Science & Technology Facilities Council

www.stfc.ac.uk

What's In a Name? Managing a controlled vocabulary for climate and forecast data

[1]Alison Pamment, [2]Calum Byrom, [1]Bryan Lawrence, [3]Roy Lowry
[1]NCAS/BADC,Science and Technology Facilities Council, [2]Tessella plc, [3]British Oceanogrphic Data Centre

Purpose and Scope of CF Metadata

❖ The Climate and Forecast (CF) metadata conventions were developed originally to describe numerical weather prediction (NWP) and climate model output.. The conventions are currently being extended to improve their usefulness for observational data.

❖The central aim is to facilitate the exchange of data between scientists worldwide for such undertakings as CMIP5 (5th Climate Model Intercomparison Project) and IPCC (Intergovernmental Panel on Climate Change) assessment reports.

❖The CF metadata conventions are generally used in conjunction with the netCDF file format, but many of the metadata concepts could equally well be applied to other file formats.

❖CF was originally developed by Brian Eaton (NCAR), Jonathan Gregory (Hadley Centre, UK Met Office), Bob Drach (PCMDI,LLNL), Karl Taylor (PCMDI,LLNL) and Steve Hankin (PMEL,NOAA).

CF Standard Names – a parameter vocabulary

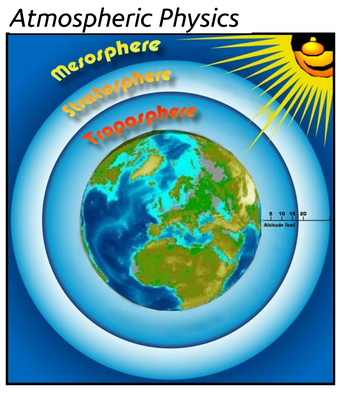
Each physical, chemical or biological quantity is given a CF standard name and assigned appropriate units. Equally importantly, a description of the quantity represented by each name is stored in the CF standard name table. The purpose of the standard name is to determine whether parameters from different data sources can be treated as directly comparable.

Standard Name	Canonical Units	AMIP	GRIB
air_potential_temperature	K	theta	13
air_pressure	Pa	plev	1
mass_fraction_of_sulfate_dry_aerosol_in_air	1		
ocean_mixed_layer_thickness	m		67

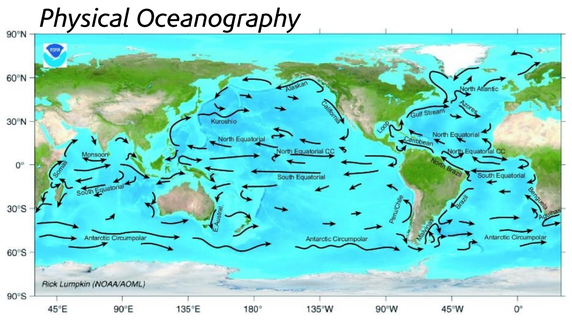
Science Domains

There are currently almost 2000 CF standard names to describe quantities in many science domains. New names are added in response to the needs of the science community.

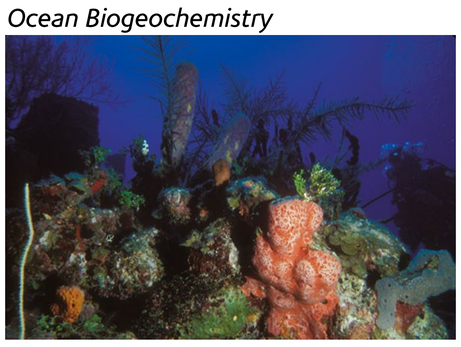
Atmospheric Physics



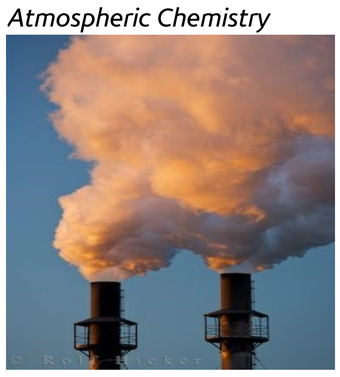
Physical Oceanography



Ocean Biogeochemistry



Atmospheric Chemistry



Cryosphere

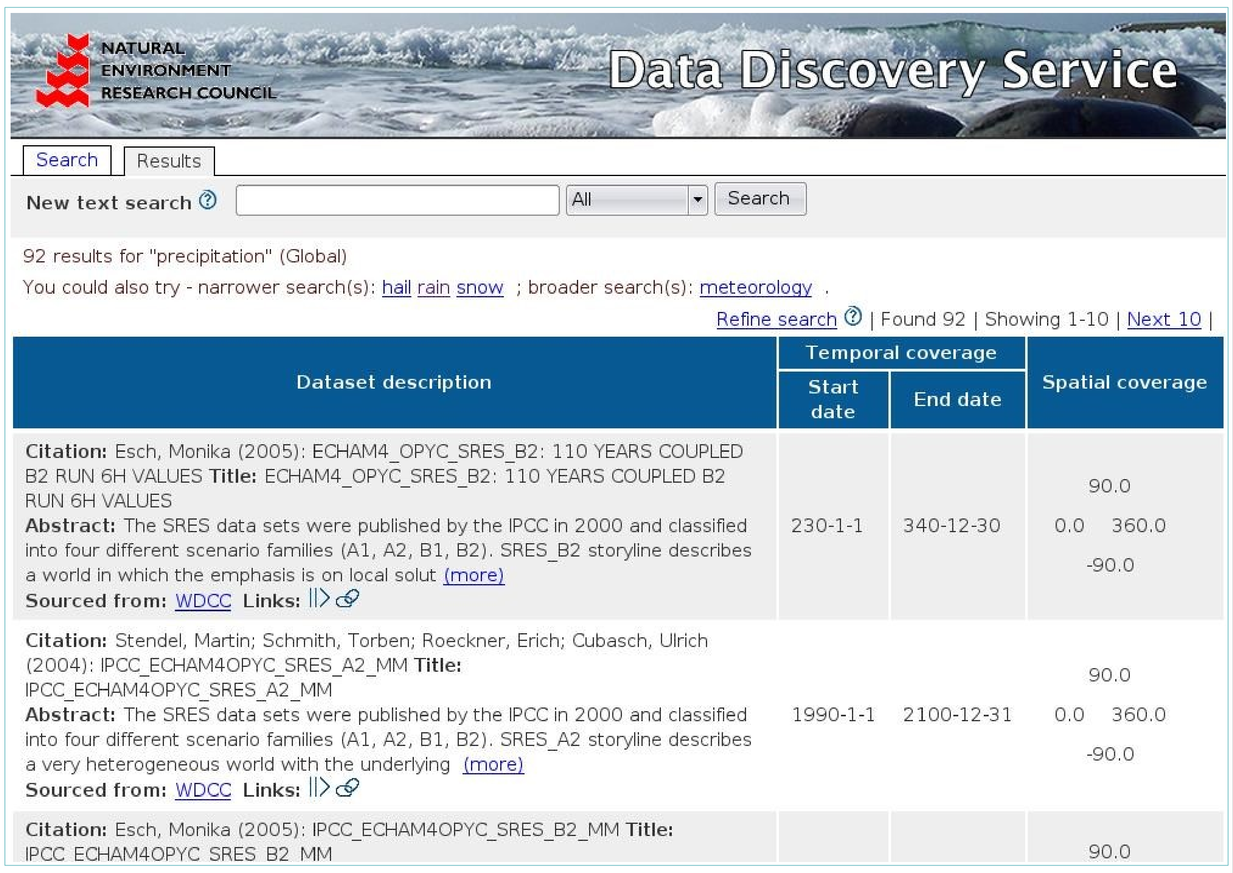
Vegetation and Carbon Cycle

Hydrology

Sea Surface Waves

Vocabulary Services

Web based services, for example, the NERC Data Discovery Service, can be built to exploit parameter names taken from controlled vocabularies such as CF standard names. Ontologies can be used to describe the relationships between terms within a vocabulary or between terms in different vocabularies. Ontologies facilitate the discovery of data through semantic searches, for example, a search for parameters containing the term 'precipitation' can yield results for exact matches, but also for related terms such as 'rain', 'snow', or 'hail'.



The NERC vocabulary server developed at the British Oceanographic Data Centre (BODC) is an underlying technology that allows semantic searching on parameter names. Currently, it holds around 150 controlled vocabularies including CF standard names. Whole vocabularies, groups of terms matching a string, or individual terms from a particular vocabulary can be requested from the server using URLs and the results are returned as XML documents.

A vocabulary editor is currently under development at BADC. This will provide a web based interface for the direct maintenance of vocabularies held in the NERC vocabulary server by the individuals responsible for vocabulary content. CF standard names have been used as a test case in the development of the editor.

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

NATURAL ENVIRONMENT RESEARCH COUNCIL

Tessella Technology & Consulting

Funding for the development of the CF metadata conventions is provided by PCMDI (Program for Climate Model Data Intercomparison), Lawrence Livermore National Laboratory, USA and by NERC (Natural Environment Research Council) through NCAS (National Centre for Atmospheric Science), UK.