`cfdm`, `cf-python` & `cf-plot`: Python data tools for CF-netCDF

ESiWACE Summer School on Effective HPC for Climate and Weather: Storage → Input/Output and Middleware
25th August 2020

Sadie Bartholomew
NCAS & University of Reading
On behalf of the NCAS-CMS team working on CF
Acknowledging the international netCDF and CF community

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Introduction & scope

- NetCDF files + CF Metadata Conventions = CF-netCDF
  - flexible self-describing storage for array-based geoscientific data
  - plus standardised metadata to facilitate comparison & processing
- From the netCDF to the CF-netCDF data model
- A suite of Python tools for working with CF-netCDF
  - `cfdm`, `cf-python`, `cf-plot` & `cf-checker`
  - built around the CF data model, so able to process any CF-compliant dataset e.g. read, write, modify, analyse, regrid & plot

There is a ~1 hour walk-through session next demonstrating use of the data tools. These slides (~30 mins) summarise the underlying concepts.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

NetCDF in geoscience: recap

Network Common Data Form

- Binary file format (.nc) adopted currently as de-facto standard for exchange & storage of earth science data
  - + supporting set of software libraries with APIs in many languages
  - originally (& still actively) developed by UCAR's Unidata project
  - netCDF-4/HDF5 backward compatible with "classic" netCDF-3

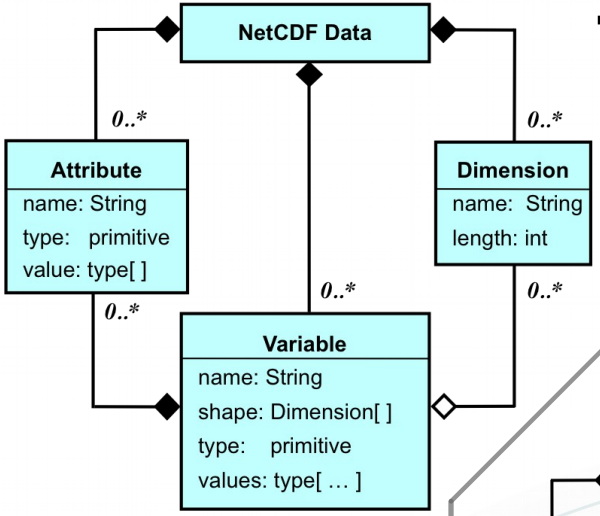
✔ self-describing (metadata categorises each data array)
✔ portable (machine independent)
✔ open source, actively maintained
✔ wide use by a diverse community
✔ very flexible (therefore...)
✖ ...requires interpretation

Recommended resource:
➔ UCAR netCDF homepage, including documentation, release & support details, a tutorial, FAQs & more: www.unidata.ucar.edu/software/netcdf/

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

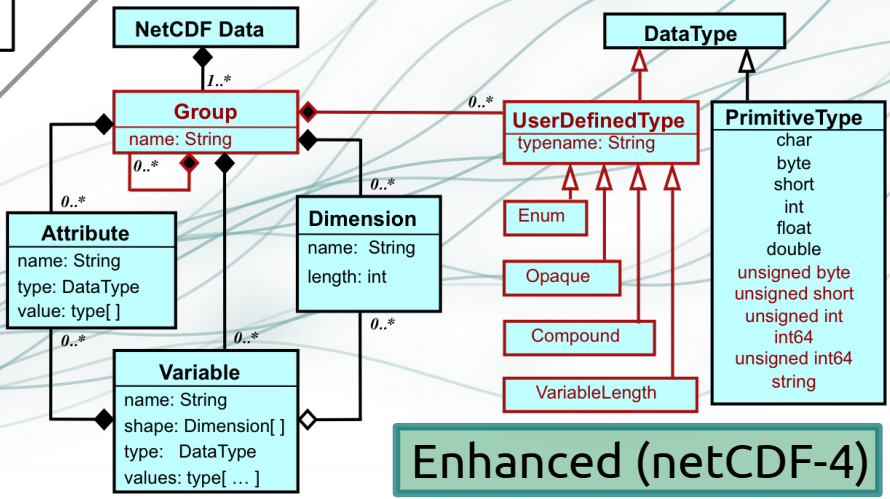
The NetCDF data models: recap

- Adds groups & user-defined types to classic model



Classic (netCDF-3)

- 3 key elements:
  - dimensions
  - variables
  - attributes



Enhanced (netCDF-4)

Diagrams by UCAR Unidata: found at www.unidata.ucar.edu/software/netcdf/papers/nc4_conventions.html

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

CF Metadata Conventions

Climate & Forecast

- Intended for climate & forecast data (model, satellite, observational, etc.) for atmosphere, surface & ocean
- Metadata rules to provide a *definitive* description of:
  - what the data in each variable represents; &
  - the spatial & temporal properties of that data.
- Updated by established community consensus process

✔ reduces interpretation requirement on netCDF

✔ enables users of data from different sources to decide which quantities are comparable

✔ human- & machine- readable

Recommended resource:

➔ CF Conventions website, including the formal convention documents & tables, links to discussions, presentations, & more: cfconventions.org

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

CF-netCDF elements

Table lifted from CF 1.9 draft document, first appearing in paper (see †). First column added & items re-ordered by SB.

Type	CF-netCDF element	Description
Dimension	Dimension	Independent axis of the domain
Variable	*Data* variable	Scientific data discretised within a domain
	Coordinate variable	Unique coordinates for a single axis
	Auxiliary coordinate variable	Additional or alternative coordinates for any axes
	Scalar coordinate variable	Coordinate for an implied size one axis
	Grid mapping variable	Horizontal coordinate system
	Boundary variable	Cell vertices
	Cell measure variable	Cell areas or volumes
	Ancillary data variable	Metadata that depends on the domain
Attribute	*Formula terms* attribute	Vertical coordinate system
	Feature type attribute	Characteristics of discrete sampling geometry
	Cell methods attribute	Description of variation within cells

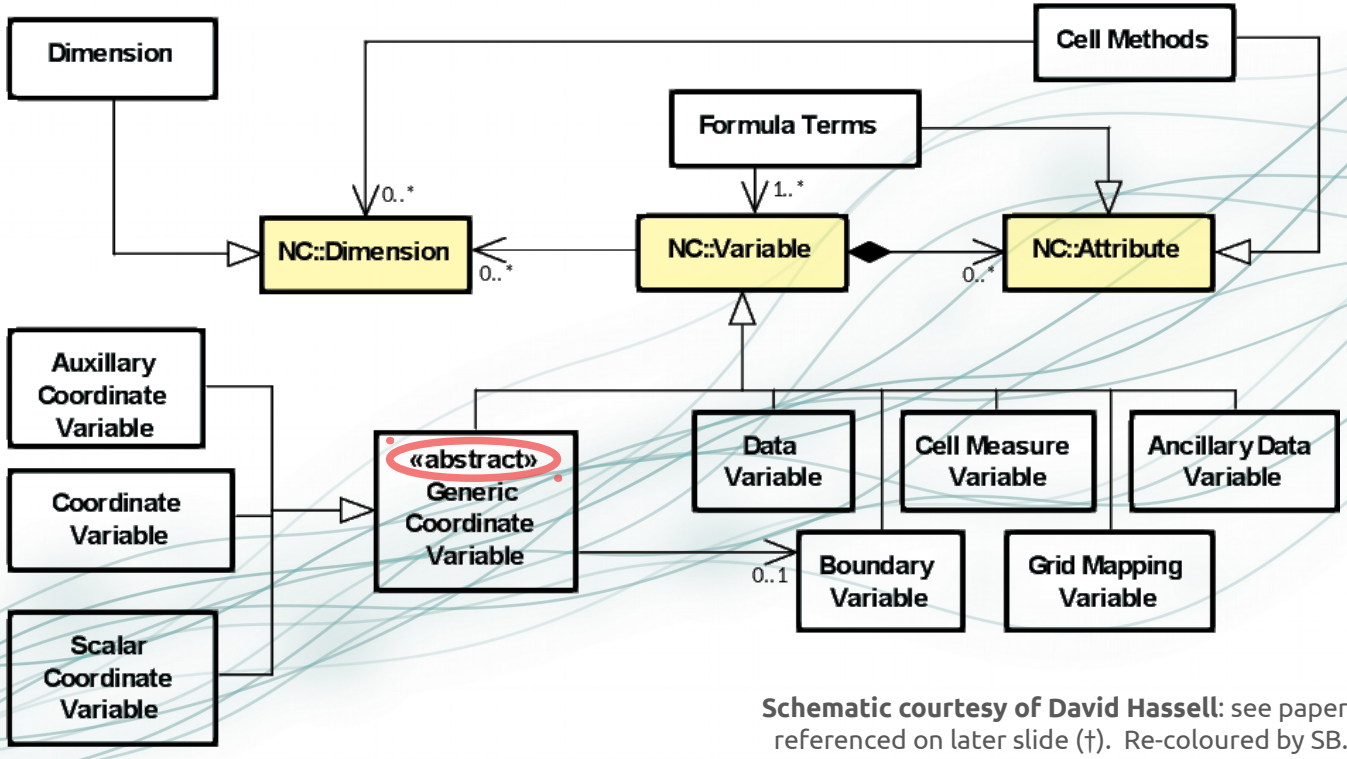
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Correspondence to netCDF



Schematic courtesy of David Hassell: see paper referenced on later slide (†). Re-coloured by SB.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

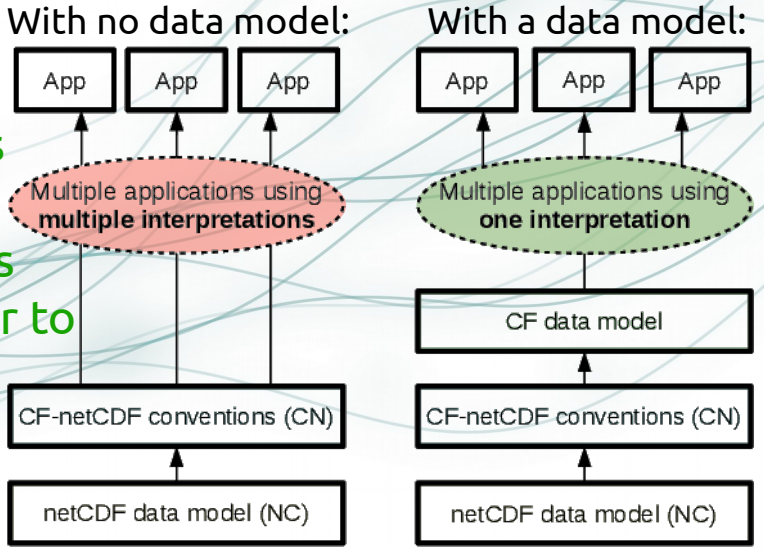
University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

A data model for CF-netCDF

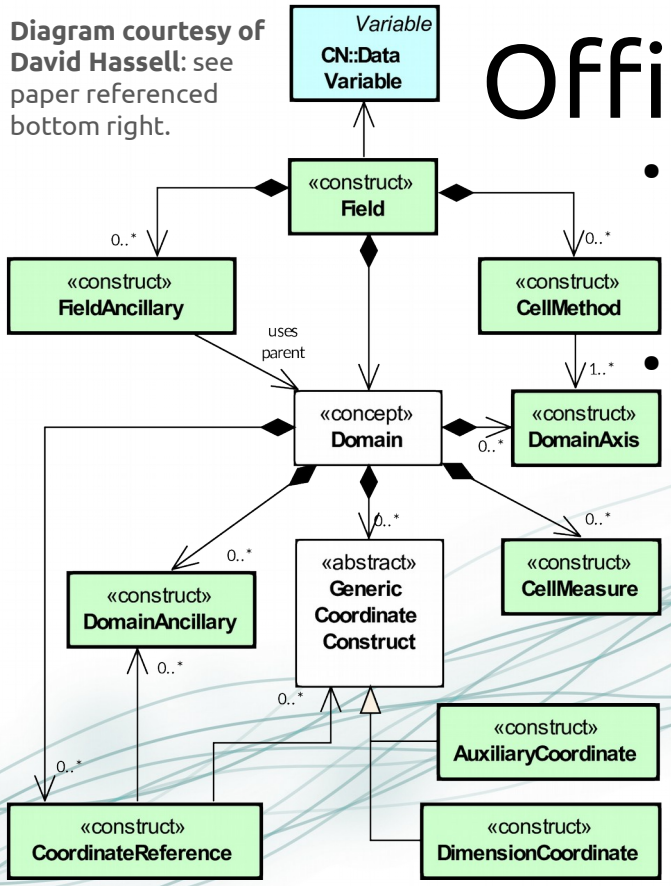
- Benefits of a formal, consistent model for CF-netCDF:
  - ✔ improve understanding of CF-netCDF by identifying distinct elements & inherent relationships
  - ✔ facilitate enhancements to the CF Conventions
  - ✔ improved software tools
  - ✔ CF-compliant data easier to represent in other file formats
  - ✔ *one* interpretation for every application:



Schematic courtesy of David Hassell: see paper referenced on later slide (†). Re-coloured by SB.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Diagram courtesy of David Hassell: see paper referenced bottom right.



Official data model

- Guaranteed to be up-to-date with the Conventions for every release (CF 1.6+)
- Is "necessary & sufficient"
  - minimal set of elements sufficient to account for all of CF...
  - ...with no additional elements

For full detail on the model, see:

➔ CF Conventions 1.9 draft: cfconventions.org/cf-conventions/cf-conventions.html#appendix-CF-data-model

➔ dedicated paper (†) by Hassell et al.: doi.org/10.5194/gmd-10-4619-2017

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

The full picture

- Encoding (netCDF) independent

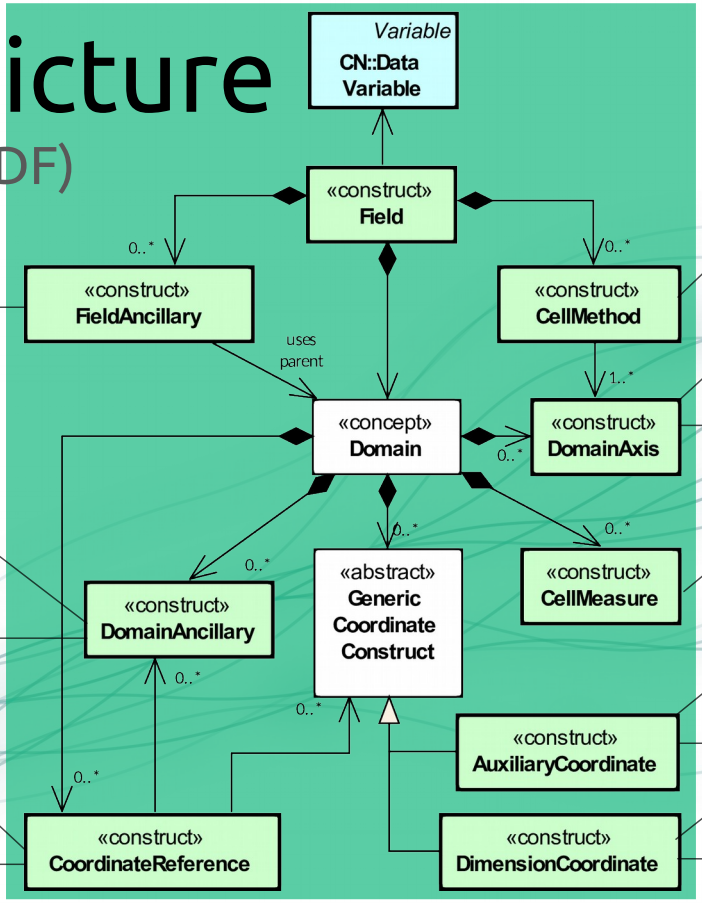


Diagram courtesy of David Hassell: see paper referenced on previous slide (†).

Using CF-compliant netCDF

- Many excellent open-source tools exist for netCDF as listed in the links below, but not all recognising CF...
- … including multiple Python libraries
- The official Python/NumPy interface to the netCDF C library is Unidata's `netcdf4-python` library:
  - NCAS CF suite of tools discussed next use this as a dependency

- `netcdf4-python`: see unidata.github.io/netcdf4-python/netCDF4/index.html
- To find and/or read about tools that can be used with (CF-)netCDF datasets:
  - ➔ Unidata's near-exhaustive list, 'Software for Manipulating or Displaying NetCDF Data': www.unidata.ucar.edu/software/netcdf/software.html
  - ➔ CF Conventions listing of 'Software that "Understands" CF Data': cfconventions.org/software.html

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

NCAS CF-netCDF Data Tools

- A small suite of compatible, complimentary tools
- All open-source (hosted on GitHub) & Python 3 based

Library	Description & purpose	Functionality
`cfdm`	Reference implementation of the CF data model	For the most part, only that required to read and write datasets, and to create, modify and inspect field constructs in memory
`cf-python`	CF-compliant geoscientific data analysis library	Much higher-level than `cfdm`, e.g. statistical operations, collapses, subspacing, regridding
`cf-plot`	Set of Python functions for making the visualisations often used by geoscientists	That for plotting e.g. contour, vector and line plots from field constructs (or `numpy` arrays)
`cf-checker`	CF compliance checking utility	Checks the CF compliance of a netCDF file

There is not sufficient time to cover the `cf-checker`, so for more info, see:

➔ the code repository e.g. to install: github.com/cedadev/cf-checker

➔ the browser-based interface: pumatest.nerc.ac.uk/cgi-bin/cf-checker.pl

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

The field construct

- Central object in `cfdm` & `cf-python` is the *field construct* → a CF-netCDF data variable with all of its metadata
- A field construct `cfdm.Field` or `cf.Field` consists of:
  - descriptive properties that apply to field construct as a whole (e.g. the standard name);
  - a data array; &
  - metadata constructs that describe the locations of each cell of the data array (the *domain*) → the eight other constructs of the data model i.e. classes in the UML diagrams on previous slides

For more information, please see:
➔ field construct breakdown within the `cfdm` documentation:
ncas-cms.github.io/cfdm/cf_data_model.html

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

`cfdm` Python library

- A reference implementation of the CF data model, hence a complete representation of CF!
- *Designed to be subclassed*, so that the creation of a new implementation of the CF data model, based on `cfdm`, is straight forward
- Includes a stand-alone core implementation, the `cfdm.core` package, that includes no functionality beyond that mandated by the CF data model

For more information, please see:

➔ The documentation, including installation information & an API reference: ncas-cms.github.io/cfdm/

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

cf-python library

- Builds upon `cfdm` to provide diverse geoscientific data analysis capability → `cfdm` with high-level functionality
- As a small sample, `cf-python` can:
  - read, inspect, & write field constructs from netCDF & CDL (& more);
  - modify & analyse field construct metadata & data;
  - perform statistical collapses on field constructs;
  - create subspaces of field constructs;
  - regrid field constructs (several interpolation methods supported);
  - combine field constructs arithmetically; &
  - read & process netCDF & CDL containing hierarchical groups.

For more information, please see:
➔ The documentation, including installation information & an API reference: ncas-cms.github.io/cf-python/

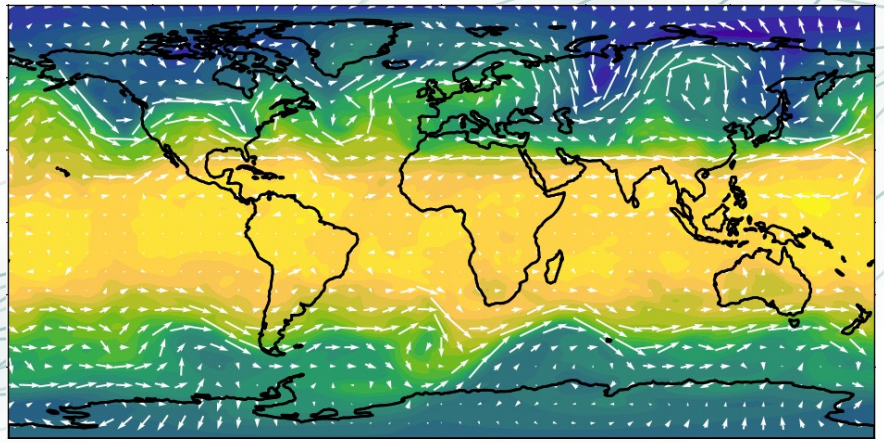
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

cf-plot Python library

- CF-aware geoscientific visualisation
- Generally uses cf-python to present the data & CF attributes for plotting (can also use numpy arrays)
  - contour plots
  - vectors plots
  - plots of trajectories
  - significance plots
  - & more...

For more information, see:

➔ The documentation, including installation information, a gallery of plots & a user guide: ajheaps.github.io/cf-plot/



Example contour plot with overlaid vectors created with cf-plot.
Colourbar & axes labels omitted as it is just for illustration.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

Summary

- NetCDF files compliant with the CF Metadata Conventions (CF-netCDF) enable *flexible self-describing storage* of array-oriented geoscientific data
- CF-netCDF has become *a community standard*
- Different data models of CF-netCDF are possible, but an *official* model exists & is up-to-date for all CF 1.6+
  - formal model is “necessary & sufficient” & netCDF-independent
- Numerous tools for working with netCDF exist, including in Python, but NCAS’s CF suite is built upon the official CF data model: CF compliance *at heart*
  - able to process any CF-compliant (or non-compliant) netCDF
  - read, write, inspect, modify, analyse, plot, check compliance & more

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of Reading

NCAS work on CF is supported by:

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

We now move onto a walk-through of the NCAS CF libraries in practice. But I welcome any questions about the concepts at this stage!

Thanks for listening (so far). Any questions?

- Quick links to useful related resources:
  - UCAR netCDF homepage: www.unidata.ucar.edu/software/netcdf/
  - CF-netCDF (Metadata) Conventions homepage: cfconventions.org
  - `cf-python` documentation: ncas-cms.github.io/cf-python/
  - `cfdm` documentation: ncas-cms.github.io/cfdm/
  - `cf-plot` documentation: ajheaps.github.io/cf-plot/
  - walk-through & lab materials: github.com/NCAS-CMS/cf-training

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL
University of Reading
NCAS work on CF is supported by:
is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING