Overview

Two broad objectives:

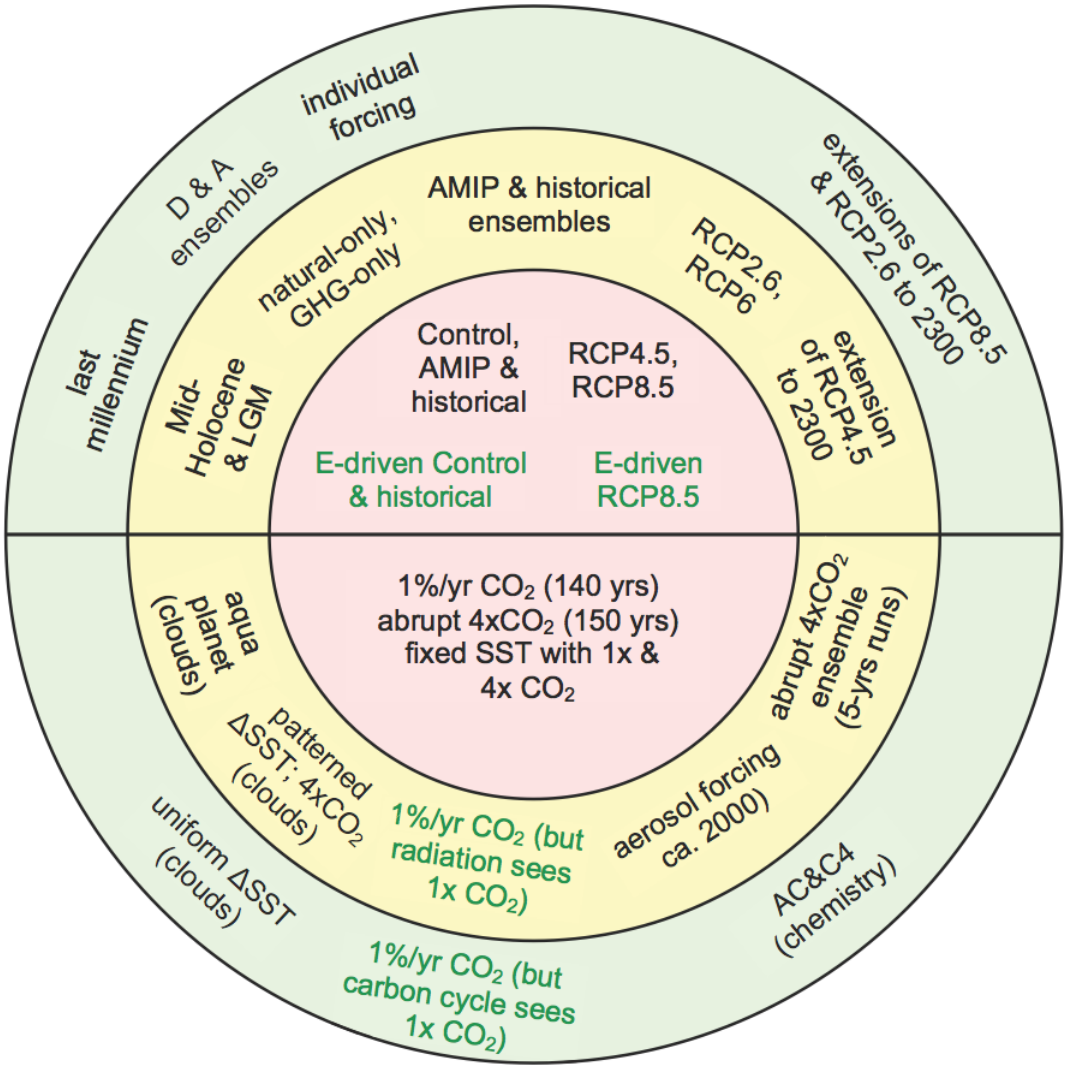
Introduction to working with climate data

1. Explore how the IPCC AR5 data are organized
2. Introduce HDF/NetCDF as a data format
3. Use Climate Data Operators (CDO) from within R to process the daily data to climate metrics
4. Make a few plots

Illustrate "Literate Programing"

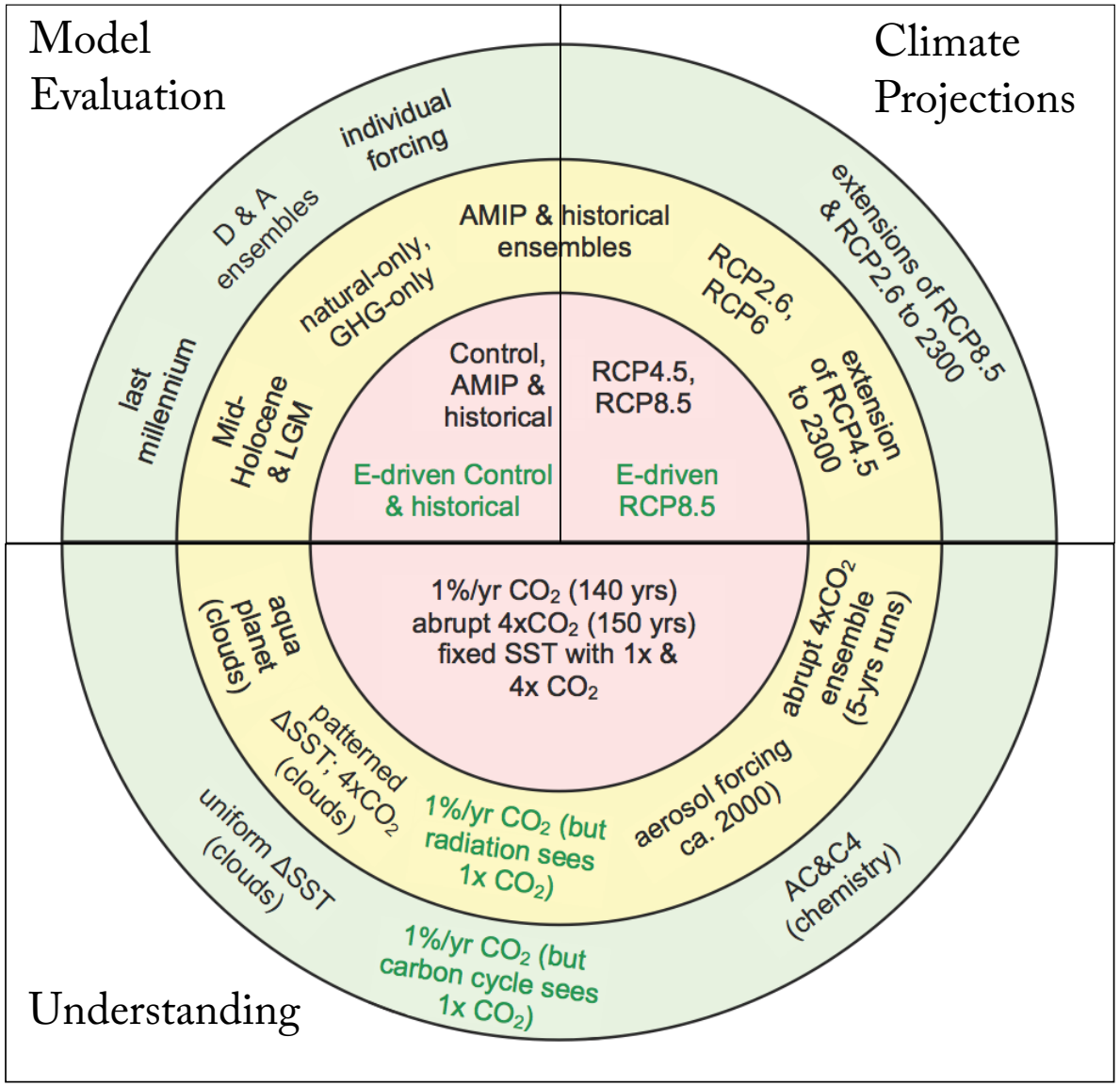
1. Use of R+Markdown to generate repeatable, "human-readable" reports of an analysis.

Coupled Model Intercomparison Project Phase 5 (CMIP5)



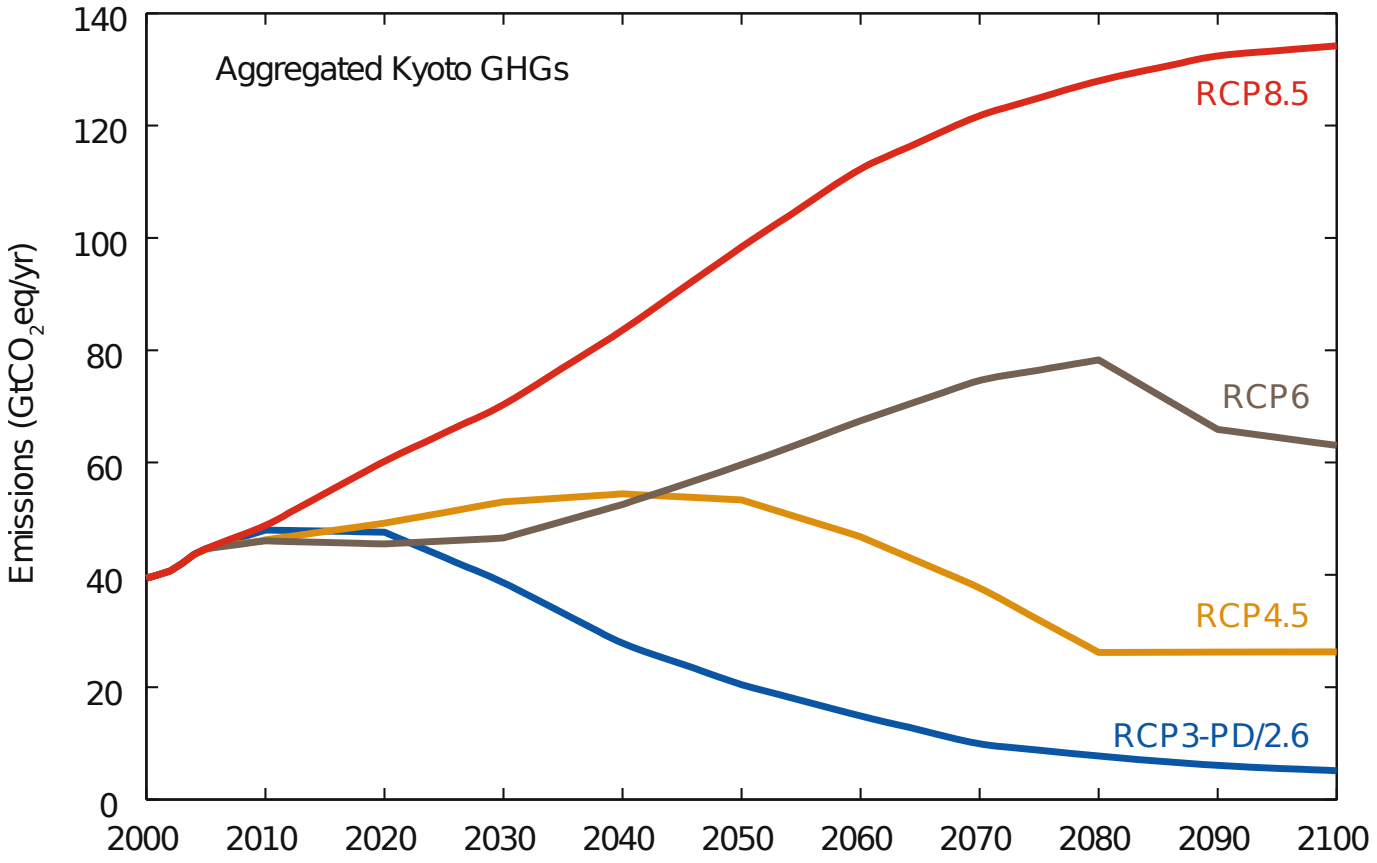
Taylor, et.al., (2012) BAMS 93(4):485–498.

Coupled Model Intercomparison Project Phase 5 (CMIP5)



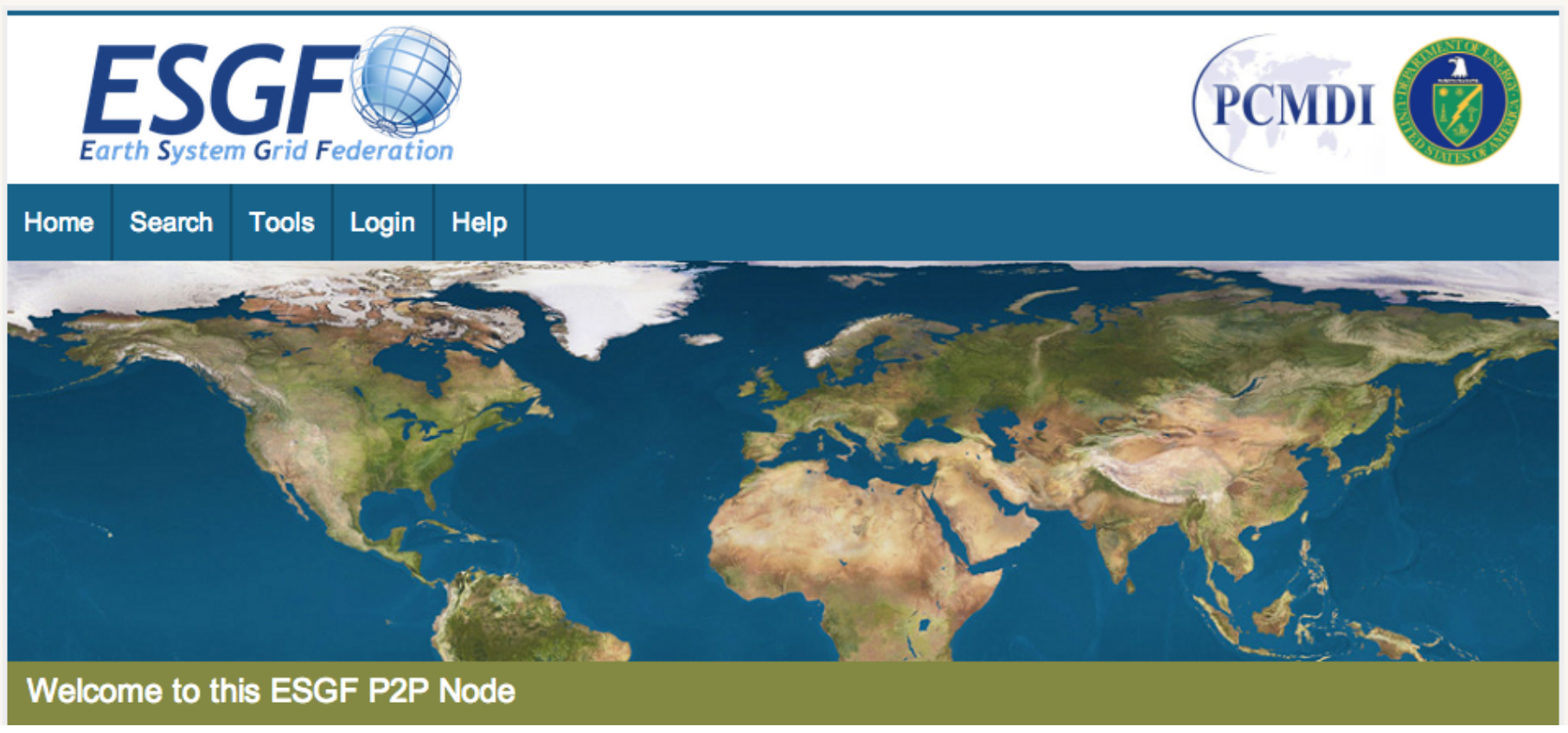
Taylor, et.al., (2012) BAMS 93(4):485–498.

Representative Concentration Pathways (RCP)



Plus several historical scenarios (‘natural’, ‘green house gases’, etc.)

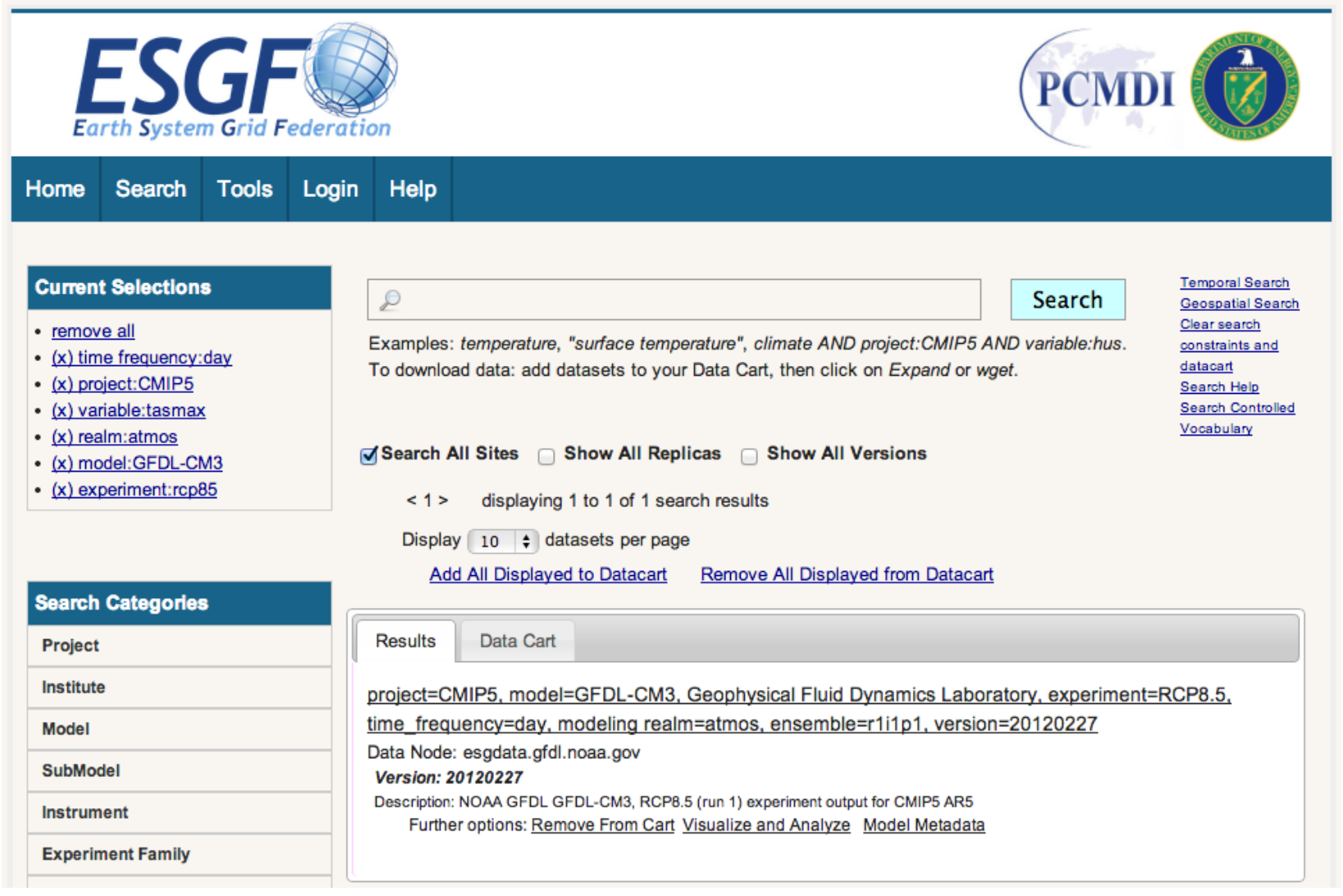
Earth System Grid Federation (ESGF)



`http://pcmdi9.llnl.gov/`

Downloads currently not available....

Filter via ESGF webpage http://pcmdi9.llnl.gov/

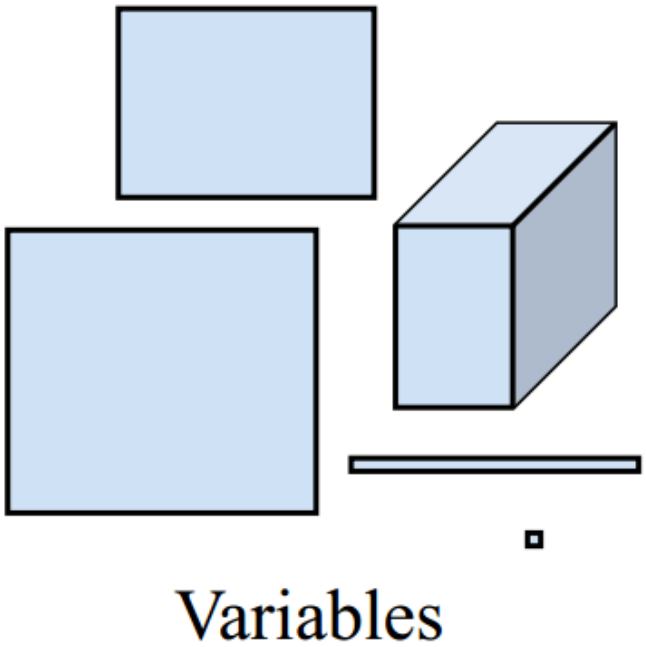


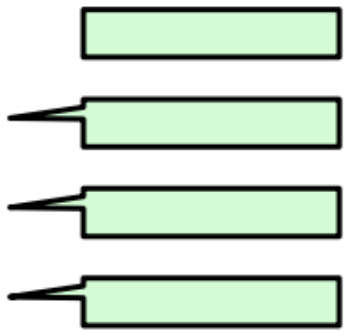
ESGF provides a `wget` script to download the selected data

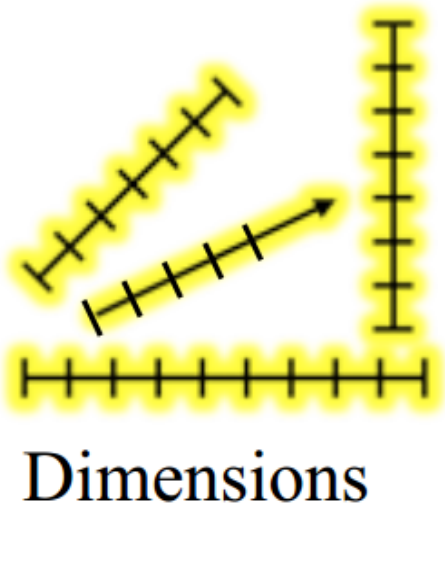
```
#!/bin/bash
##################################################################
# ESG Federation download script
#
# Template version: 1.2
# Generated by pcmdi9.llnl.gov - 2013/10/09 12:02:14
# Search URL: http://pcmdi9.llnl.gov/esg-search/wget/?query=*&
#       dataset_id= cmip5.output1.NOAA-GFDL.GFDL-CM3 \\
#      .rcp85.day.atmos.day.r1i1p1.v20120227|esgdata.gfdl.noaa.gov
#
##################################################################
. . .
version=1.3.2
CACHE_FILE=.$(basename $0).status
openId=
search_url='http://pcmdi9.llnl.gov/esg-search/wget/?query=*&
       dataset_id=cmip5.output1.NOAA-GFDL.GFDL-CM3.rcp85.day.atmos.day.r1i1p1.v20120227| \\
       esgdata.gfdl.noaa.gov'
#These are the embedded files to be downloaded
download_files="$(cat <<EOF--dataset.file.url.chksum_type.chksum
'clt_day_GFDL-CM3_rcp85_r1i1p1_20960101-21001231.nc'
'http://esgdata.gfdl.noaa.gov/thredds/fileServer/gfdl_dataroot/NOAA-GFDL/GFDL-CM3/
rcp85/day/atmos/day/r1i1p1/v20110601/clt/clt_day_GFDL-CM3_rcp85
_r1i1p1_20960101-21001231.nc' 'MD5' '801fc44c406377a084f87c4f509e9da8'
'clt_day_GFDL-CM3_rcp85_r1i1p1_20910101-20951231.nc'
. . .
```

What is NetCDF?

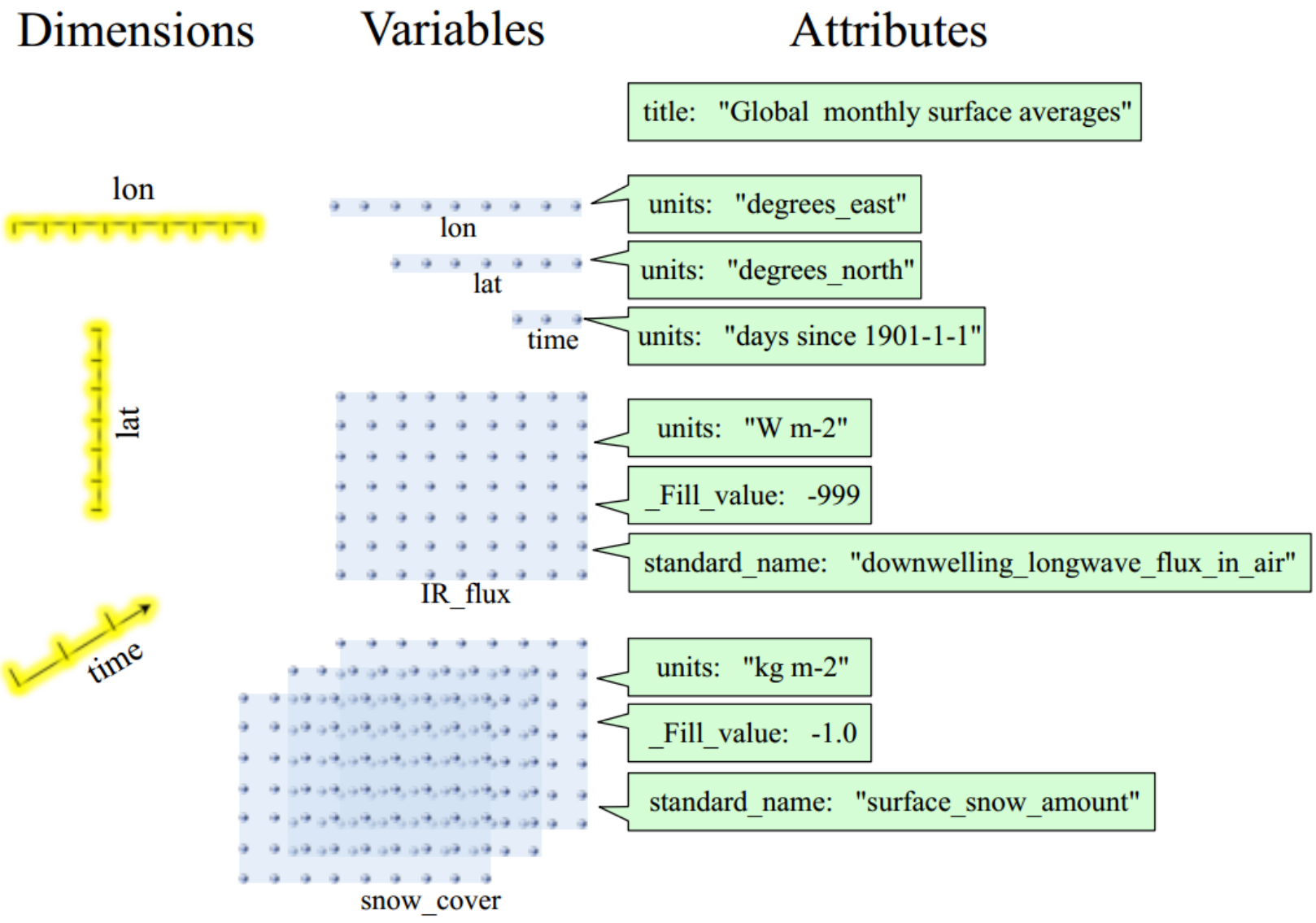
- A netCDF file has named variables,attributes, and dimensions.
- Variables → data
- Attibutes → metadata (for file or variable)
- Dimensions → Shapes of variables
- Variables may share dimensions, indicating a common grid
- Each variable or attribute ∈ char, byte, short, int, float, double







What is NetCDF?



What is NetCDF?

Notation for NetCDF in CDL (Common Data Language)

```
netcdf snow{
    dimensions:
        lon= 9 ;
        lat= 7 ;
        time = unlimited ;  // 3 currently
    variables:
        float IR\_flux(lon, lat) ;
        IR_flux:units = "W m-2" ;
        IR_flux:_Fill_value = -999 ;
        IR_flux:standard_name=  "downwelling_longwave";
        float snow_cover(time, lon, lat) ;
        snow_cover:units = "kg m-2" ;
        ...
        // global attributes
        :title  =  "simple  example,  lacks  some  conventions"  ;
    data:
        IR_flux = 200, 201, 202 ;
        snow_cover = 0.1, 0.2, 0.0, ... ;
      }
```

NetCDF Characteristics?

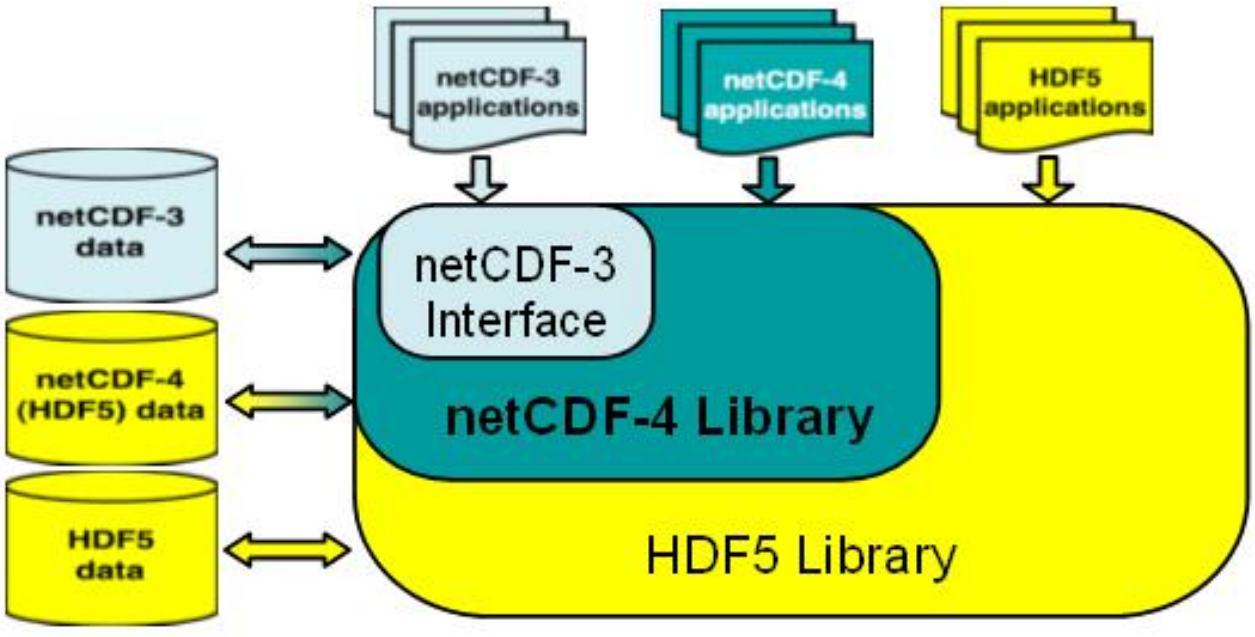
- Self-Describing: A netCDF file includes all metadata needed to identify the data in time and space, units of measure, and other useful information.
- Portable: Data written on one platform can be read on other platforms
- Direct-access: A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- Appendable: Data may be efficiently added to a netCDF file without copying the dataset or redefining its structure.
- Extensible: Adding new dimensions, variables, or attributes to netCDF files does not require changes to existing programs that read the files.
- Sharable: One writer and multiple readers may simultaneously access the same netCDF file. With Parallel netCDF, multiple writers may efficiently and concurrently write into the same netCDF file.
- Archivable: Access to all earlier forms of netCDF data will be supported by current and future versions of the software.
- Networkable: The netCDF library provides client access to structured data on remote servers through OPeNDAP protocols.

Commitment to Compatibility

To ensure future access to existing data archives, Unidata is committed to compatibility of:

- **Data access**: new versions of netCDF software will provide read and write access to previously stored netCDF data.
- **Programming interfaces**: C and Fortran programs using documented netCDF interfaces from previous versions will work without change with new versions of netCDF software.
- **Future versions**: Unidata will continue to support both data access compatibility and program compatibility in future netCDF releases.

NetCDF3, NetCDF4, HDF4, HDF5, HDF-EOS?



http://www.unidata.ucar.edu/projects/THREDDS/CDM/CDM-TDS.htm

NetCDF Climate and Forecast (CF) Metadata Conventions

Provide a definitive description of what the data in each variable represents, and of the spatial and temporal properties of the data.

Attribute	Type	Use	Links	Description
add_offset	N	D	~~NUG (8.1)~~ NUG (8.1), Section 8.1, "Packed Data"	If present for a variable, this number is to be added to the data after it is read by an application. If both scale_factor and add_offset attributes are present, the data are first scaled before the offset is added.
ancillary_variables	S	D	Section 3.4, "Ancillary Data"	Identifies a variable that contains closely associated data, e.g., the measurement uncertainties of instrument data.
axis	S	C	Chapter 4, *Coordinate Types*	Identifies latitude, longitude, vertical, or time axes.
bounds	S	C	Section 7.1, "Cell Boundaries"	Identifies a boundary variable.
calendar	S	C	Section 4.4.1, "Calendar"	Calendar used for encoding time axes.
cell_measures	S	D	Section 7.2, "Cell Measures"	Identifies variables that contain cell areas or volumes.
cell_methods	S	D	Section 7.3, "Cell Methods", Section 7.4, "Climatological Statistics"	Records the method used to derive data that represents cell values.

http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.6/cf-conventions.html

Disadvantages

- NetCDF3 works everywhere, but NetCDF4 must run in Cygwin on Window$ (though this may change very soon)
- Less familiar (though this is changing)
- More complicated for simple datasets

CDO - NetCDF Climate Data Operators

```
bash $ cdo <options> <operator> input.nc out.nc
```

It's that easy!

- file information:
  `cdo sinfo file.nc`
- file operations (copy, split, merge):
  `cdo mergetime y2000.nc y2001.nc output.nc`
- arithmetic, daily/monthly/annual summaries, interpolation, etc:
  `cdo monmean input.nc output.nc`

`https://code.zmaw.de/embedded/cdo/1.5.5/cdo.html`

CDO Climate indices

2.16.1 ECACDD - Consecutive dry days index per time period
2.16.2 ECACFD - Consecutive frost days index per time period
2.16.3 ECACSU - Consecutive summer days index per time period
2.16.4 ECACWD - Consecutive wet days index per time period
2.16.5 ECACWDI - Cold wave duration index w.r.t. mean of reference period
2.16.6 ECACWFI - Cold-spell days index w.r.t. 10th percentile of reference period
2.16.7 ECAETR - Intra-period extreme temperature range
2.16.8 ECAFD - Frost days index per time period
2.16.9 ECAGSL - Thermal Growing season length index
2.16.10 ECAHD - Heating degree days per time period
2.16.11 ECAHWDI - Heat wave duration index w.r.t. mean of reference period
2.16.12 ECAHWFI - Warm spell days index w.r.t. 90th percentile of reference period
2.16.13 ECAID - Ice days index per time period
2.16.14 ECAPD - Precipitation days index per time period
2.16.15 ECAR75P - Moderate wet days w.r.t. 75th percentile of reference period
2.16.16 ECAR75PTOT - Precipitation percent due to R75p days
2.16.17 ECAR90P - Wet days w.r.t. 90th percentile of reference period
2.16.18 ECAR90PTOT - Precipitation percent due to R90p days
2.16.19 ECAR95P - Very wet days w.r.t. 95th percentile of reference period
2.16.20 ECAR95PTOT - Precipitation percent due to R95p days
2.16.21 ECAR99P - Extremely wet days w.r.t. 99th percentile of reference period
2.16.22 ECAR99PTOT - Precipitation percent due to R99p days
2.16.23 ECARR1 - Wet days index per time period
2.16.24 ECARX1DAY - Highest one day precipitation amount per time period
2.16.25 ECARX5DAY - Highest five-day precipitation amount per time period
2.16.26 ECASDII - Simple daily intensity index per time period
2.16.27 ECASU - Summer days index per time period
2.16.28 ECATG10P - Cold days percent w.r.t. 10th percentile of reference period
2.16.29 ECATG90P - Warm days percent w.r.t. 90th percentile of reference period
2.16.30 ECATN10P - Cold nights percent w.r.t. 10th percentile of reference period
2.16.31 ECATN90P - Warm nights percent w.r.t. 90th percentile of reference period
2.16.32 ECATR - Tropical nights index per time period
2.16.33 ECATX10P - Very cold days percent w.r.t. 10th percentile of reference period
2.16.34 ECATX90P - Very warm days percent w.r.t. 90th percentile of reference period

CDO: piping commands

Calculate the number of consecutive frost days over several years:

```
cdo selname,tmin input.nc output1.nc
cdo selyear,2000 output1.nc output2.nc ...
cdo eca_cfd output2.nc output3.nc
cdo mergetime output3.nc output.nc
```

Or shorter using operator piping:

```
cdo mergetime -eca_cfd -selyear,2000 -selname,tmin input.nc
-selyear,2001 -selname,tmin input.nc ...  output.nc
```

Improves the performance by:

- reducing unnecessary disk I/O
- parallel processing

NCO - NetCDF Operators

netCDF Operators (NCO) Software Stack

- hyperslabbing
- processing over arbitrary dimensions
- attribute editing
- scripting for complicated (or simple) operations
  `ncap -s "RAIN=RAINC+RAINNC" in.nc out.nc`
- access network files via Open-source Project for a Network Data Access Protocol (OPeNDAP)

```
ncwa -C -a lat,lon,time -d lon,-10.,10.  -d lat,-10.,10.  -p
http://.../pres.sfc.1969.nc foo.nc
```

Other Software of interest

- GDAL/RGDAL (http://www.gdal.org/frmt_netcdf.html)

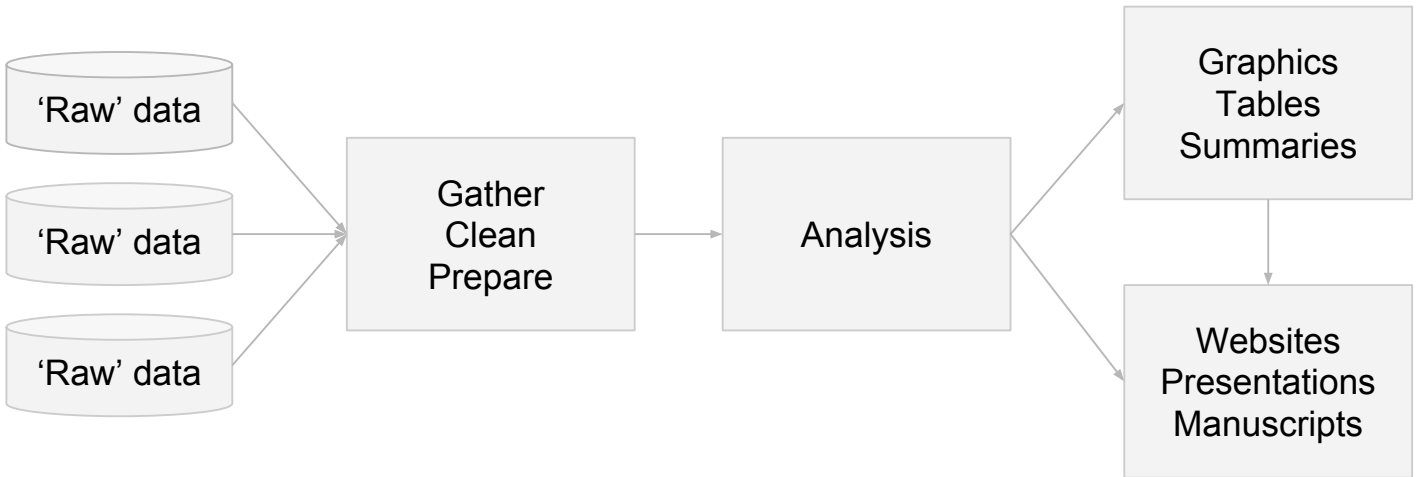
```
$ gdalinfo sst.nc
Driver:  netCDF/Network Common Data Format
Size is 512, 512
Metadata:
NC_GLOBALtitle=IPSL model output for IPCC FAR SRES A2
Subdatasets:
SUBDATASET_1_NAME=NETCDF:"sst.nc":lon_bnds
...
```

- R: raster package

```
ne=brick("mod09_CT.nc",varname="ndvi")
ne2=mean(ne,na.rm=T)
plot(ne2)
```

- ArcGIS
- ParaView (http://www.paraview.org/), Panoply (http://www.giss.nasa.gov/tools/panoply/)
- And more: http://www.unidata.ucar.edu/software/netcdf/software.html

Example Workflow (again...)



`wget`
`APIs`

`gdal`
`pktools`
`GRASS`
`merge`
`reshape`

`BUGS/JAGS`
`Models`

`xtable`
`knitr/sweave`

Let's link the various steps together

Literate Programming

Literate Programming with R

- RStudio $\rightarrow$ nice tools
- markdown or LaTeX