

CMIP5
ooooo

Intro to NetCDF
oooooooo

NetCDF Software
ooooo

Literate Programming
oooo

Introduction to NetCDF, Climate Data Processing, and “Literate Programming”

Adam M. Wilson

November 15, 2012

Overview

Two broad objectives:

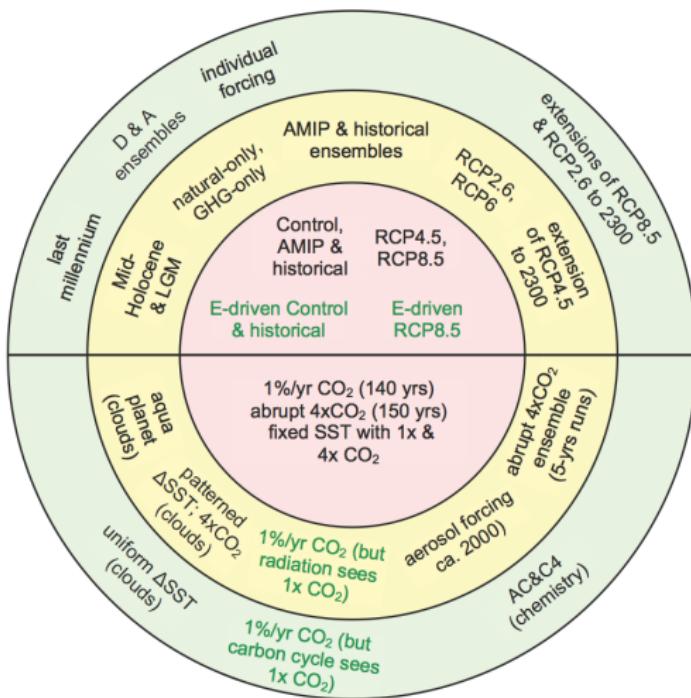
Introduction to working with climate data

1. Explore how the IPCC AR5 data are organized
2. Introduce HDF/NetCDF as a data format
3. Introduce Climate Data Operators (CDO)
4. Call CDO from R to process daily data into climate metrics
5. Do a few simple analyses (plots, tables, etc.)

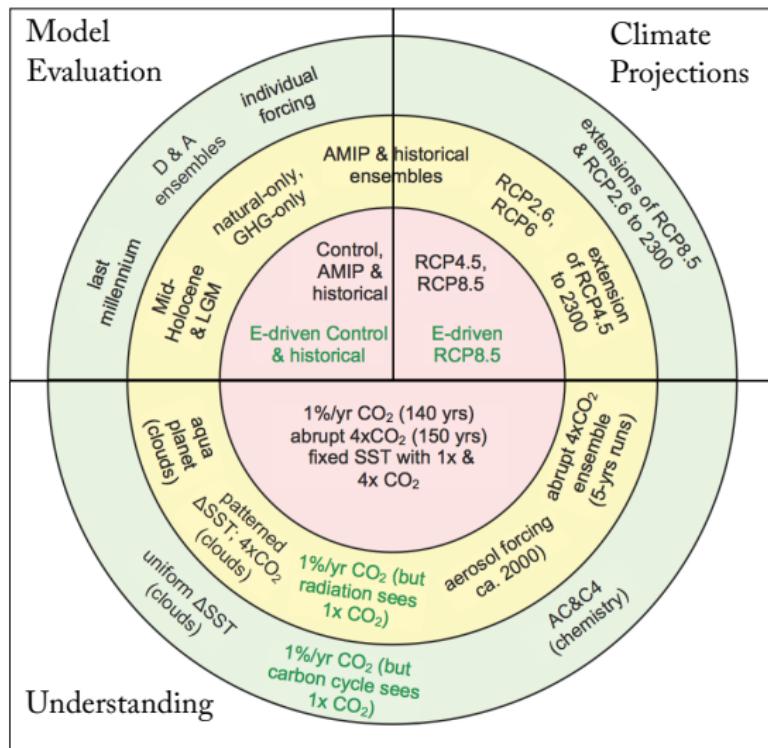
Illustrate “Literate Programming”

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

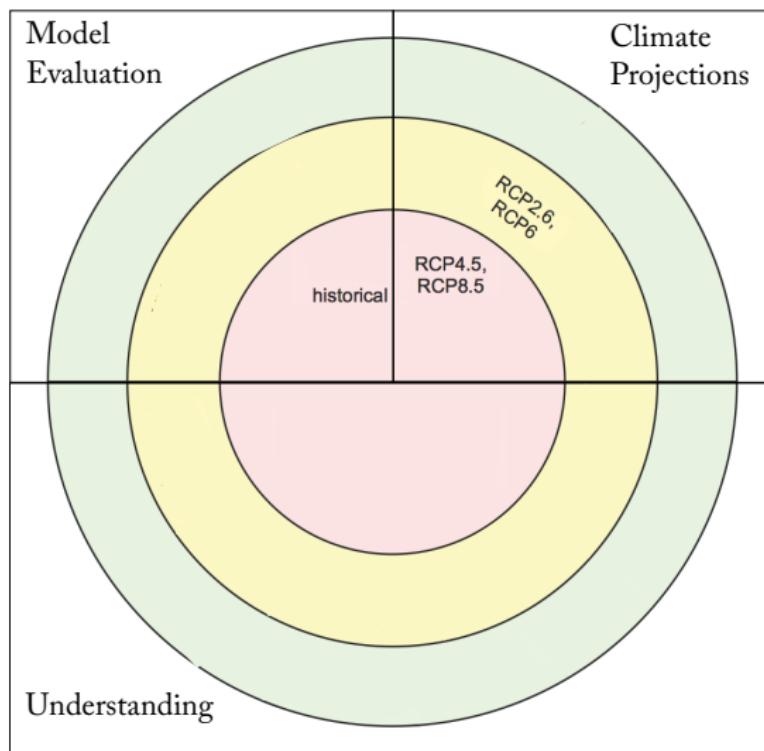
Coupled Model Intercomparison Project Phase 5 (CMIP5)



Coupled Model Intercomparison Project Phase 5 (CMIP5)



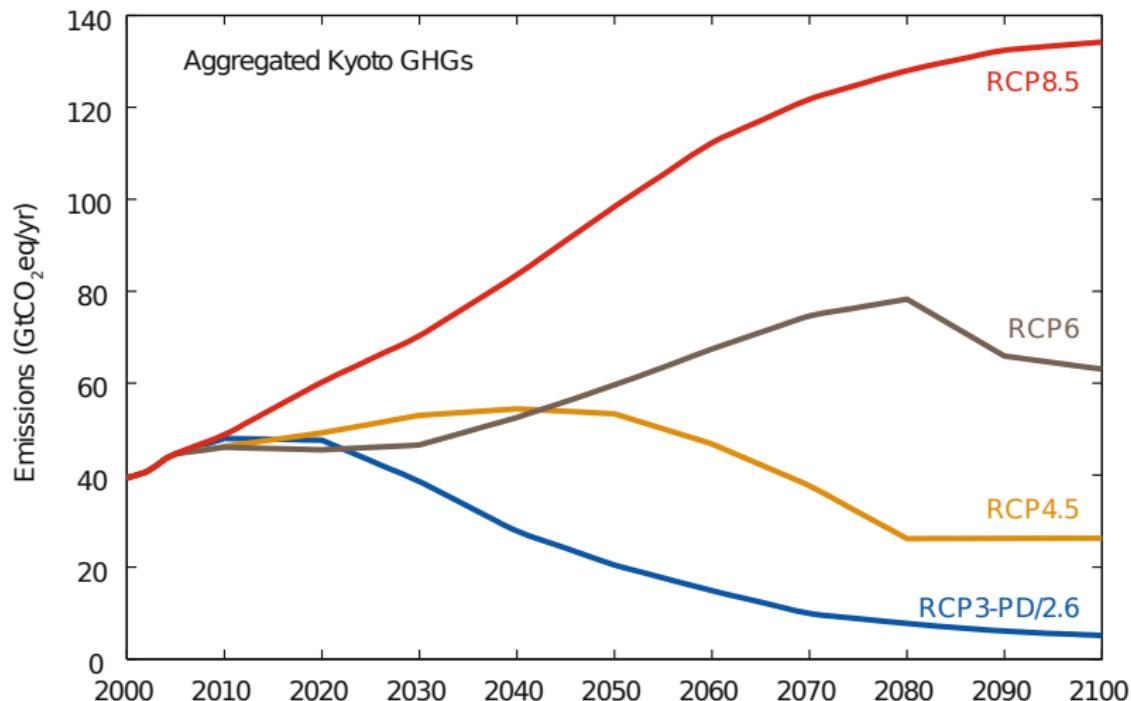
Coupled Model Intercomparison Project Phase 5 (CMIP5)



RCP8.5 increases to 8.5 W m^{-2} in 2100

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

Representative Concentration Pathways (RCP)



Earth System Grid Federation (ESGF)

The screenshot shows the homepage of the Earth System Grid Federation (ESGF). At the top left is the ESGF logo with the text "Earth System Grid Federation". To the right are the PCMDI logo (a globe with "PCMDI" overlaid) and the University of California Berkeley logo. A blue navigation bar at the top contains links for "Home", "Search", "Tools", "Login", and "Help". Below the navigation bar is a large world map showing land and ocean. At the bottom of the page, a green banner displays the text "Welcome to this ESGF P2P Node".

<http://pcmdi9.llnl.gov/>
Downloads currently not available....

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



[Home](#) [Search](#) [Tools](#) [Login](#) [Help](#)

Current Selections

- [remove all](#)
- (x) time frequency:day
- (x) project:CMIP5
- (x) variable:tasmax
- (x) realm:atmos
- (x) model:GFDL-CM3
- (x) experiment:rcp85

Examples: *temperature*, "surface temperature", climate AND project:CMIP5 AND variable:hus.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Search All Sites Show All Replicas Show All Versions

< 1 > displaying 1 to 1 of 1 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Search Categories

Project
Institute
Model
SubModel
Instrument
Experiment Family

Results Data Cart

[project=CMIP5, model=GFDL-CM3, Geophysical Fluid Dynamics Laboratory, experiment=RCP8.5, time_frequency=day, modeling realm=atmos, ensemble=r1i1p1, version=20120227](#)

Data Node: esgdata.gfdl.noaa.gov
Version: 20120227

Description: NOAA GFDL GFDL-CM3, RCP8.5 (run 1) experiment output for CMIP5 AR5

Further options: Remove From Cart Visualize and Analyze Model Metadata

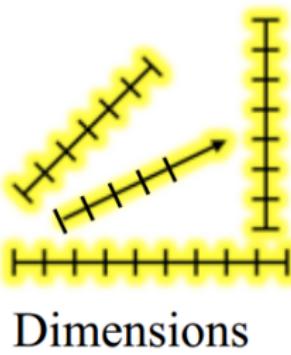
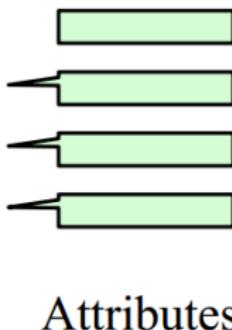
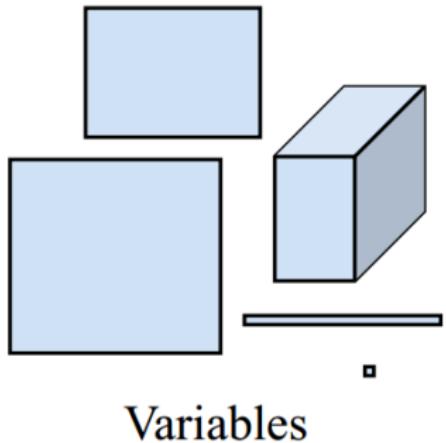
Temporal Search
[Geospatial Search](#)
[Clear search](#)
[constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

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'
.
.
```

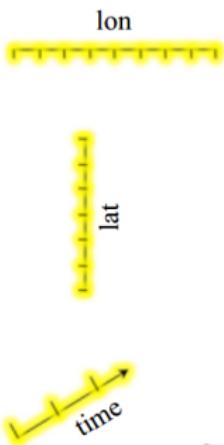
What is NetCDF?

- A netCDF file has named variables, attributes, and dimensions.
- Variables → data
- Attributes → 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

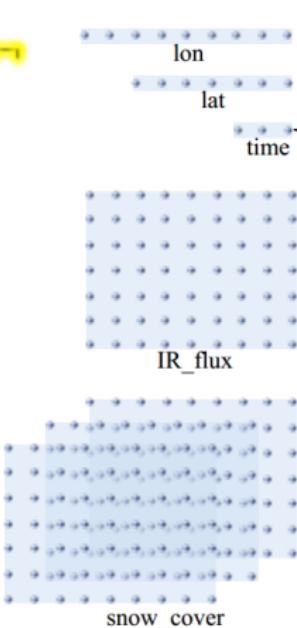


What is NetCDF?

Dimensions



Variables



Attributes

title: "Global monthly surface averages"	
units: "degrees_east"	units: "degrees_north"
units: "days since 1901-1-1"	
units: "W m-2"	
_Fill_value: -999	
standard_name: "downwelling_longwave_flux_in_air"	
units: "kg m-2"	
_Fill_value: -1.0	
standard_name: "surface_snow_amount"	

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.

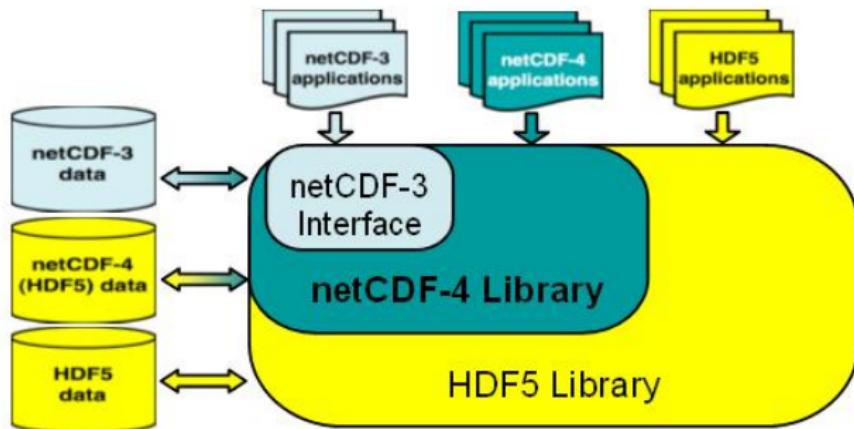
Commitment to Compatibility

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?



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
<code>add_offset</code>	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 <code>scale_factor</code> and <code>add_offset</code> attributes are present, the data are first scaled before the offset is added.
<code>ancillary_variables</code>	S	D	Section 3.4, "Ancillary Data"	Identifies a variable that contains closely associated data, e.g., the measurement uncertainties of instrument data.
<code>axis</code>	S	C	Chapter 4, Coordinate Types	Identifies latitude, longitude, vertical, or time axes.
<code>bounds</code>	S	C	Section 7.1, "Cell Boundaries"	Identifies a boundary variable.
<code>calendar</code>	S	C	Section 4.4.1, "Calendar"	Calendar used for encoding time axes.
<code>cell_measures</code>	S	D	Section 7.2, "Cell Measures"	Identifies variables that contain cell areas or volumes.
<code>cell_methods</code>	S	D	Section 7.3, "Cell Methods" , Section 7.4, "Climatological Statistics"	Records the method used to derive data that represents cell values.

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`

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
- 2.16.6 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 ECAFDF - 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
- 2.16.20 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
- 2.16.26 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

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



- 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/>)

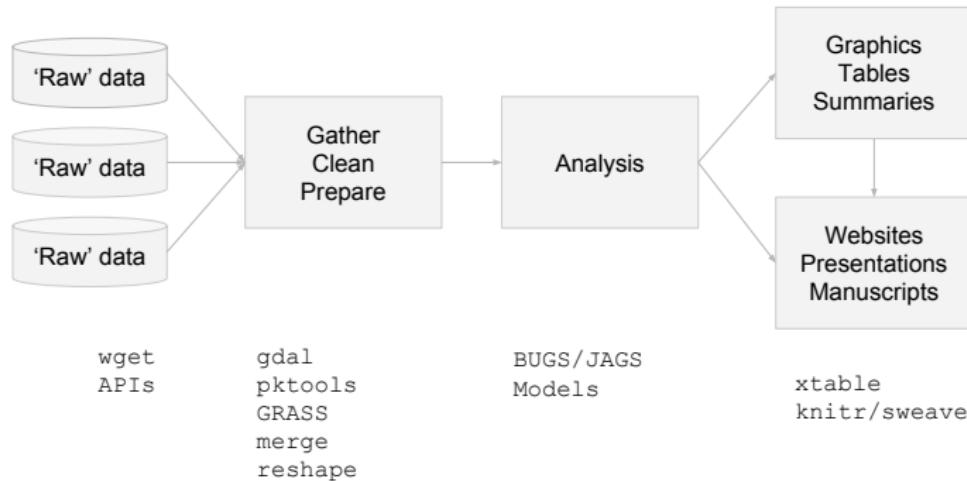
- And more: <http://www.unidata.ucar.edu/software/netcdf/software.html>

Literate Programming

"... an explanation of a program in a natural language, such as English, interspersed with snippets of code"

```
210  ### Calculate the mean change BY SEASON for all the variables in the GFDL_RCM3 dataset
211  The desired output will be a netcdf file with 4 time steps, the mean change for each of the 4 seasons.
212
213 - ````{r}
214  system("cdo sub -yseasmean ../data/GFDL_Future.nc -yseasmean ../data/GFDL_Current.nc gfdl_seasdif.nc")
215  ...
216
217 Read the file into R using the 'brick' command to read in all 4 time steps from tmean as separate raster images. Then update
update the column names (to spring, summer, etc.) and plot the change in each of the four seasons.
```

Example Workflow (again...)



Let's link the various steps together

Literate Programming with R

text+code → html/docx/pdf with RStudio

```
1 Title
2 =====
3
4 This is an R Markdown document. Markdown is a simple formatting syntax for authoring web
pages (click the **MD** toolbar button for help on Markdown).
5
6 When you click the **Knit HTML** button a web page will be generated that includes both
content as well as the output of any embedded R code chunks within the document. You can
embed an R code chunk like this:
7
8 ````{r}
9 summary(cars)
10 ...
11
12 You can also embed plots, for example:
13
14 ````{r fig.width=7, fig.height=6}
15 plot(cars)
16 ...
17
18 |
```

Exercise in Climate Data Processing

Before Class (or now):

```
cd ~/ost4sem/exercise/
```

```
git clone
```

```
https://github.com/adammwilson/SpatialAnalysisTutorials.git
```

Everyone:

```
cd ~/ost4sem/exercise/
```

```
git pull
```

```
Open RStudio → /ost4sem/exercise/SpatialAnalysisTutorials/  
climate/code/CDO_Process.Rmd
```