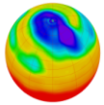
CF Conventions

Ocean Data Dojo
December 2022

Alison Pamment


Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL


National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL


National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

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



UKRI
Science and
Technology
Facilities Council
Natural
Environment
Research Council

CF overview

The CF (Climate - Forecast) metadata conventions

http://cfconventions.org/

- Originated in 2000 for sharing of climate model data
- Define wide-ranging rules for describing geoscientific data in netCDF files
- Provide a description of the data in each variable, including spatiotemporal and other dimensional properties
- Enable users to identify comparable data held in different files
- Facilitate the development of software to extract, process, regrid, analyse and display data

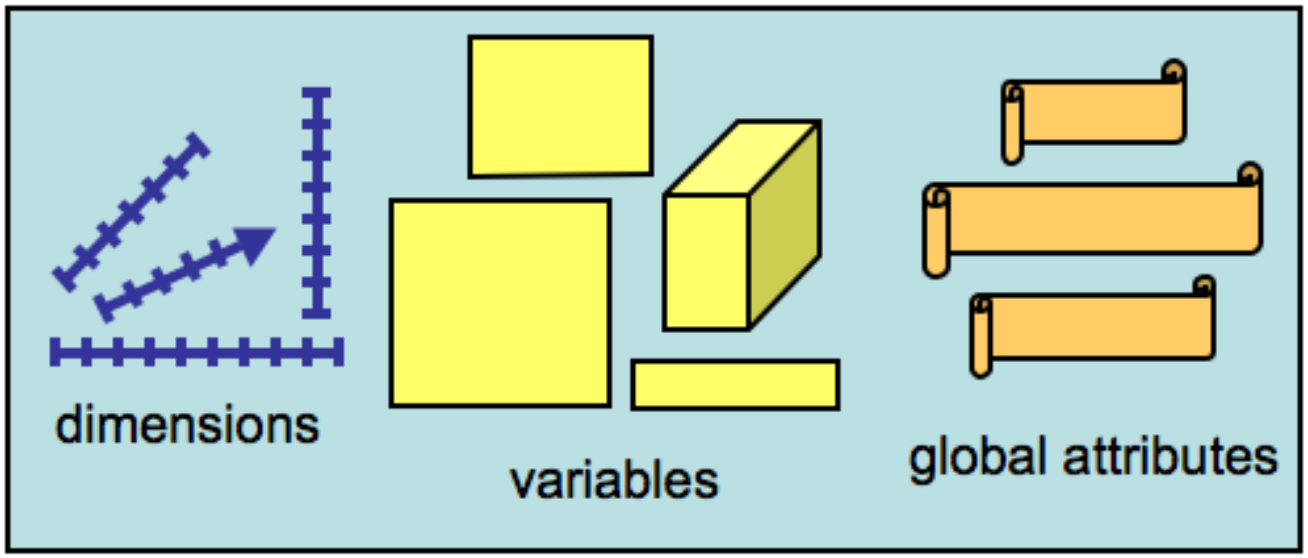
The CF (Climate and Forecast) metadata conventions

http://cfconventions.org/

- Adopted by many international modelling projects, such as CMIP (Coupled Model Intercomparison Project)
  https://www.wcrp-climate.org/wgcm-cmip
- Increasingly being used for observational data, e.g. SeaDataNet oceanographic data and GOES series satellite products
- WMO (World Meteorological Organisation) is developing experimental CF profiles for radar data, marine and atmospheric trajectories:
  https://community.wmo.int/activity-areas/wis/wmo-cf-extensions

What's in a NetCDF file?

NetCDF files are containers for Dimensions, Variables, and Global Attributes.



A netCDF file has a **.nc** file extension and contains some **dimensions**, **variables**, **global** (file-level) **attributes**, and **data values** associated with the variables.

Examples of CF global attributes

Attribute	Meaning
conventions	CF version, e.g. Conventions = "CF-1.10"
title	What's in the file
institution	Where it was produced
source	E.g. Name of model, instrument
history	Audit trail of processing
references	Publications, web pages
comment	Miscellaneous information

Examples of variable attributes

Attribute	Meaning
standard_name	Taken from **standard name table**
units	Mandatory unless dimensionless quantity
long_name	Not standardised
cell_methods	Variation within a cell e.g. max, mean
cell_measures	Area or volume of a cell
valid_max, valid_min, valid_range	“Acceptable” data values
_FillValue	Value to use for missing data
flag_values, flag_meanings	to make "flag" variables self-describing

Coordinate Variables in CF

Coordinate variables have the same name as their dimension:

```
dimensions: longitude = 36;


longitude(longitude)
    :units = "degrees_east";
    :axis = "X";
    :standard_name = "longitude";
```

CF standard names

What is a CF standard name?

A standard name identifies the geophysical quantity in a data variable, e.g. air_temperature.

Attach the standard name to a data variable using the CF standard_name attribute:

```
float psl(lat,lon) ;
   psl:units = “hPa” ;
   psl:standard_name = “pressure_at_mean_sea_level” ;
```

Standard names facilitate data exchange by providing unambiguous identification of variables.

Why use a standard name?

The names of data variables are **not** standardized in netCDF files.
For example:

- alison1, alison2, xyz345blah
- Temperature, temp, T

Standard names facilitate data exchange by providing unambiguous identification of variables.

We can tell whether variables from different data sources can be compared.

CF standard name table

http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

Standard Name	Canonical Units
air_temperature_anomaly "Anomaly" means difference from climatology. Air temperature is the bulk temperature of the air, not the surface (skin) temperature.	**K**
surface_upward_latent_heat_flux The surface latent heat flux is the exchange of heat between the surface and the air on account of evaporation (including sublimation). In accordance with common usage in geophysical disciplines, "flux" implies per unit area, called "flux density" in physics.	**W m-2**

CF standard names: basic rules

- Any variable labelled with the `standard_name` attribute **must** use a value from the published standard name table
- Standard names consist of letters, digits and underscores, no whitespace.
- English language with US spellings
- Case sensitive
  - Mixed case used for chemical element symbols, e.g.

  integral_wrt_time_of_radioactivity_concentration_of_112Ag_in_air
- (Almost) all standard names have an accompanying description
- Names are never removed once they have been added to the table
  - Name can be modified using an 'alias'

The units attribute used in a data file must be physically consistent with the canonical unit…

… but it doesn't have to be identical!

This is because software implementations of the CF conventions can utilise the UniData UDUnits2 package to convert between equivalent units.

What isn't described in the standard name

- Vertical level and geolocation, e.g. 2m air temperature
  - Use coordinate variables or region labels
- Statistical processing, e.g. mean, maximum, etc.
  - Use cell_methods attribute
- Portions of grid cell, e.g. mean surface albedo over snow area
  - Use cell_methods attribute plus area_type coordinate variable
- Units
  - Use units attribute

Standard names for marine environment

Property	Number of names
Ocean / sea_water physics/chemistry/biology	579
Sea surface waves	80
Sea surface height / tidal	31
Mean sea level change	19
Sea ice	90
Total	**799**

Ocean physics / chemistry / biology names

Property	Number of names
Ocean dynamics and transport	104
Ocean thermodynamics / salinity	111
Radiation / surface fluxes	93
Ocean model parameters	40
Ocean chemistry / biology	231
Total	**579**

Examples – physics and chemistry

tendency_of_sea_water_temperature (K s-1)

northward_ocean_salt_transport_due_to_diffusion (kg s-1)

mole_concentration_of_ammonium_in_sea_water (mol m-3)

mole_concentration_of_calcite_expressed_as_carbon_in_sea_water (mol m-3)

mole_concentration_of_cfc11_in_sea_water (mol m-3)

Biological taxa

Use a generic standard name for the data variable, e.g. mass_concentration_of_biological_taxon_expressed_as_carbon_in_sea_water (kg m-3)

To specify the taxon, attach a *coordinate* variable to the data variable:
biological_taxon_name, e.g. "Calanus finmarchicus"

Optionally, provide a Life Science Identifier
biological_taxon_lsid, e.g. "urn:lsid:marinespecies.org:taxname:104464"

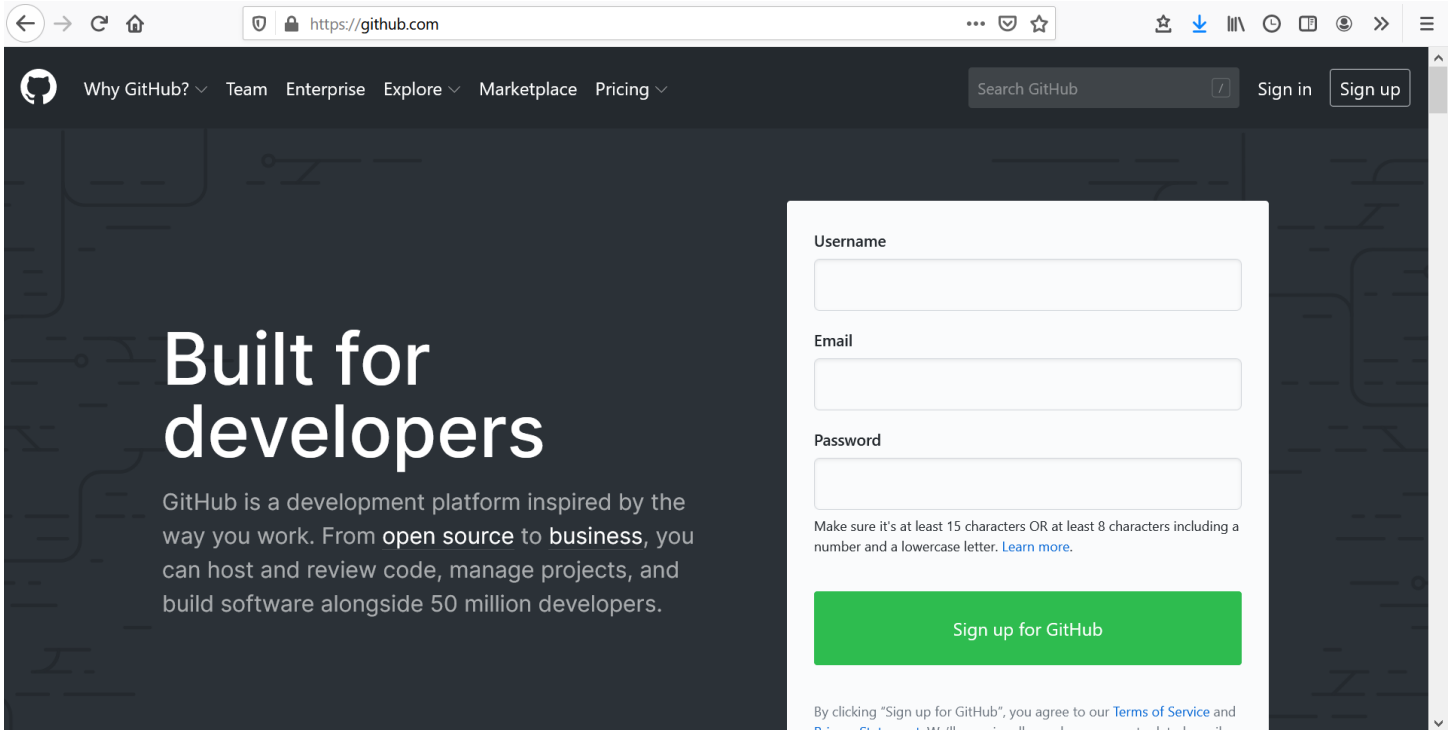
CF recommends World Register of Marine Species (WoRMS) for oceanographic data and Integrated Taxonomic Information System (ITIS) for freshwater and terrestrial data.

Standard names process

How do I ask for a new standard name?

STEP 1

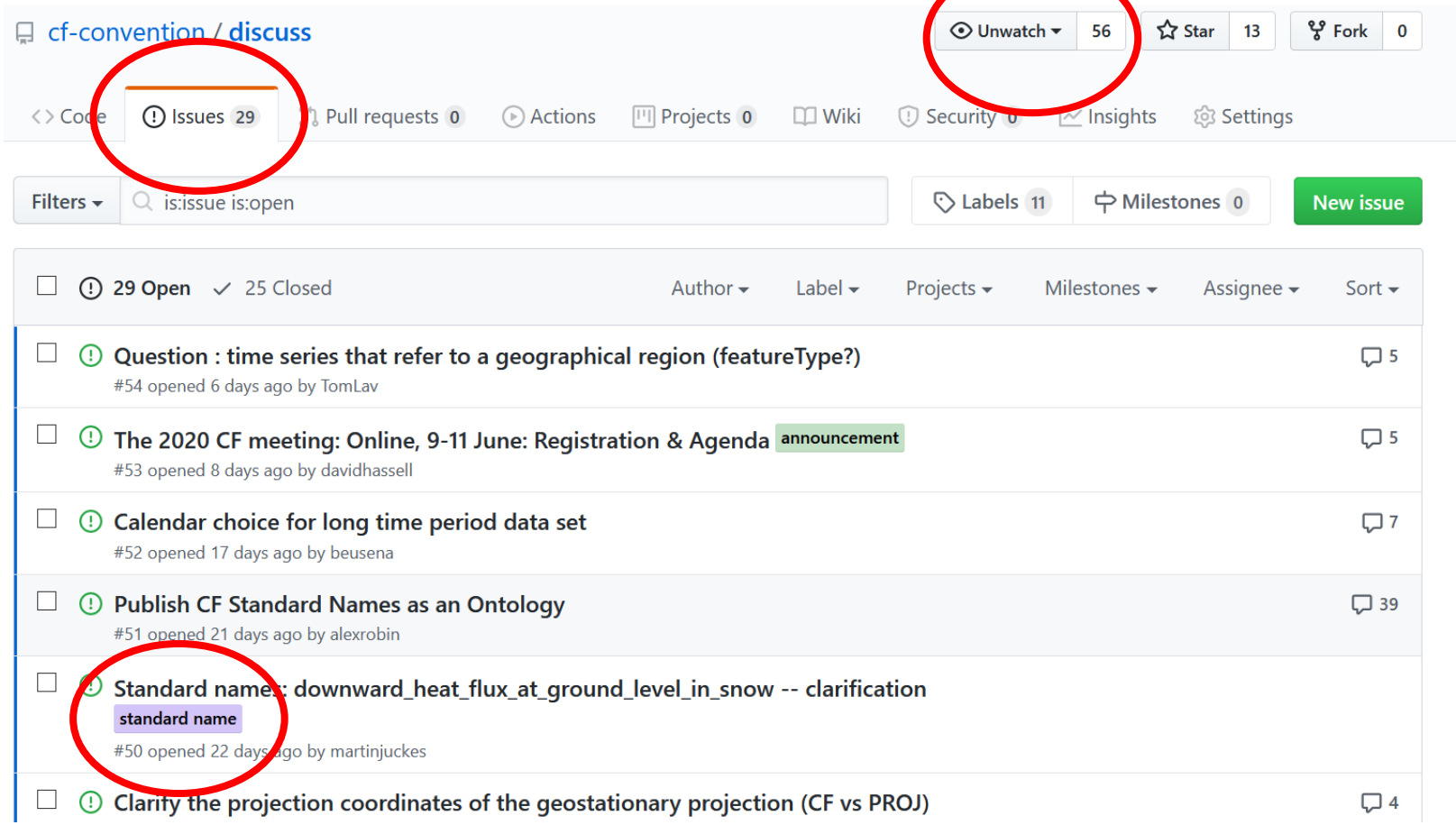
Sign up for an account on GitHub
https://github.com/



How do I ask for a new standard name?

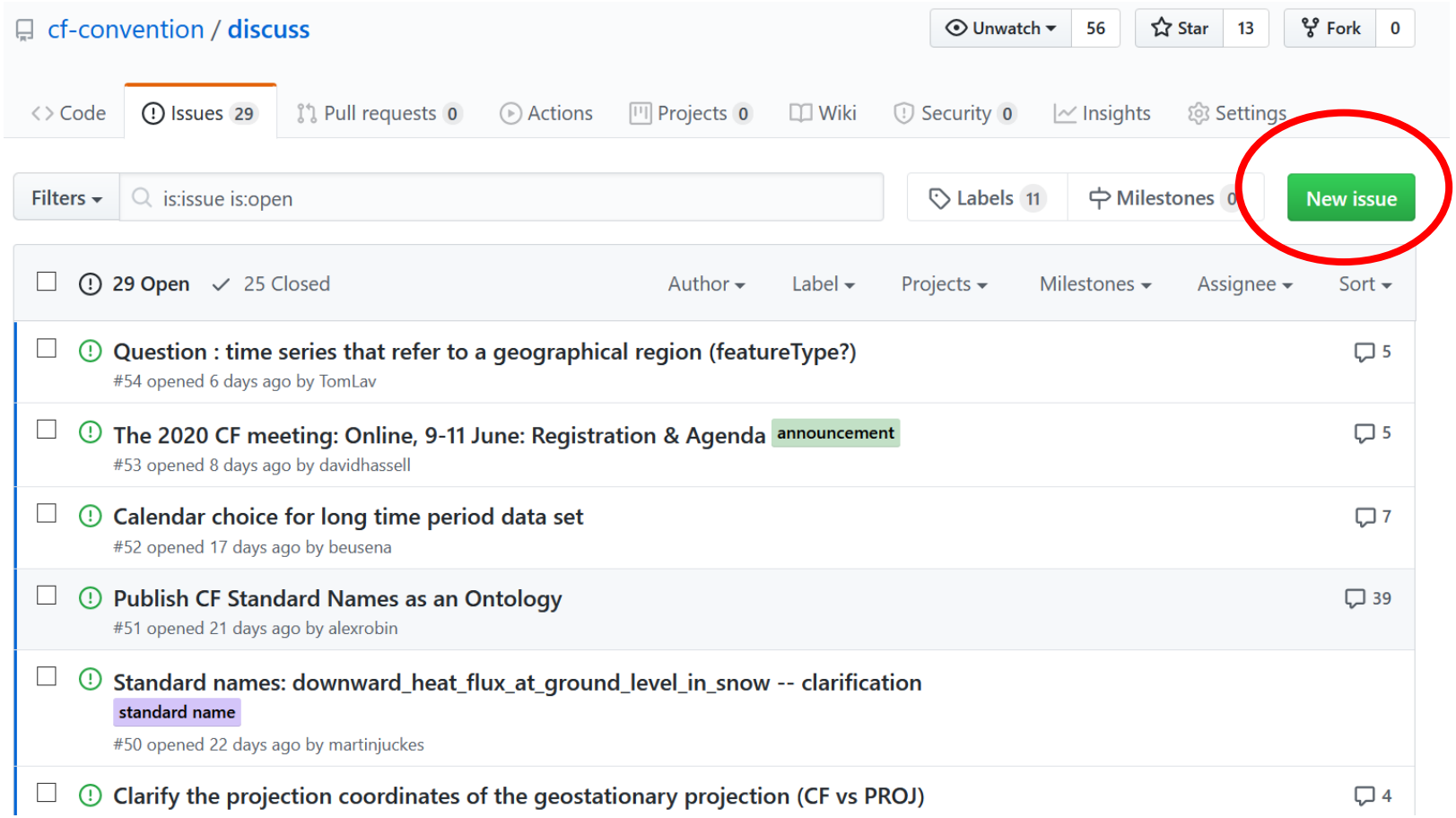
STEP 2

Navigate to the CF “discuss” repository (repo)
https://github.com/cf-convention/discuss



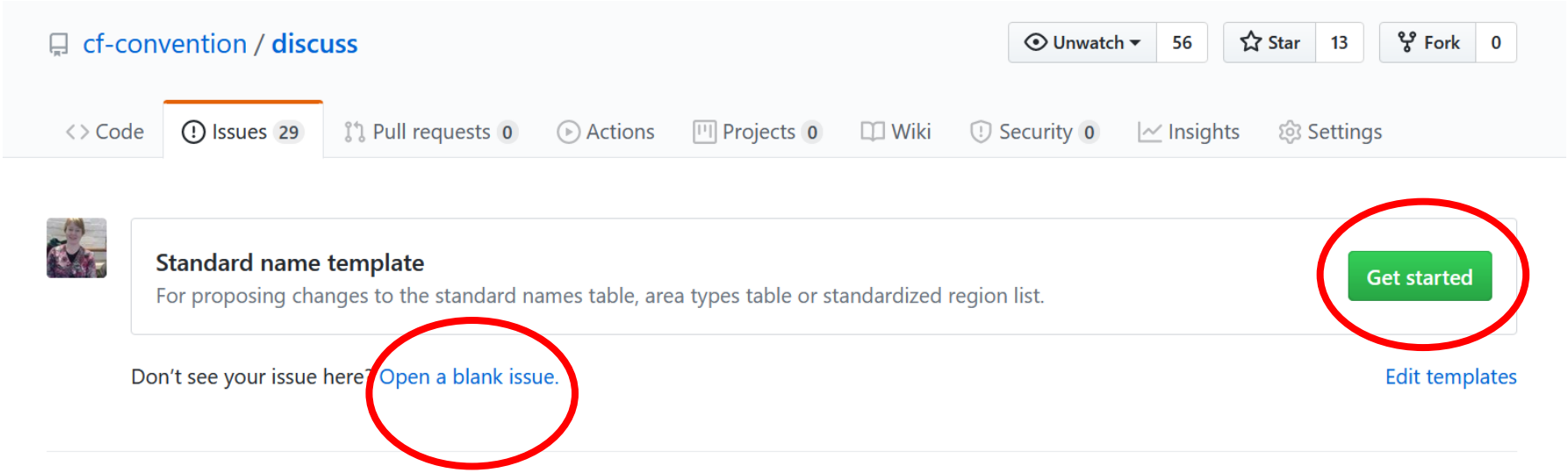
How do I ask for a new standard name?

STEP 3 Open a new issue



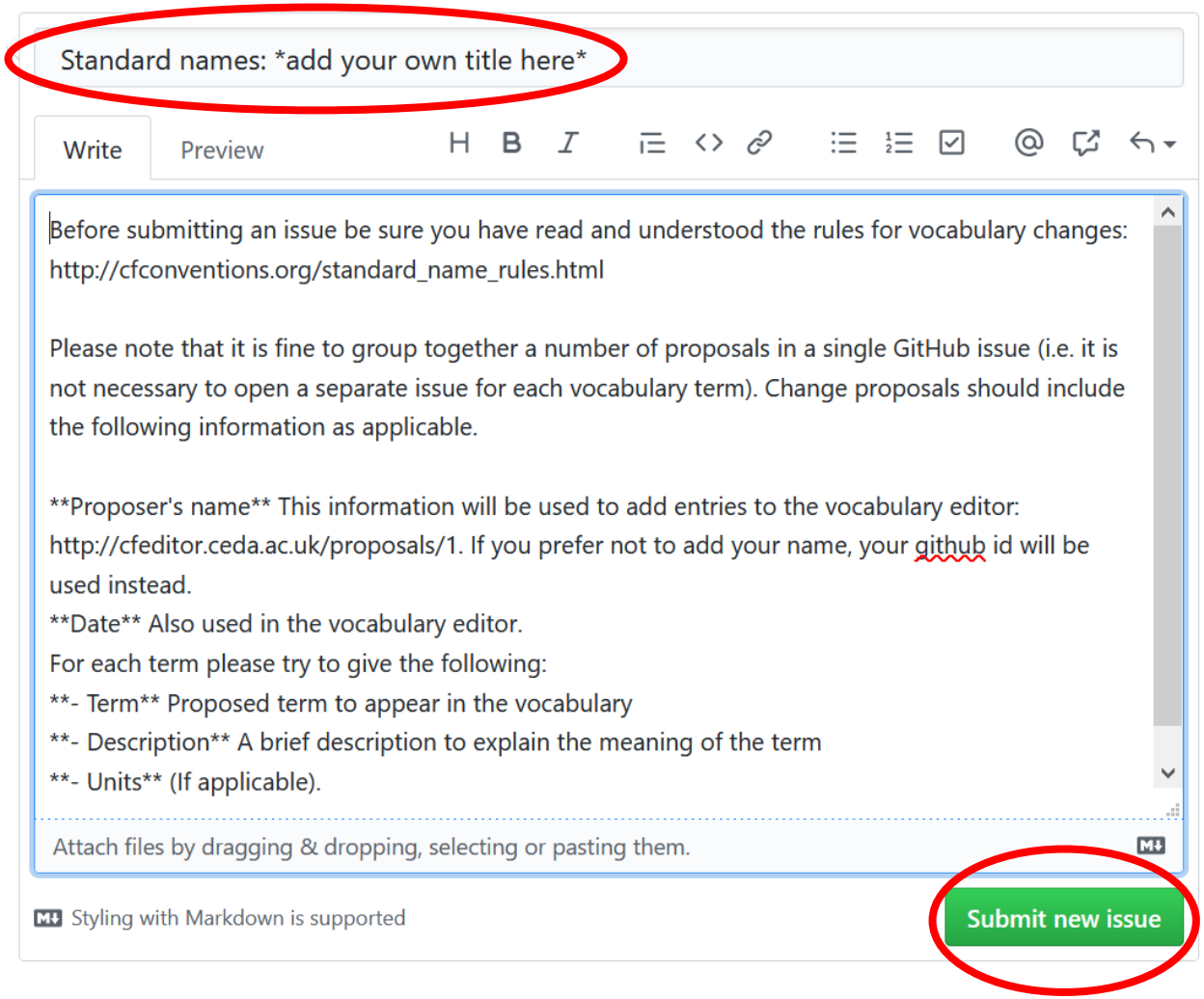
How do I ask for a new standard name?

STEP 4 Select the standard name template



How do I ask for a new standard name?

STEP 5 Use the template to write your proposal



CF standard name rules

- Any member of the community may comment on proposals
- Aim of the discussion is to achieve consensus (we now have two moderators!)
- Rules are laid out at http://cfconventions.org/standard_name_rules.html
- Provision for “fast tracking” new names that are very similar to existing terms, subject to checking
- CF standard names committee can be asked to vote if consensus cannot be achieved

Rules for acceptance / rejection of proposals

- A proposal will be **accepted** if one of the following is true:
  (a) it is similar to existing terms and has been checked for consistency by the moderator;
  (b) consensus has been reached in favour of the proposal;
  (c) the moderator's summary indicates that consensus in favour of the proposal has nearly been achieved;
  (d) a majority of the standard names committee vote to accept the proposal.

- A proposal will be **rejected** if one of the following is true:
  (a) it duplicates an existing vocabulary term;
  (b) consensus has been reached against the proposal;
  (c) the moderator's summary indicates that consensus against the proposal has nearly been achieved;
  (d) a majority of the standard names committee vote to reject the proposal;
  (e) the proposer withdraws the proposal.

CF standard name table

http://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html

Standard Name	Canonical Units
air_temperature_anomaly "Anomaly" means difference from climatology. Air temperature is the bulk temperature of the air, not the surface (skin) temperature.	**K**
surface_upward_latent_heat_flux The surface latent heat flux is the exchange of heat between the surface and the air on account of evaporation (including sublimation). In accordance with common usage in geophysical disciplines, "flux" implies per unit area, called "flux density" in physics.	**W m-2**

NERC Vocabulary Server (NVS2)

Collection (complete vocabulary):
vocab.nerc.ac.uk/collection/P07/current/
(CF standard names)

Single term:
http://vocab.nerc.ac.uk/collection/P07/current/CFV16A1/
(age_of_sea_ice)

OR

Single term:
http://vocab.nerc.ac.uk/standard_name/age_of_sea_ice/

UKRI Science and Technology Facilities Council | Natural Environment Research Council

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

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

CF collections in NVS2

CF standard_names: P07

CF area_type list: P29
CF standardized region list: P30

CF cell_methods list: P15
CF calendars: P37
CF vertical coverages: P38

Access methods: URL, SOAP, SPARQL endpoint
SPARQL: XML, JSON, text, CSV and TSV