National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

The CF data model and CF aggregation rules

David Hassell and Jonathan Gregory

National Centre for Atmospheric Science

University of Reading

Contributions from Jon Blower, Mark Hedley, Bryan Lawrence + many others in the CF community

GO-ESSP, 25 February 2015

Proposed CF data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- A minimal, logical representation of the content contained in the CF-netCDF conventions
  - *minimal* → as simple as possible
  - *logical* → independent of any implementation or encoding
  - *content contained* → what CF is, not what it could/should be
- It differs from the Unidata common data model (no CF information) and the OGC CF-netCDF data model (not encoding independent)
- First proposed in August 2011. Not yet adopted by CF, though much debated ...

```
http://cf-trac.llnl.gov/trac/ticket/68
http://cf-trac.llnl.gov/trac/ticket/95
http://cf-trac.llnl.gov/trac/ticket/107
```

Motivation for a data model

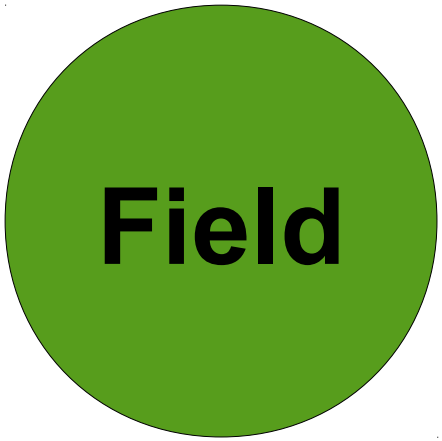
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- `http://cf-trac.llnl.gov/trac/ticket/88`
  - Provide an orientation guide to the CF conventions document
  - Guide the development of software compatible with CF
  - Facilitate the creation of an API which 'behaves/feels like CF' and is intuitive to use.
  - Provide a reference point for gap analysis and conflict analysis of the CF specification
  - Provide a communication tool for discussing CF concepts and proposals for changes to the CF specification
  - Set the ground work to expand CF beyond netCDF files

Constructs of the data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

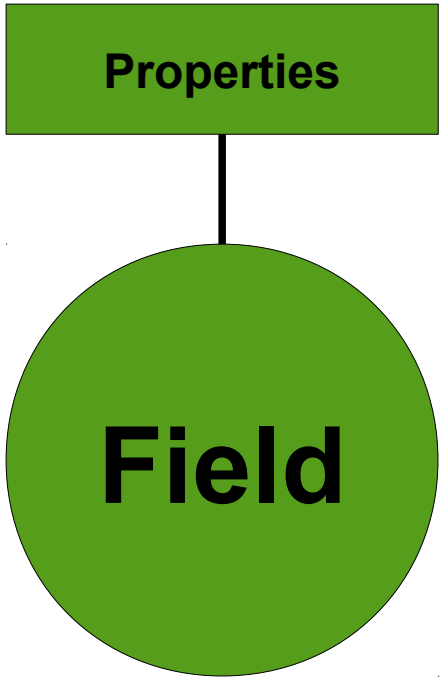
Field construct = array described by metadata



Constructs of the data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

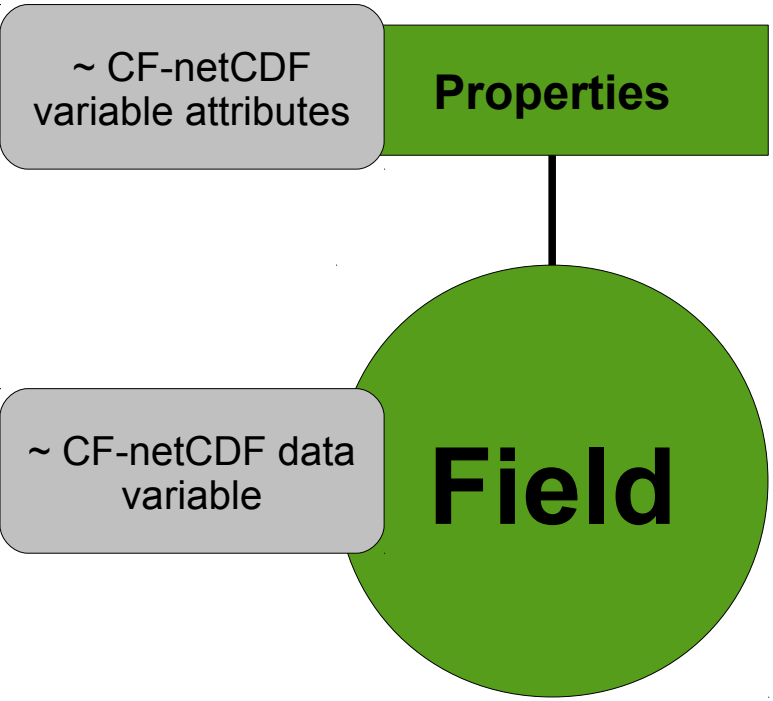
Field construct = array described by metadata



Constructs of the data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

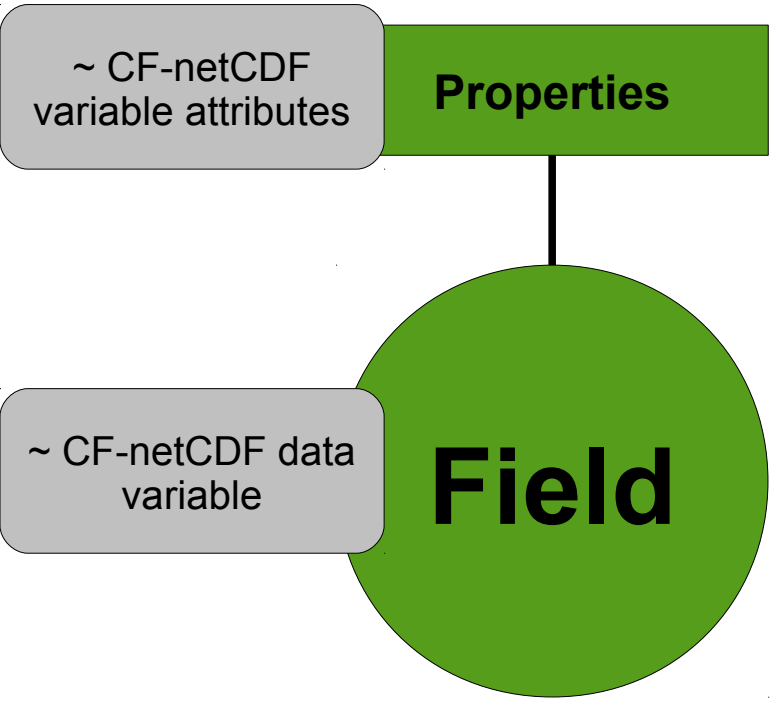
Field construct = array described by metadata

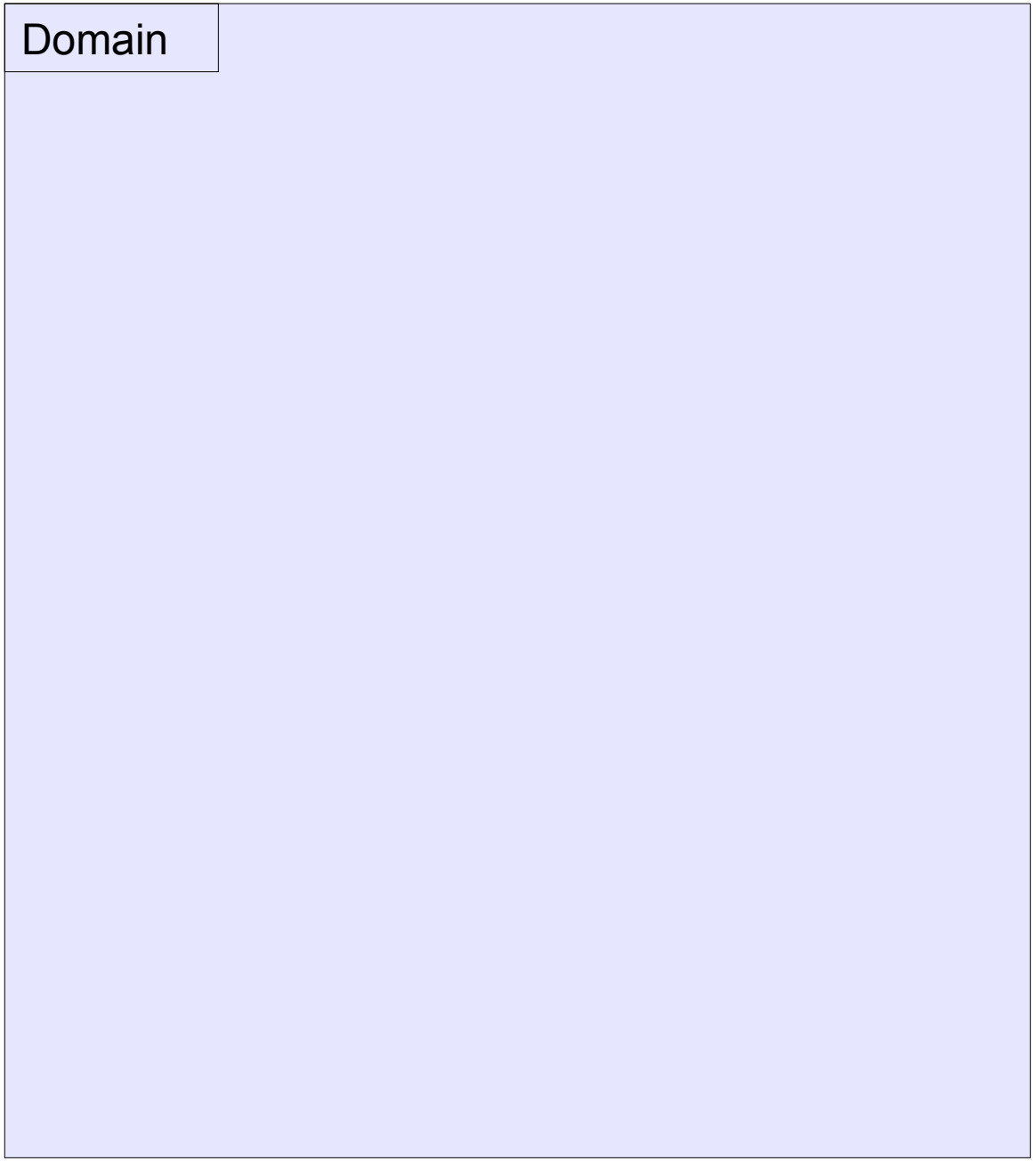


Constructs of the data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

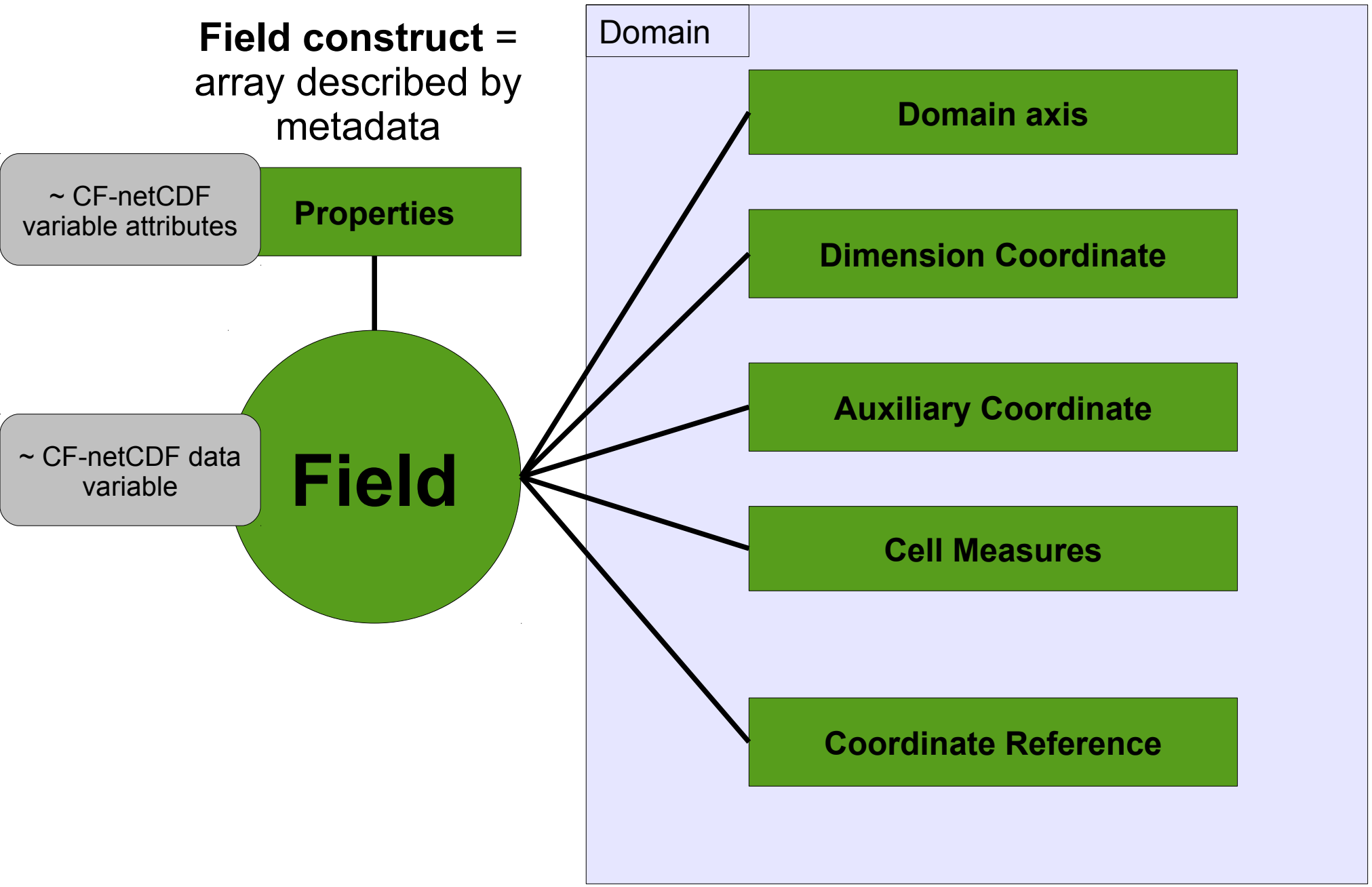
Field construct = array described by metadata





Constructs of the data model

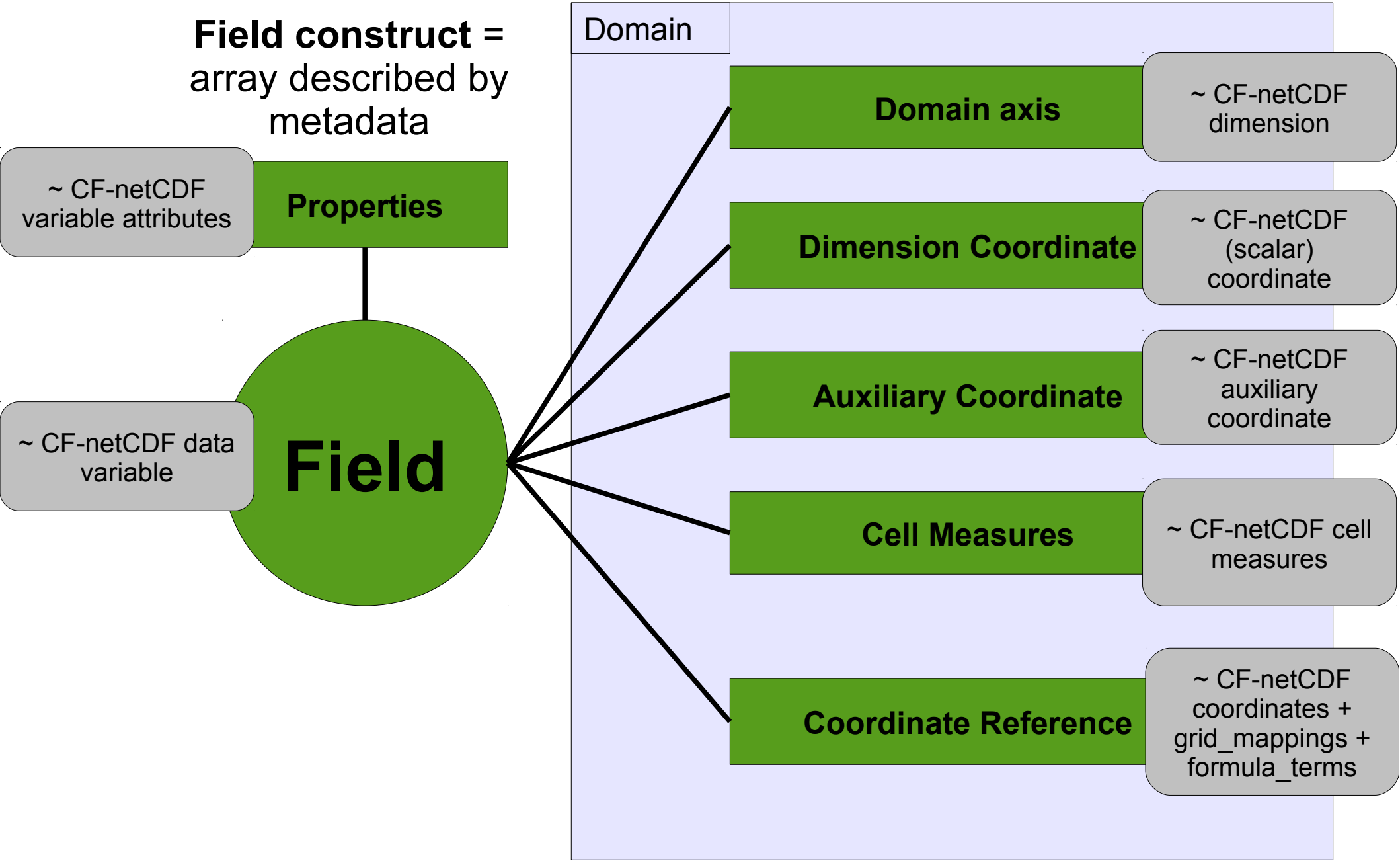
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



Constructs of the data model

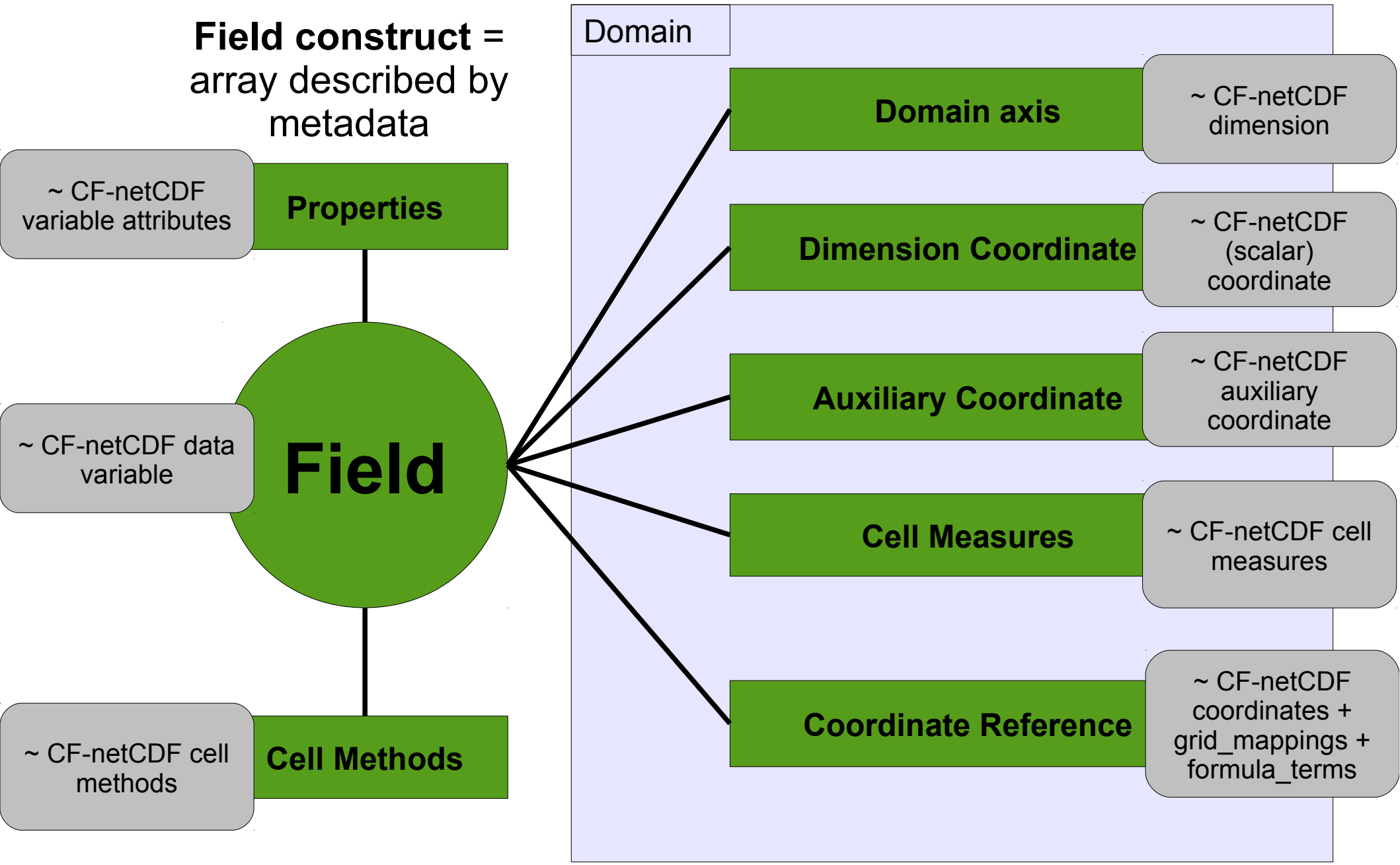
National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

Field construct = array described by metadata



Constructs of the data model

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL



Coordinate reference construct

- A coordinate reference construct relates the field's coordinate values to locations in a planetary reference frame.

- Conflates **`grid_mapping`** (horizontal) and **`formula_terms`** (vertical) cases by noting that each is just a means of defining, more accurately, where grid cells are located.

- Together with some coordinate constructs, it may define a coordinate reference system:
  - Coordinate values and units
  - A datum
  - A formula for converting coordinates to a different coordinate system.

Grid mapping

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

```
int crs ;
    crs:grid_mapping_name = "transverse_mercator";
    crs:semi_major_axis = 6377563.396 ;
    crs:semi_minor_axis = 6356256.910 ;
    crs:inverse_flattening = 299.3249646 ;
    crs:latitude_of_projection_origin = 49.0 ;
    crs:longitude_of_projection_origin = -2.0 ;
    crs:false_easting = 400000.0 ;
    crs:false_northing = -100000.0 ;
    crs:scale_factor_at_central_meridian = 0.9996012717 ;
```

- Coordinate reference construct relates:
  - Coordinate constructs (**`projection_x_coordinate`**, **`projection_y_coordinate`**, **`latitude`**, **`longitude`**)
  - Datum (**`semi_major_axis`**, *etc.)*
  - Formula name (**`transverse_mercator`**)
  - Formula parameters (**`scale_factor_at_central_meridian`**, *etc.)*

formula_terms

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

```
float lev(lev) ;
  lev:positive = up" ;
  lev:standard_name = "atmosphere_hybrid_height_coordinate" ;
  lev:formula_terms = "a: A b: B orog: OROG” ;
```

- Coordinate reference construct relates:
  - Coordinate constructs (**`atmosphere_hybrid_height_coordinate`**, **`height`**)
  - Formula name (**`atmosphere_hybrid_height`**)
  - Formula parameters:
    - **`a`** (auxiliary coordinate construct)
    - **`b`** (auxiliary coordinate construct)
    - **`orog`** (field construct)

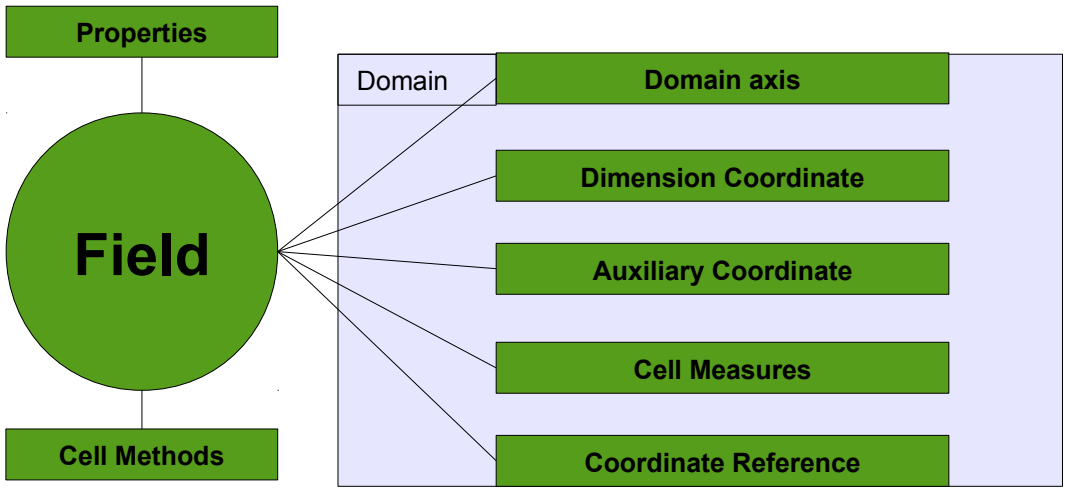
Proposed CF aggregation rules

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Application of the CF data model

`http://cf-trac.llnl.gov/trac/ticket/78`

- Aggregation: combining multiple datasets into a single dataset with a larger domain
  - E.g. a timeseries split across several files may be logically viewed as a single dataset
- Having the data model makes it feasible to define all of the conditions under which any CF fields are aggregatable



CF aggregation rules (#78)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

1. Both fields have identical standard name properties.

2. Both fields have the same number of coordinate constructs, all of which have a standard name property, and each coordinate construct's standard name is unique within its field. Each coordinate construct in one field forms a pair of matching coordinate constructs with a unique coordinate construct in the other field.

3. Each domain axis in both fields has at least one associated 1D coordinate construct.

4. Each domain axis in one field forms a pair of matching domain axes with a unique domain axis in the other field.

5. There is exactly one pair of matching domain axes for which one or more of the 1D matching coordinate constructs have different values for their coordinate arrays and, if present, their boundary coordinate arrays.

6. Both fields have the same number of cell measure constructs, all of which have a units property. Each cell measure construct in either field forms a pair of matching cell measure constructs with a unique cell measure construct in the other field.

7. Each pair of matching coordinate constructs and matching cell measure constructs that do not span their aggregating domain axes have identical values for their coordinate arrays and, if present, their boundary coordinate arrays.

8. If the pair of matching aggregating domain axes has a pair of associated dimension coordinate constructs, then there are no common values in their coordinate arrays. If the matching dimension coordinate constructs have boundary coordinate arrays then no cells from one dimension coordinate construct lie entirely within any cell of the other dimension coordinate construct. b

9. If one field has a cell methods construct then so does the other field, with the equivalent methods in the same order. Corresponding domain axes in each cell methods are matching pairs.

10. Both fields have the same number of coordinate reference constructs. For each coordinate reference construct in one field there is a coordinate reference construct in the other field with identical coordinate reference name and the same set of terms, taking optional terms into account. Corresponding terms of scalar parameters (as found in CF-netCDF grid_mapping variables) are identical. Corresponding terms of pointers to coordinate constructs or other fields (as found in CF-netCDF formula_terms attributes) are related in the same manner as pairs of matching coordinate constructs.

CF aggregation rules (#78)

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

1. Both **fields** have identical standard name **properties**.

2. Both **fields** have the same number of **coordinate constructs**, all of which have a standard name **property**, and each **coordinate construct's** standard name is unique within its field. Each **coordinate construct** in one field forms a pair of matching **coordinate constructs** with a unique **coordinate construct** in the other **field**.

3. Each **domain axis** in both **fields** has at least one associated 1D **coordinate construct**.

4. Each **domain axis** in one **field** forms a pair of matching **domain axes** with a unique **domain axis** in the other **field**.

5. There is exactly one pair of matching **domain axes** for which one or more of the 1D matching **coordinate constructs** have different values for their **coordinate arrays** and, if present, their **boundary coordinate arrays**.

6. Both **fields** have the same number of **cell measure constructs**, all of which have a units **property**. Each **cell measure construct** in either **field** forms a pair of matching **cell measure constructs** with a unique **cell measure construct** in the other **field**.

7. Each pair of matching **coordinate constructs** and matching **cell measure constructs** that do not span their aggregating **domain axes** have identical values for their **coordinate arrays** and, if present, their **boundary coordinate arrays**.

8. If the pair of matching aggregating **domain axes** has a pair of associated **dimension coordinate constructs**, then there are no common values in their **coordinate arrays**. If the matching **dimension coordinate constructs** have **boundary coordinate arrays** then no cells from one **dimension coordinate construct** lie entirely within any cell of the other **dimension coordinate construct**.

9. If one **field** has a **cell methods construct** then so does the other **field**, with the equivalent methods in the same order. Corresponding **domain axes** in each **cell methods** are matching pairs.

10. Both **fields** have the same number of **coordinate reference constructs**. For each **coordinate reference construct** in one **field** there is a **coordinate reference construct** in the other **field** with identical **coordinate reference** name and the same set of terms, taking optional terms into account. Corresponding terms of scalar parameters (as found in CF-netCDF grid_mapping variables) are identical. Corresponding terms of pointers to **coordinate constructs** or other **fields** (as found in CF-netCDF formula_terms attributes) are related in the same manner as pairs of matching **coordinate constructs**.

cf-python

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- Application of the CF data model and aggregation rules

http://cfpython.bitbucket.org

```
$ ls tasmax_day*.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20051201-20101130.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20101201-20151130.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20151201-20201130.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20201201-20251130.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20251201-20301130.nc
tasmax_day_IPSL-CM5A-MR_rcp85_r1i1p1_20301201-20351130.nc

$ ipython
In [1]: import cf

In [2]: field = cf.read('tasmax_day*.nc')

In [3]: print field
air_temperature field summary
-----------------------------
Data         : air_temperature(time(10800), latitude(145), longitude(192)) K
Cell methods : time: maximum
Axes         : time(10800) = [2005-12-01 12:00:00, ..., 2035-11-30 12:00:00] 360_day
             : latitude(145) = [-90.0, ..., 90.0] degrees_north
             : longitude(192) = [0.0, ..., 358.125] degrees_east
             : height(1) = [1.5] m

In [4]: cf.write(field, 'tasmax_day_20051201-20351130.nc')
```

The future

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

1. To agree on a CF data model

2. The model, and any downstream applications, must evolve with the conventions

Summary

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

- A minimal, logical data model for CF-netCDF has been proposed, but has not yet been accepted

- Working applications of the proposed data model exist: the proposed CF aggregation rules and the cf-python data analysis library