-08-23T15:40:37+00:00";
:creator_name = "Alexander Boehmlaender";
:creator_email = "alexander.boehmlaender@kit.edu";

:creator_affiliation = "Karlsruhe Institute of Technology";

:history = "2024-05-10: raw_data analysed with PIA 2.0.2 (dev_custom_settings branch)";
:institution = "Karlsruhe Institute of Technology";

:title = "Level1 data from PINE-04-01 during PaCE22, heated total inlet, see also A
:references = "Moehler et al. 2021, doi:10.5194/amt-14-1143-2021";
:Researcher_name = "Alexander Boehmlaender";

:Researcher_affiliation = "Karlsruhe Institute of Technology";

:instrument = "PINE-04-01";
:station_name = "Pallas Atmosphere-Ecosystem Supersite";
```

1. Why we want to harmonize Metadata workflows

Our scientists produce cool data, without properly describing them, because of a lack of **guidance**.

```
// global attributes:  
:ContactPerson_name = "Alexander Boehmlaender";  
:ContactPerson_id = "https://orcid.org/0009-0007-3485-2139";  
:ContactPerson_affiliation = Institute of Meteorology and Climate Research Atmosphere  
:ContactPerson_affiliation_id = "https://ror.org/04t3en479";  
:Conventions = "CF-1.12, ACDD-1.3, DataCite 4.6";  
:creation_date = "2024-08-23T15:40:37+00:00";  
:creator_name = "Alexander Boehmlaender";  
:creator_email = "alexander.boehmlaender@kit.edu";  
:creator_id = "https://orcid.org/0009-0007-3485-2139";  
:creator_affiliation = "Karlsruhe Institute of Technology";  
:creator_affiliation_id = "https://ror.org/04t3en479";  
:history = "2024-05-10: raw_data analysed with PIA 2.0.2 (dev_custom_settings branch)";  
:institution = "Karlsruhe Institute of Technology";  
:license = "CC-BY-SA 4.0";  
:source = "surface observation";  
:featureType = "timeSeries";  
:title = "Level1 data from PINE-04-01 during PaCE22, heated total inlet, see also A  
:references = "Moehler et al. 2021, doi:10.5194/amt-14-1143-2021";  
:Researcher_name = "Alexander Boehmlaender";  
:Researcher_id = "https://orcid.org/0009-0007-3485-2139";  
:Researcher_affiliation = "Karlsruhe Institute of Technology";  
:Researcher_affiliation_id = "https://ror.org/04t3en479";  
:instrument = "PINE-04-01";  
:station_name = "Pallas Atmosphere-Ecosystem Supersite";
```

1. Why we want to harmonize Metadata workflows

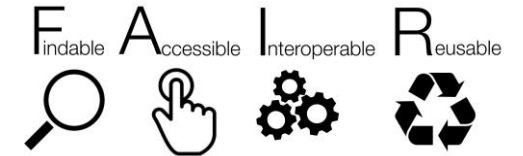
Harmonized Metadata in NetCDF files help us to improve data **FAIR**ness.

Metadata enrichment improves...

...machine readability

- good for Data and Metadata Repositories
- good for Software engineering and Algorithms

...human readability and data discovery



Findable

- Identifier
- Rich metadata

Accessible

- (Meta-)data online
- Explained restrictions

Interoperable

- Standards & vocabularies
- Common format

Reusable

- Rich documentation
- Licensing & Provenance

2. How to tackle the problem

The problem: A lack of **guidance**.

The solution: A collaborative initiative...

2. How to tackle the problem

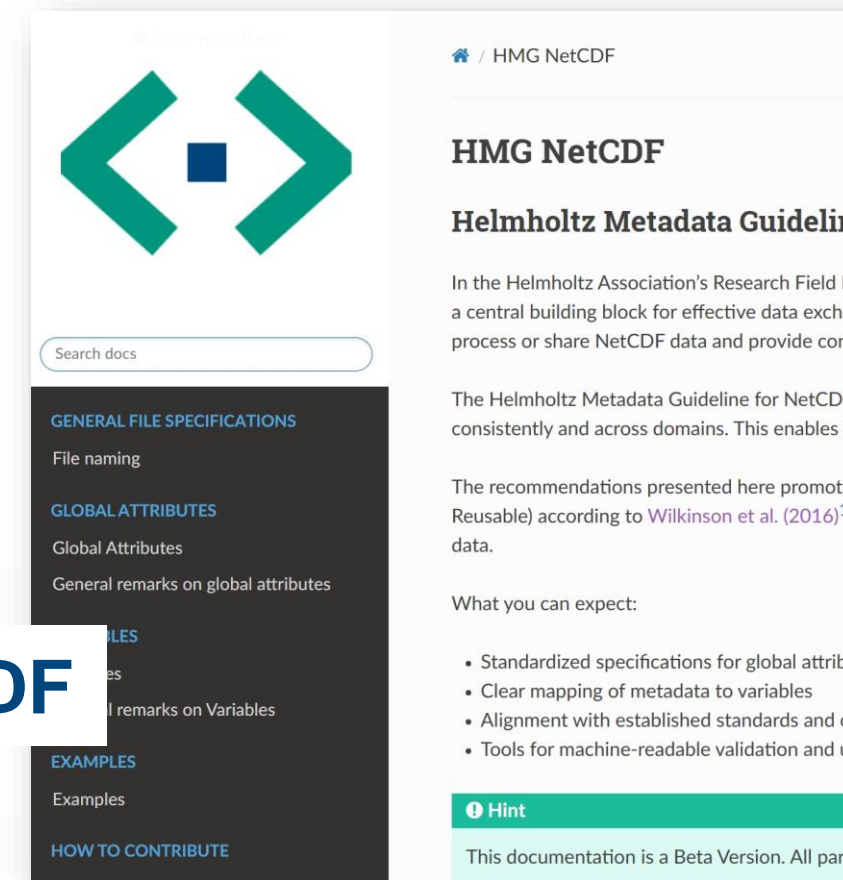
The problem: A lack of **guidance**.

The solution: A collaborative initiative...

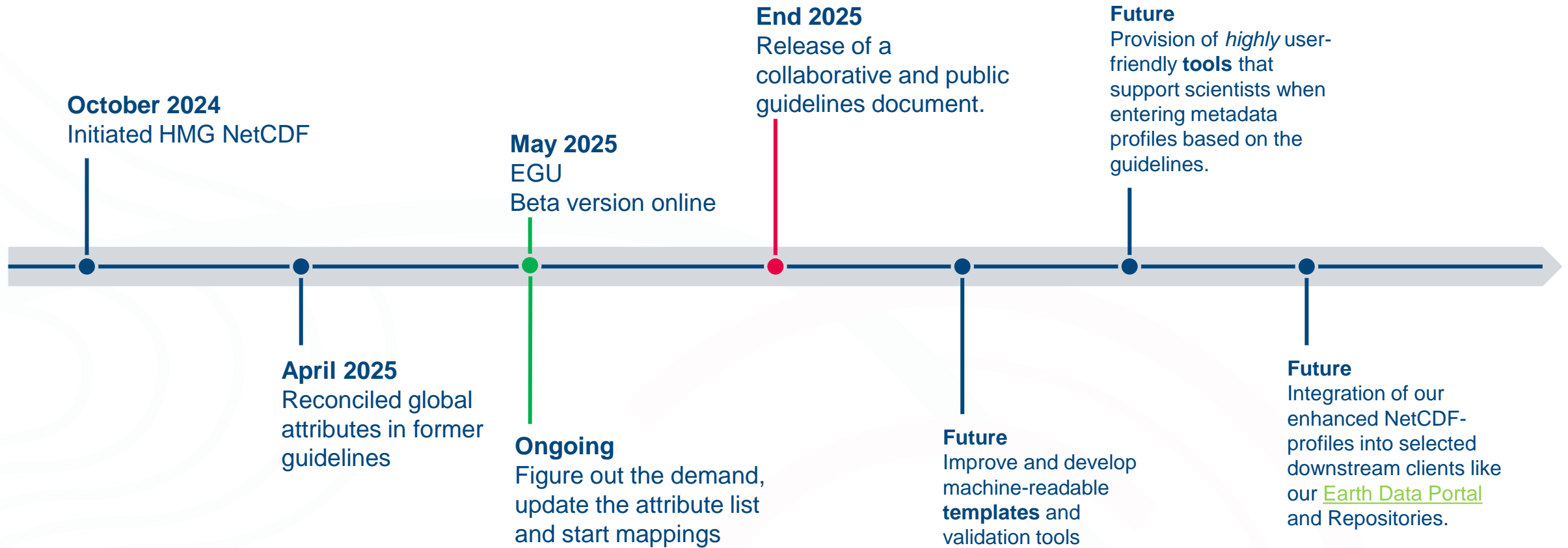
...which develops the

(Helmholtz) Metadata Guidelines for NetCDF

short: HMG NetCDF



3. The solution: Metadata guidelines



4. Challenges overcome during the first harmonization milestone

1. Get everyone around the table and start harmonizing approaches.
2. Create a living document spirit, to overcome the problem of never-ending discussions.
3. Involve Scientists and ask for their input and their demands.

5. Positive side effects of the harmonizing process

- We started **mappings** between discovery schemas.
 - CF is the foundation
 - ACDD is a good start but lacks precise identifiers
 - DataCite seems to fill the lack of attributes holding identifiers
- We collect and connect to ongoing but hidden discussions in the community, task forces and project groups.



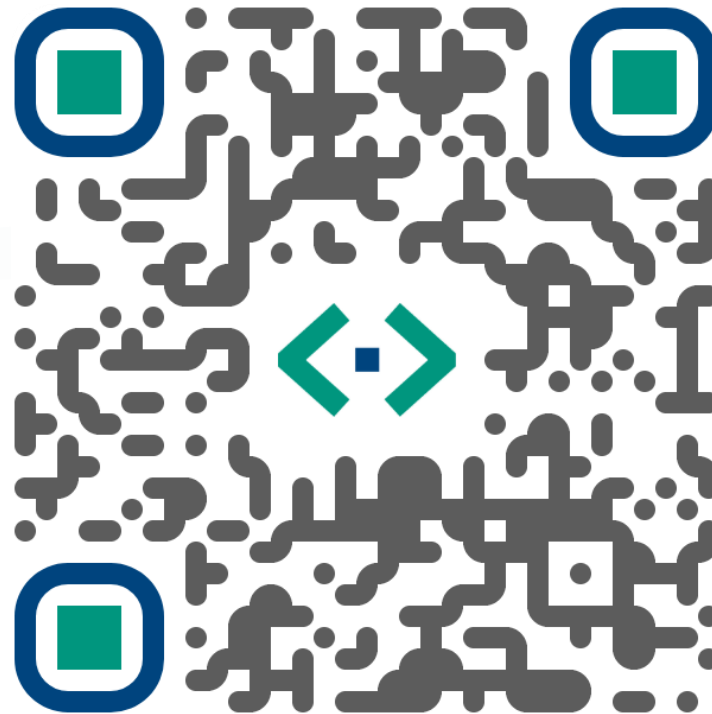
6. We invite you to participate



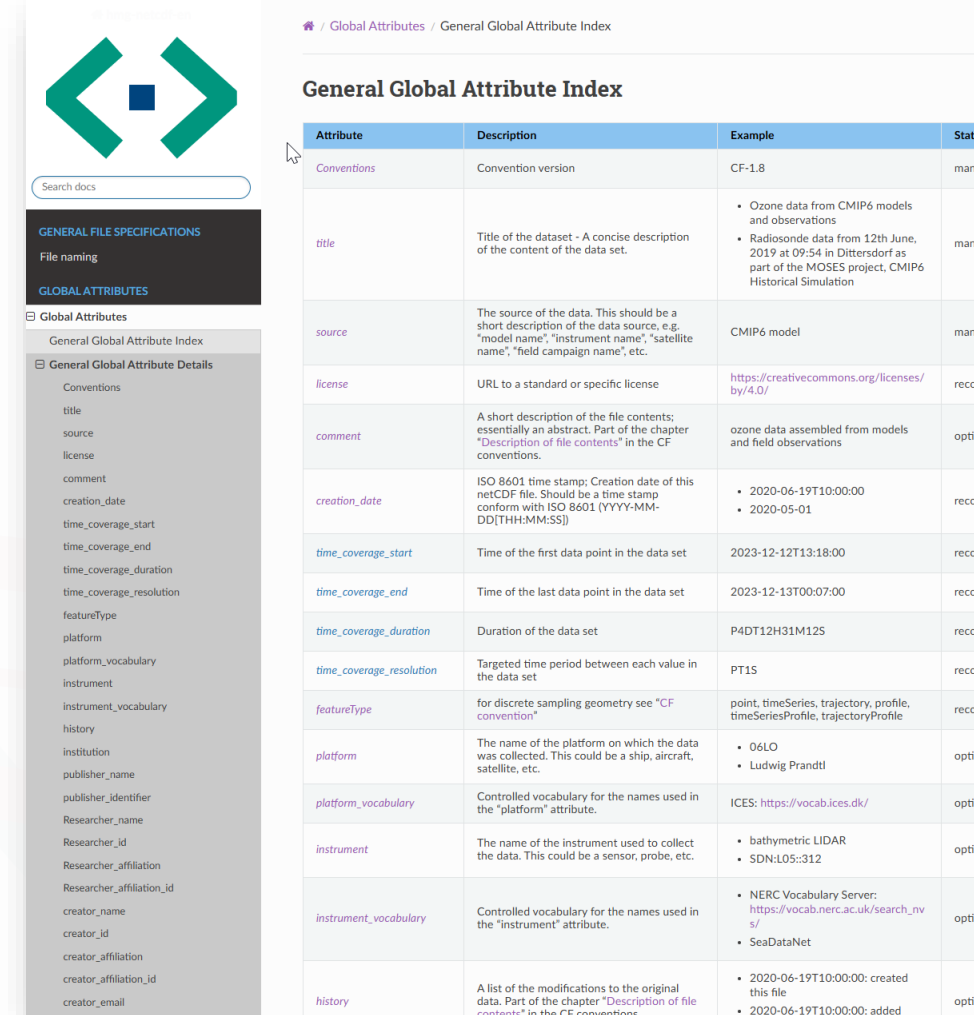
Version beta-0.1.0 is online

And we are happy to get your input.

- [GitLab Codebase](#)
- [Feedback via email](#)
- Closer participation after a Deep Talk



hmg-netcdf.readthedocs.io

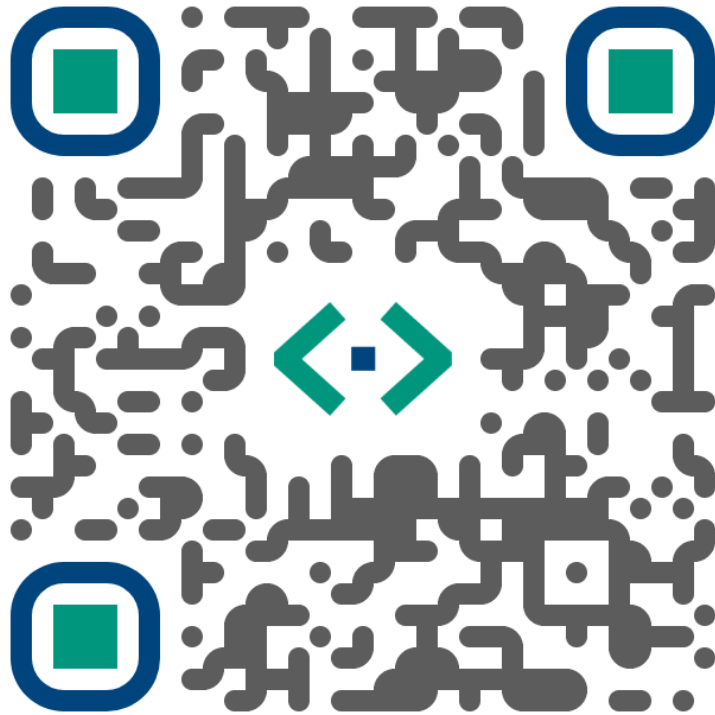


The screenshot shows the 'General Global Attribute Index' website. It features a search bar at the top, a sidebar with navigation links like 'GENERAL FILE SPECIFICATIONS', 'GLOBAL ATTRIBUTES', and 'General Global Attribute Details', and a main table of attributes. The table lists various attributes such as 'Conventions', 'title', 'source', 'license', 'comment', 'creation_date', 'time_coverage_start', 'time_coverage_end', 'time_coverage_duration', 'time_coverage_resolution', 'featureType', 'platform', 'platform_vocabulary', 'instrument', 'instrument_vocabulary', 'history', 'publisher_name', 'publisher_identifier', 'Researcher_name', 'Researcher_id', 'Researcher_affiliation', 'Researcher_affiliation_id', 'creator_name', 'creator_id', 'creator_affiliation', 'creator_affiliation_id', and 'creator_email'. Each row provides a description and an example value.

Attribute	Description	Example	Status
Conventions	Convention version	CF-1.8	mandatory
title	Title of the dataset - A concise description of the content of the data set.	<ul style="list-style-type: none">Ozone data from CMIP6 models and observationsRadiosonde data from 12th June, 2019 at 09:54 in Dittersdorf as part of the MOSES project, CMIP6 Historical Simulation	mandatory
source	The source of the data. This should be a short description of the data source, e.g. "model name", "instrument name", "satellite name", "field campaign name", etc.	CMIP6 model	mandatory
license	URL to a standard or specific license	https://creativecommons.org/licenses/by/4.0/	recommended
comment	A short description of the file contents; essentially an abstract. Part of the chapter "Description of file contents" in the CF conventions.	ozone data assembled from models and field observations	optional
creation_date	ISO 8601 time stamp: Creation date of this netCDF file. Should be a time stamp conform with ISO 8601 (YYYY-MM-DD[THH:MM:SS])	<ul style="list-style-type: none">2020-06-19T10:00:002020-05-01	recommended
time_coverage_start	Time of the first data point in the data set	2023-12-12T13:18:00	recommended
time_coverage_end	Time of the last data point in the data set	2023-12-13T00:07:00	recommended
time_coverage_duration	Duration of the data set	P4DT12H31M12S	recommended
time_coverage_resolution	Targeted time period between each value in the data set	PT1S	recommended
featureType	for discrete sampling geometry see "CF convention"	point, timeSeries, trajectory, profile, timeSeriesProfile, trajectoryProfile	recommended
platform	The name of the platform on which the data was collected. This could be a ship, aircraft, satellite, etc.	<ul style="list-style-type: none">06LOLudwig Prandtl	optional
platform_vocabulary	Controlled vocabulary for the names used in the "platform" attribute.	ICES: https://vocab.ices.dk/	optional
instrument	The name of the instrument used to collect the data. This could be a sensor, probe, etc.	<ul style="list-style-type: none">bathymetric LIDARSDN.L05::312	optional
instrument_vocabulary	Controlled vocabulary for the names used in the "instrument" attribute.	<ul style="list-style-type: none">NERC Vocabulary Server: https://vocab.nerc.ac.uk/search_nv/s/SeaDataNet	optional
history	A list of the modifications to the original data. Part of the chapter "Description of file contents" in the CF conventions	<ul style="list-style-type: none">2020-06-19T10:00:00: created this file2020-06-19T10:00:00: added	optional

Thanks

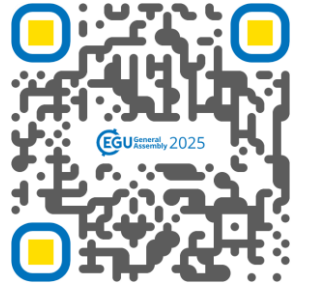
and we value your Feedback!



hmg-netcdf.readthedocs.io



Sharing is
encouraged



Abstract: EGU25-6484