Netcdf: combining spatial and temporal data with metadata

Dasapta Erwin Irawan, R. Willem Vervoort & Gene Melzack

19 January 2018

Storing and sharing spatio-temporal data

- Spatial and temporal data are challenging to store
- It is 2 or 3 dimensional and can even be 4 dimensional (for example groundwater data)
- We could use complex spreadsheets to share data, or build interconnected text files (by site, by time)
  - This is not necessarily easy to share, you would need an additional read me file to describe the data (i.e. it is not self-describing)
  - Is not necessarily system independent (proprietry software)
- Solution: NetCDF formats

NetCDF: history and principles

- NetCDF is the most widely used file format in climate and global studies
  - Almost all Global climate change model data is in NetCDF
  - Specifically good for multi-dimensional arrays
- Key strenghts are that NetCDF files are *self-describing* and *machine independent*
- Libraries and protocols are maintained by Unidata
- The latest protocol is NetCDF4, but older versions NetCDF3 are still around
- The typical file extension used for NetCDF files is **.nc**
- We will use the package `ncdf4` in R to read and create NetCDF4 files

NetCDF: Self-describing

- Self-describing means that within the file the metadata are included
- Here is an example from ET data from NCI Thredds server ET data
- This is output from the AWRA model

```
## File ../exercise/netcdf_data/AWRA05_data.nc (NC_FORMAT_C
##
##      1 variables (excluding dimension variables):
##         float actual_evapotranspiration[lon,lat,time]
##             _FillValue: -999
##             long_name: Daily evapotranspiration, 50th pe
##             cell_methods: day
##             grid_mapping: crs
##             standard_name: actual_evapotranspiration
##             coordinates: time lat lon
##             units: mm day-1
##             _ChunkSizes: 1
```