HMG NetCDF 2.0

DataHub
Research Field Earth & Environment

Building a Community-Driven Metadata Guideline for FAIR and AI-Ready NetCDF Data

Björn Saß

Helmholtz Coastal Data Center
Helmholtz-Zentrum Hereon

Romy Fösig

Institute of Meteorology and Climate Research
Karlsruhe Institute for Technology

DSS 2026 @ GEOMAR, Kiel

Karlsruher Institut für Technologie

Helmholtz-Zentrum hereon

1 The Metadata Problem

Scientific Data floods the infrastructures.

- massive data volumes
- increasing data sharing
- growing automation

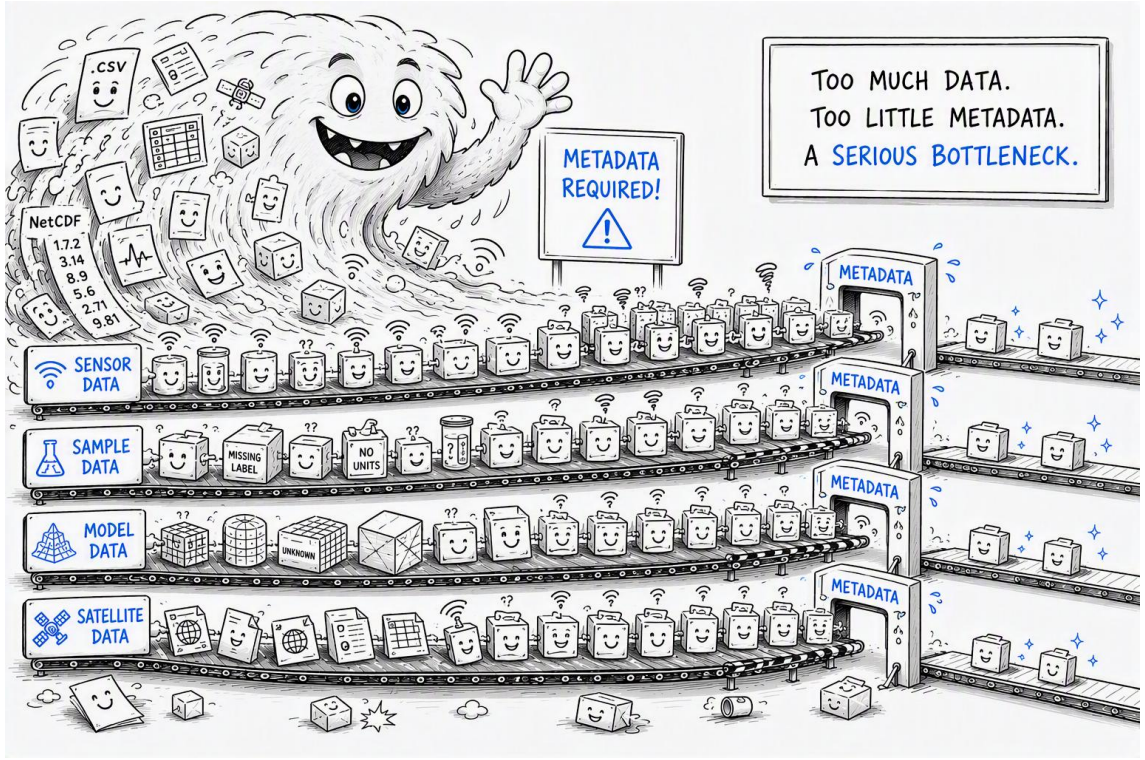


Image: generated with AI

Standards Exist. Practice Varies.

CF, ACDD and DataCite provide strong foundations.

Yet datasets still differ substantially in

- completeness
- consistency
- discoverability
- interoperability

Consequences:

- reduced reuse
- integration effort
- manual interpretation
- limited automation

Helmholtz-Zentrum hereon

Act 1: Introducing HMG NetCDF

What is HMG NetCDF?

H elmholtz
M etadata
G uidelines
for NetCDF

Helmholtz-Zentrum hereon

What is the HMG NetCDF?

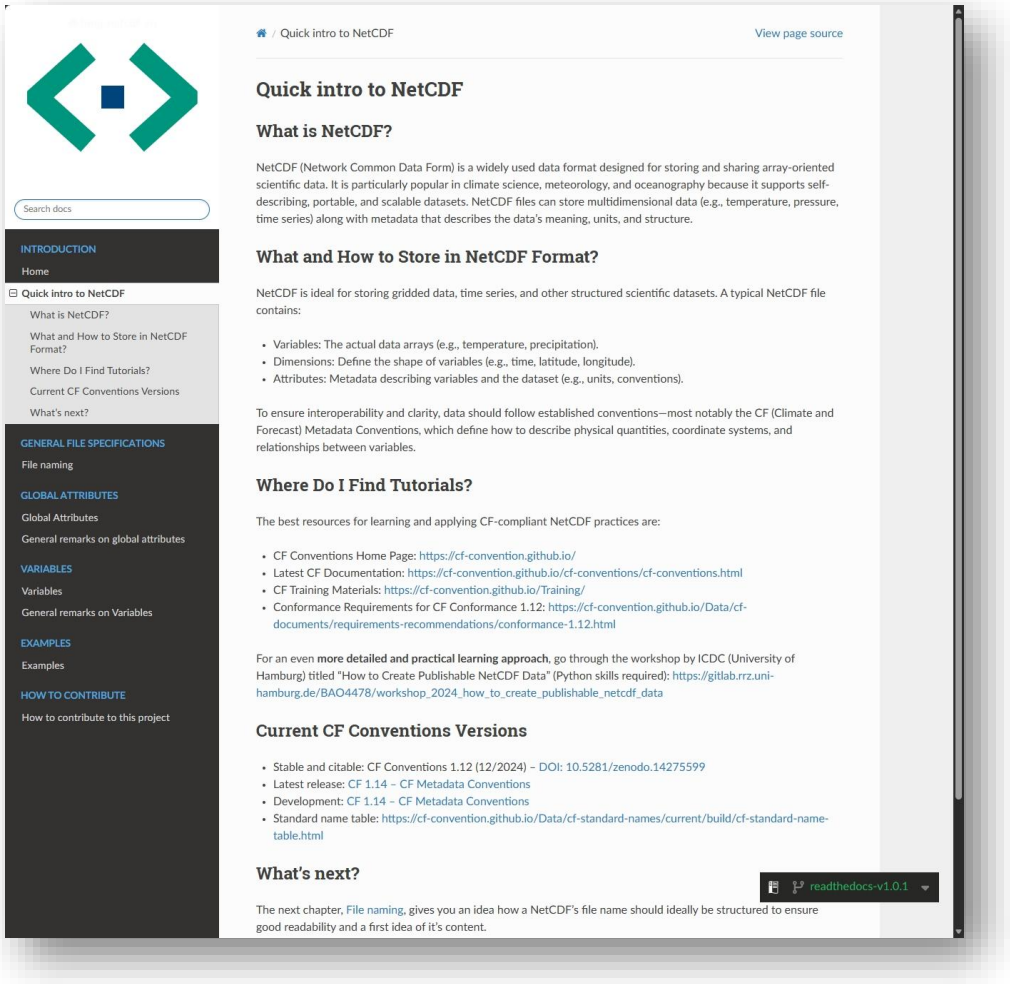
https://hmg-netcdf.readthedocs.io/

HMG NetCDF is a **community-driven metadata guideline** initiative for Helmholtz Earth and Environment sciences which currently concentrates on the NetCDF data format.

It builds upon:

- NetCDF introduction & best practices
- Climate & Forecast (CF) Conventions
- Attribute Convention for Data Discovery (ACDD)
- DataCite Schema
- Organizational Profiles

and translates existing standards into practical recommendations.



DOI 10.5281/zenodo.19691977

21 people involved

Helmholtz Metadata Guidelines for NetCDF

Evolving network. Basis for sustainability.

Repositories

We are in exchange with PANGAEA and the DKRZ.

Target: Reduce curation workload and simplify data submission.



HMG NetCDF

DKRZ
DEUTSCHES KLIMARECHENZENTRUM

CF Community

We are involved in discussions, take part on workshops and are in exchange with the foundational organisations in UK.

UKRI NERC Environmental Data Service

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

DataHub
Research Field Earth & Environment

KIT Karlsruher Institut für Technologie

AWI

JÜLICH Forschungszentrum

GEOMAR

GFZ

marum

Helmholtz-Zentrum hereon

Helmholtz Metadata Guidelines for NetCDF

Why HMG NetCDF 2.0?

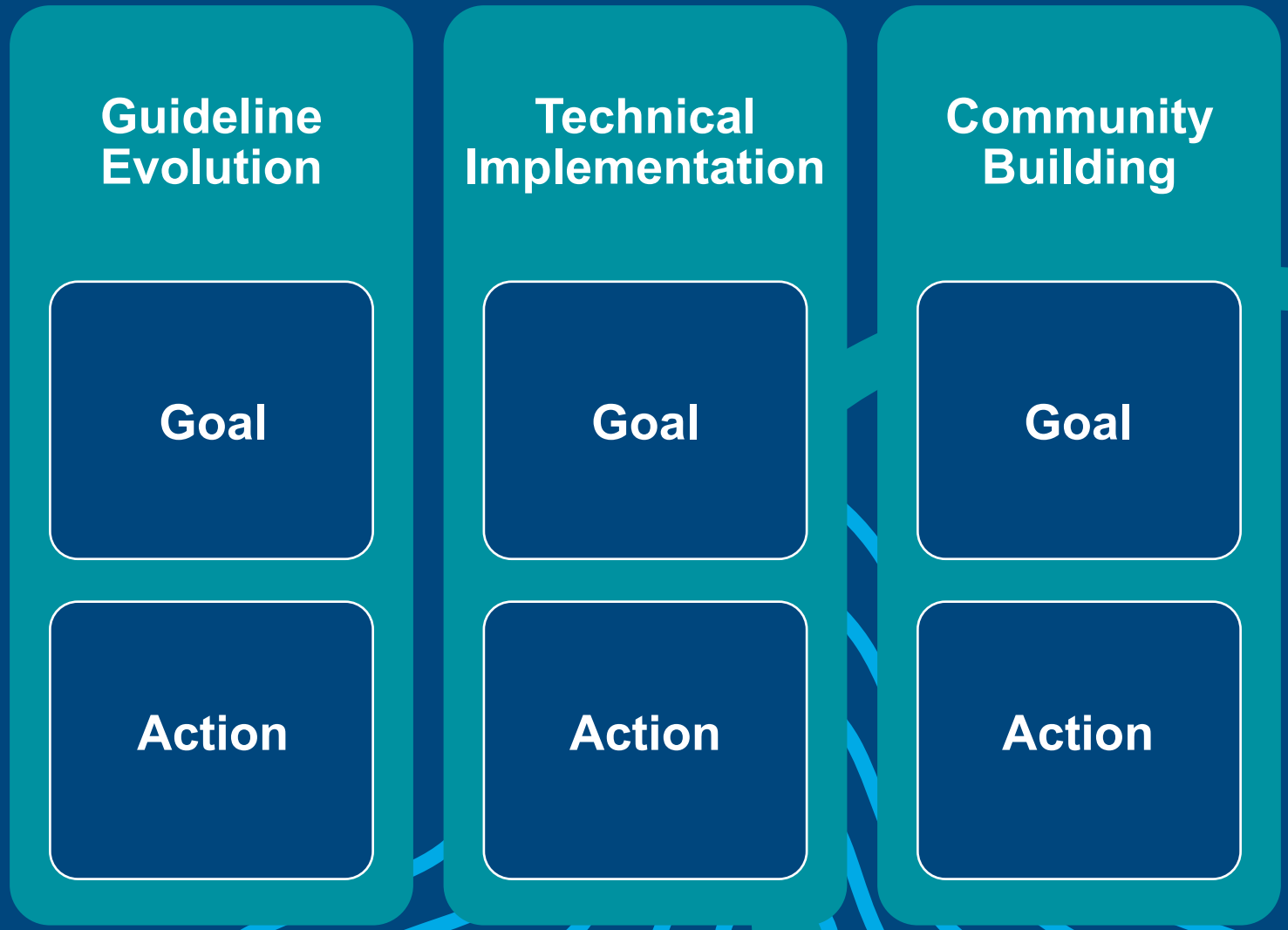
Version 1 delivered metadata guidance.
Mainly global attributes and a network across „all“ centers.

Version 2 aims to turn the guidelines into:

- a living document which serves as a standard for onboarding on NetCDF
- a collection of organizational metadata profiles
- a validation ecosystem
- an active community
- a sustainable governance model

Join us on codebase

Act 2: The Three Pillars for 2.0



Helmholtz-Zentrum hereon

	Guideline Evolution	Technical Implementation	Community Building
Goal	Consistent metadata Recommendations for emerging needs Reusable profiles	Turn recommendations into practice Automated metadata assessment Integration into workflows	Different users, different needs Shared practices Sustainable ownership
Action	Extend guideline content Provenance & quality information AI-readiness concepts Profile development	Validation ecosystem Checker tools Repository integration CI/CD integration	User engagement Governance Reusable metadata profiles Feedback processes Sustainable maintenance

Helmholtz-Zentrum hereon

	Guideline Evolution	Technical Implementation	Community Building
Goal	Consistent metadata Recommendations for emerging needs Reusable profiles	Turn recommendations into practice Automated metadata assessment Integration into workflows	Different users, different needs Shared practices Sustainable ownership
Action	Extend guideline content Provenance & quality information AI-readiness concepts Profile development	Validation ecosystem Checker tools Repository integration CI/CD integration	User engagement Governance Reusable metadata profiles Feedback processes Sustainable maintenance

Helmholtz-Zentrum hereon

	Guideline Evolution	Technical Implementation	Community Building
Goal	Consistent metadata Recommendations for emerging needs Reusable profiles	Turn recommendations into practice Automated metadata assessment Integration into workflows	Different users, different needs Shared practices Sustainable ownership
Action	Extend guideline content Provenance & quality information AI-readiness concepts Profile development	Validation ecosystem Checker tools Repository integration CI/CD integration	User engagement Governance Reusable metadata profiles Feedback processes Sustainable maintenance

Helmholtz-Zentrum hereon

Act 3: Looking Beyond NetCDF

The ideas are larger than the file format.

How does that sound:
Helmholtz Metadata Guidelines.

Helmholtz Metadata Guidelines
Beyond NetCDF

Transferability will be explored, e.g. for:

- Zarr
- HDF5
- RO-Crate
- Helmholtz Knowledge Graph

The overarching objective is metadata harmonisation across infrastructures and disciplines by facing the challenges brought by Machine Learning, AI and Digital Twins.

Join us on codebase

Helmholtz-Zentrum hereon

Summary

HMG NetCDF 2.0

Towards Machine-Actionable Data Ecosystems

- NetCDF provides storage.
- CF provides semantics.
- HMG provides guidance.
- Validation provides trust.
- Community provides sustainability.

Take-Home Message

Standards alone are not enough.

To achieve interoperable and AI-ready data ecosystems we need:

- shared metadata practices
- validation mechanisms
- reusable profiles
- active and transparent communities

HMG NetCDF 2.0 combines all four (for NetCDF alike data formats).

Thanks

and we value your Feedback!

hmg-netcdf.readthedocs.io

And we love to discuss questions during coffee and lunch breaks.

Contact

Björn Saß
bjoern.sass@hereon.de

Romy Fösig
romy.foesig@kit.edu

KIT Karlsruher Institut für Technologie

Helmholtz-Zentrum hereon