input4MIPs: Boundary Conditions and Forcing Datasets for CMIP6

DOE ESGF F2F 2016, Marriott Metro Center, Washington D.C.

Paul J. Durack, Karl E. Taylor, Sasha Ames, Denis Nadeau, Tony Hoang and many others..

Wednesday 7th December 2016

PROGRAM FOR CLIMATE MODEL DIAGNOSIS AND INTERCOMPARISON (PCMDI)
pcmdi.llnl.gov/home

U.S. DEPARTMENT OF ENERGY | Office of Science

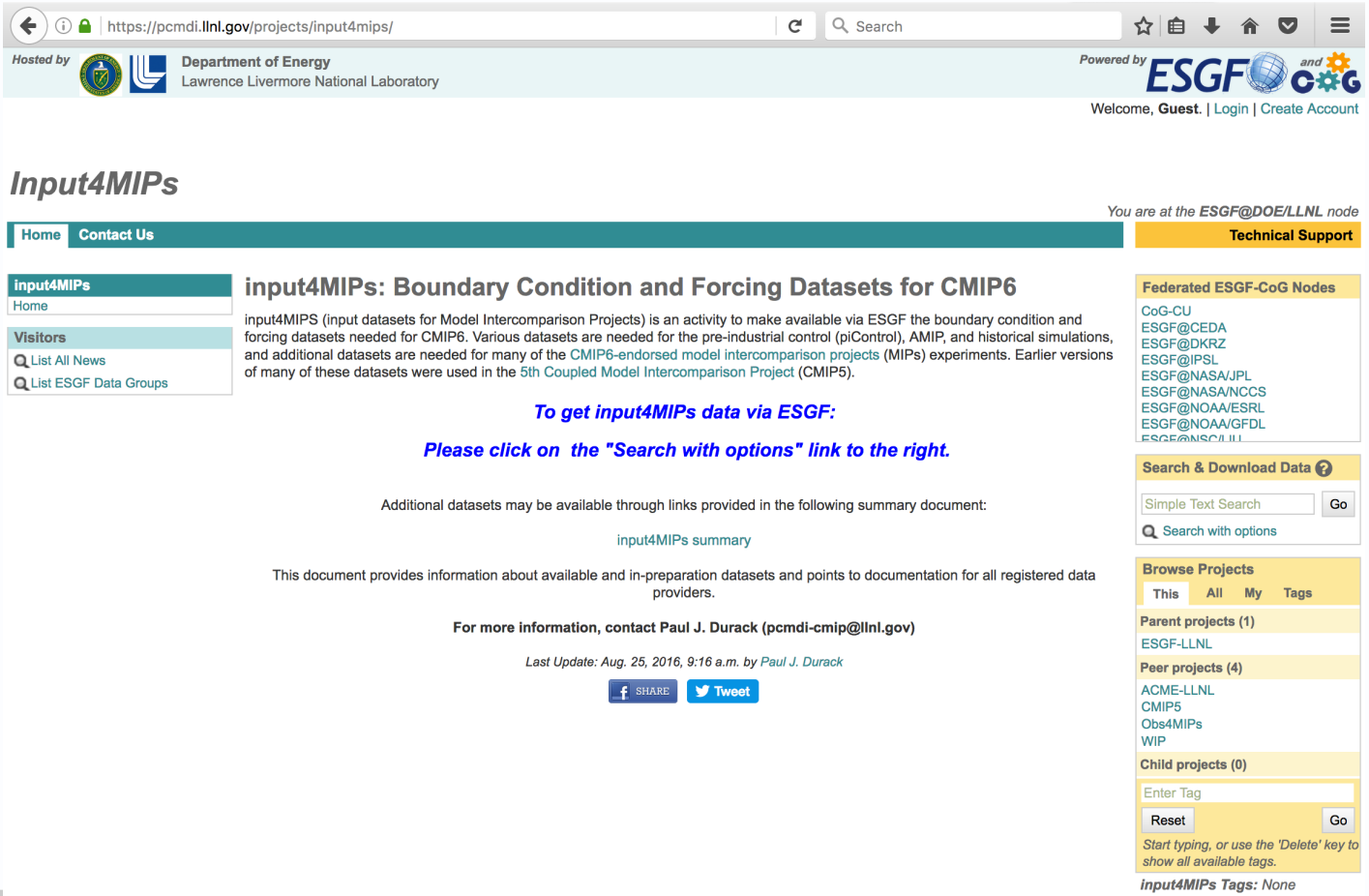
LLNL IM Release#: LLNL-PRES-712497

Section 0:
What is input4MIPs
Observationally-derived input datasets for Model Intercomparison Projects (MIPs)

PCMDI

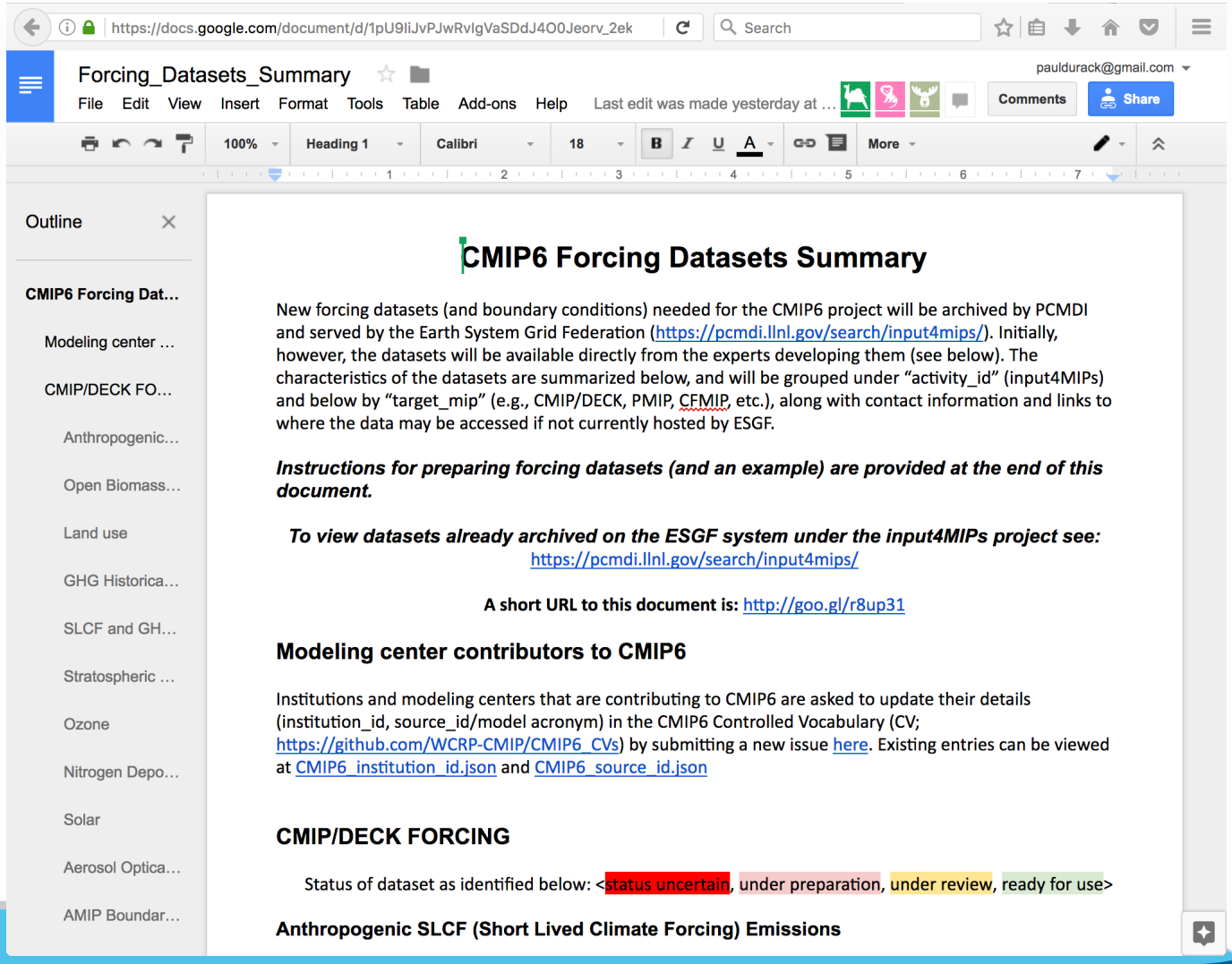
What is input4MIPs:

- All required forcing datasets for CMIP (DECK) and satellite MIP experiments
- User base is likely small ~100 users max?



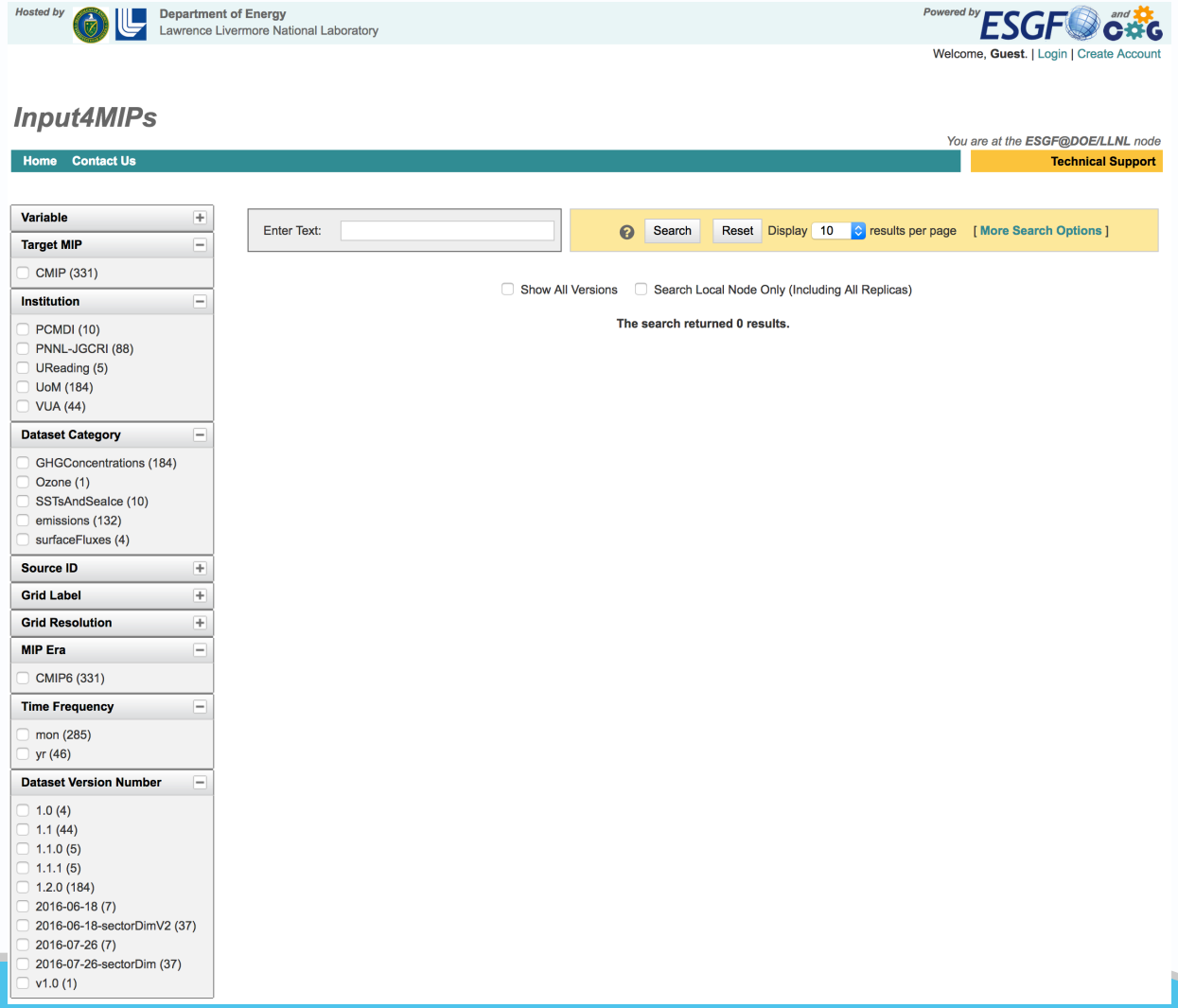
What is input4MIPs:

- **Moving target – using "live" google doc to keep track of changing datasets**



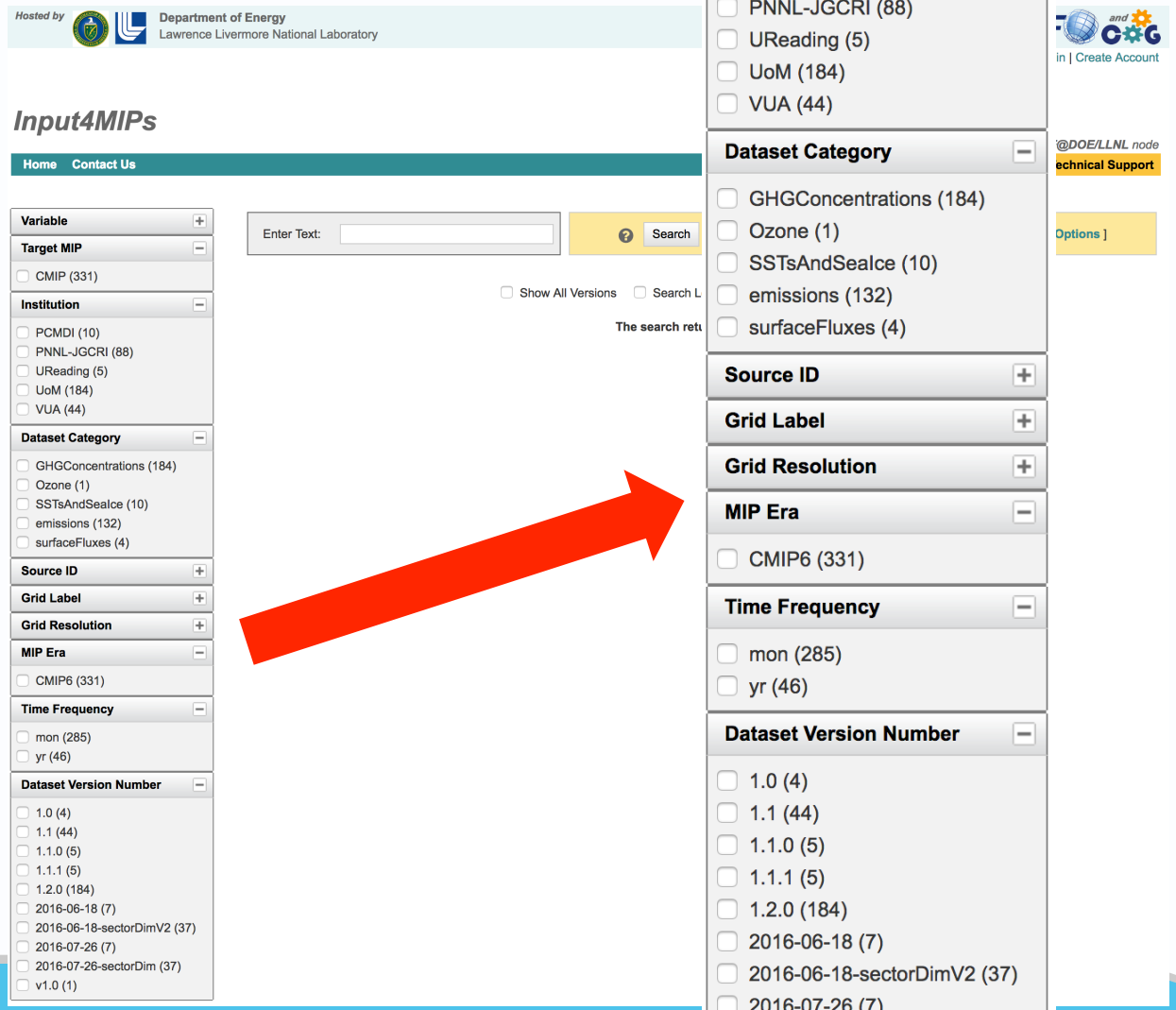
What is input4MIPs:

- **Search facets – have attempted to conform as closely to CMIP6 as possible**



What is input4MIPs:

- **Search facets – have attempted to conform IP6 as possible**



Section 1:
Problem data
Expanding ESGF support for more data formats

PCMDI

Problematic data:

- **Large single files vs temporal chunking (1 x 16GB vs 8 x 2GB time chunks)**
- ~16GB (UMD) – user downloads in 3rd world countries

New! LUH2 Historical datasets now available

LUH2 v2h Release (10/14/16): The updated release of the historical land-use forcing dataset (LUH2 v2h) covers the period 850-2015 and corrects all known issues and notices identified with the previous version (LUH2 v1.0h). This dataset replaces the previously released dataset (LUH2 v1.0h). This product is the result of a series of prototypes released previously, uses the established data format, and will connect smoothly to gridded products for the future.

- Historic Data (850 - 2015 AD)
  - states.nc (5.8 GB)
  - transitions.nc (16 GB)
  - management.nc (1.4 GB)
- Supporting Files
  - staticData_quarterdeg.nc (1 MB)
- Data Documentation
  - LUH2_v2h README

- Will large files be an issue for users behind firewalls?
- What is the experience with download restarts with **wget** (what about rsync?)

Problematic data:

- **Non-CMIP data: Multiple variables per file (UoM) – rewrite if possible**

```
[durack1@oceanonly CMIP6]$ ncdump -h input4MIPs/UoM/GHGConc/CMIP/yr/atmos/UoM-CMIP-1-1-0/GHGCon
c/gr3-GMNHSH/v20160701/mole_fraction_of_carbon_dioxide_in_air_input4MIPs_GHGConcentrations_CMIP
_UoM-CMIP-1-1-0_gr3-GMNHSH_0000-2014.nc
netcdf mole_fraction_of_carbon_dioxide_in_air_input4MIPs_GHGConcentrations_CMIP_UoM-CMIP-1-1-0_
gr3-GMNHSH_0000-2014 {
dimensions:
        lat = 1 ;
        bnds = 2 ;
        time = 2015 ;
variables:
        double lat(lat) ;
                lat:units = "degrees_north" ;
                lat:long_name = "latitude" ;
                lat:standard_name = "latitude" ;
                lat:axis = "Y" ;
                lat:bounds = "lat_bnds" ;
        double lat_bnds(lat, bnds) ;
        double time(time) ;
                time:units = "days since 1850-01-01 00:00:00" ;
                time:calendar = "365_day" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
                time:axis = "T" ;
                time:bounds = "time_bnds" ;
        double time_bnds(time, bnds) ;
        float carbon_dioxide_GM(time) ;
                carbon_dioxide_GM:long_name = "Global Mean Mole Fraction of CO2" ;
                carbon_dioxide_GM:original_name = "CO2_GM" ;
                carbon_dioxide_GM:standard_name = "mole_fraction_of_carbon_dioxide_in_air" ;
                carbon_dioxide_GM:units = "1.e-6" ;
                carbon_dioxide_GM:cell_methods = "time: mean area: mean" ;
                carbon_dioxide_GM:lat = "0.0" ;
                carbon_dioxide_GM:lat_bnds = "-90.0, 90.0" ;
        float carbon_dioxide_NH(time) ;
                carbon_dioxide_NH:long_name = "Northern Hemisphere Mean Mole Fraction of CO2" ;
                carbon_dioxide_NH:original_name = "CO2_NH" ;
                carbon_dioxide_NH:standard_name = "mole_fraction_of_carbon_dioxide_in_air" ;
                carbon_dioxide_NH:units = "1.e-6" ;
                carbon_dioxide_NH:cell_methods = "time: mean area: mean" ;
                carbon_dioxide_NH:lat = "30.0" ;
                carbon_dioxide_NH:lat_bnds = "0.0, 90.0" ;
        float carbon_dioxide_SH(time) ;
                carbon_dioxide_SH:long_name = "Southern Hemisphere Mean Mole Fraction of CO2" ;
                carbon_dioxide_SH:original_name = "CO2_SH" ;
                carbon_dioxide_SH:standard_name = "mole_fraction_of_carbon_dioxide_in_air" ;
                carbon_dioxide_SH:units = "1.e-6" ;
                carbon_dioxide_SH:cell_methods = "time: mean area: mean" ;
                carbon_dioxide_SH:lat = "-30.0" ;
                carbon_dioxide_SH:lat_bnds = "-90.0, 0.0" ;

// global attributes:
                :title = "UoM-CMIP-1-1-0: historical GHG concentrations: global and hemispheric
 means of CO2 prepared for input4MIPs" ;
                :institution_id = "UoM" ;
                :dataset_category = "GHGConcentrations" ;
                :dataset_version_number = "1.1.0" ;
```

PCMDI

Problematic data:

- **Non-CMIP data: Multiple variables per file (UoM) – rewrite if possible**

```
[durack1@oceanonly CMIP6]$ ncdump -h input4MIPs/UoM/GHGConc/CMIP/yr/atmos/UoM-CMIP-1-1-0/GHGCon
c/gr3-GMNHSH/v20160701/mole_fraction_of_carbon_dioxide_in_air_input4MIPs_GHGConcentrations_CMIP
_UoM-CMIP-1-1-0_gr3-GMNHSH_0000-2014.nc
netcdf mole_fraction_of_carbon_dioxide_in_air_input4MIPs_GHGConcentrations_CMIP_UoM-CMIP-1-1-0_
gr3-GMNHSH_0000-2014 {
dimensions:
        lat = 1 ;
        bnds = 2 ;
        time = 2015 ;
variables:
        double lat(lat) ;
                lat:units = "degrees_north" ;
                lat:long_name = "latitude" ;
                lat:standard_name = "latitude" ;
                lat:axis = "Y" ;
                lat:bounds = "lat_bnds" ;
        double lat_bnds(lat, bnds) ;
        double time(time) ;
                time:units = "days since 1850-01-01 00:00:00" ;
                time:calendar = "365_day" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
                time:axis = "T" ;
                time:bounds = "time_bnds" ;
        double time_bnds(time, bnds) ;
        float carbon_dioxide_GM(time) ;
                carbon_dioxide_GM:long_name = "Global Mean Mole Fraction of CO
                carbon_dioxide_GM:original_name = "CO2_GM" ;
                carbon_dioxide_GM:standard_name = "mole_fraction_of_carbon_dio
                carbon_dioxide_GM:units = "1.e-6" ;
                carbon_dioxide_GM:cell_methods = "time: mean area:
                carbon_dioxide_GM:lat = "0.0" ;
                carbon_dioxide_GM:lat_bnds = "-90.0, 90.0" ;
        float carbon_dioxide_NH(time) ;
                carbon_dioxide_NH:long_name = "Northern Hemisphere Mean Mole F
                carbon_dioxide_NH:original_name = "CO2_NH" ;
                carbon_dioxide_NH:standard_name = "mole_fraction_of_carbon_dio
                carbon_dioxide_NH:units = "1.e-6" ;
                carbon_dioxide_NH:cell_methods = "time: mean area: mean" ;
                carbon_dioxide_NH:lat = "30.0" ;
                carbon_dioxide_NH:lat_bnds = "0.0, 90.0" ;
        float carbon_dioxide_SH(time) ;
                carbon_dioxide_SH:long_name = "Southern Hemisphere Mean Mole F
                carbon_dioxide_SH:original_name = "CO2_SH" ;
                carbon_dioxide_SH:standard_name = "mole_fraction_of_carbon_dio
                carbon_dioxide_SH:units = "1.e-6" ;
                carbon_dioxide_SH:cell_methods = "time: mean area: mean" ;
                carbon_dioxide_SH:lat = "-30.0" ;
                carbon_dioxide_SH:lat_bnds = "-90.0, 0.0" ;

// global attributes:
                :title = "UoM-CMIP-1-1-0: historical GHG concentrations: global and hemispheric
 means of CO2 prepared for input4MIPs" ;
                :institution_id = "UoM" ;
                :dataset_category = "GHGConcentrations" ;
                :dataset_version_number = "1.1.0" ;
```

Rewritten file..

```
[durack1@oceanonly CMIP6]$ ncdump -h input4MIPs/UoM/GHGConcentrations/CMIP/yr/atmos/UoM-CMIP-1-
2-0/mole_fraction_of_carbon_dioxide_in_air/gr1-GMNHSH/v20160830/mole_fraction_of_carbon_dioxide
_in_air_input4MIPs_GHGConcentrations_CMIP_UoM-CMIP-1-2-0_gr1-GMNHSH_0000-2014.nc
netcdf mole_fraction_of_carbon_dioxide_in_air_input4MIPs_GHGConcentrations_CMIP_UoM-CMIP-1-2-0_
gr1-GMNHSH_0000-2014 {
dimensions:
        time = UNLIMITED ; // (2015 currently)
        bound = 2 ;
        sector = 3 ;
variables:
        float time(time) ;
                time:bounds = "time_bnds" ;
                time:long_name = "time" ;
                time:standard_name = "time" ;
                time:units = "days since 0-1-1" ;
                time:calendar = "gregorian" ;
                time:axis = "T" ;
        double time_bnds(time, bound) ;
        int sector(sector) ;
                sector:bounds = "sector_bnds" ;
                sector:lat_bnds = "0: -90.0, 90.0; 1: 0.0, 90.0; 2: -90.0, 0.0" ;
                sector:long_name = "sector" ;
                sector:ids = "0: Global; 1: Northern Hemisphere; 2: Southern Hemisphere" ;
                sector:original_names = "0: CO2_GM; 1: CO2_NH; 2: CO2_SH" ;
        double sector_bnds(sector, bound) ;
        float mole_fraction_of_carbon_dioxide_in_air(time, sector) ;
                mole_fraction_of_carbon_dioxide_in_air:_FillValue = 1.e+20f ;
                mole_fraction_of_carbon_dioxide_in_air:missing_value = 1.e+20f ;
                mole_fraction_of_carbon_dioxide_in_air:long_name = "mole" ;
                mole_fraction_of_carbon_dioxide_in_air:cell_methods = "time: mean area: mean" ;
                mole_fraction_of_carbon_dioxide_in_air:units = "1.e-6" ;

// global attributes:
                :Conventions = "CF-1.6" ;
                :comment = "Data provided are global and hemispheric area-weighted means. Zonal
 means for 15-degree lat bands or 0.5-degree lat bands are available in gn-15x360 or gr-0p5x360
 files respectively" ;
                :variable_id = "mole_fraction_of_carbon_dioxide_in_air" ;
```

Problematic data:

- **Non-CF compliant (non-gridded) netcdf (RFMIP)**
- **Multiple vars limits ESGF functionality, no openDAP..**

```
[durack1ml:16/150128_CMIP6/RFMIP] durack1% ncdump -h 161122_RobertPincus_multiple_input4MIPs_radiation_RFMIP_UColorado-RFMIP-20161122_none.nc | more
netcdf \161122_RobertPincus_multiple_input4MIPs_radiation_RFMIP_UColorado-RFMIP-20161122_none {
dimensions:
        expt = 18 ;
        level = 61 ;
        layer = 60 ;
        site = 100 ;
variables:
        float lon(site) ;
                lon:long_name = "ERA-Interim longitude" ;
                lon:units = "degree_north" ;
                lon:standard_name = "longitude" ;
        float lat(site) ;
                lat:long_name = "ERA-Interim latitude" ;
                lat:units = "degree_east" ;
                lat:standard_name = "latitude" ;
        float time(site) ;
                time:long_name = "ERA-Interim fractional day of the year 2014" ;
                time:units = "days since 2014-1-1 0:0:0" ;
                time:standard_name = "time" ;
                time:calendar = "gregorian" ;
        float sst(site) ;
                sst:title = "sea surface temperature" ;
                sst:units = "K" ;
                sst:long_name = "ERA-Interim sea surface temperature (= \"missing_value\" over land)" ;
                sst:standard_name = "sea_surface_temperature" ;
                sst:missing_value = -9.99f ;
                sst:FillValue = -9.99f ;
                sst:coordinates = "lon lat time" ;
        string expt_label(expt) ;
                expt_label:long_name = "experiment description" ;
        float pres_layer(site, layer) ;
```

```
[durack1ml:16/150128_CMIP6/RFMIP] durack1% ncdump -h 161122_RobertPincus_multiple_input4MIPs_radiation_RFMIP_UColorado-RFMIP-20161122_none.nc | grep float
        float lon(site) ;
        float lat(site) ;
        float time(site) ;
        float sst(site) ;
        float pres_layer(site, layer) ;
        float pres_level(site, level) ;
        float surface_emissivity(site) ;
        float surface_albedo(site) ;
        float solar_zenith_angle(site) ;
        float total_solar_irradiance(site) ;
        float profile_weight(site) ;
        float oxygen_GM(expt) ;
        float nitrogen_GM(expt) ;
        float temp_layer(expt, site, layer) ;
        float temp_level(expt, site, level) ;
        float surface_temperature(expt, site) ;
        float water_vapor(expt, site, layer) ;
        float ozone(expt, site, layer) ;
        float carbon_monoxide_GM(expt) ;
        float c2f6_GM(expt) ;
        float c3f8_GM(expt) ;
        float c4f10_GM(expt) ;
        float c5f12_GM(expt) ;
        float c6f14_GM(expt) ;
        float c7f16_GM(expt) ;
        float c8f18_GM(expt) ;
        float c_c4f8_GM(expt) ;
        float carbon_dioxide_GM(expt) ;
        float carbon_tetrachloride_GM(expt) ;
        float cf4_GM(expt) ;
        float cfc113_GM(expt) ;
        float cfc114_GM(expt) ;
        float cfc115_GM(expt) ;
        float cfc11_GM(expt) ;
        float cfc11eq_GM(expt) ;
        float cfc12_GM(expt) ;
        float cfc12eq_GM(expt) ;
        float ch2cl2_GM(expt) ;
        float ch3ccl3_GM(expt) ;
        float chcl3_GM(expt) ;
        float halon1211_GM(expt) ;
        float halon1301_GM(expt) ;
        float halon2402_GM(expt) ;
        float hcfc141b_GM(expt) ;
        float hcfc142b_GM(expt) ;
        float hcfc22_GM(expt) ;
        float hfc125_GM(expt) ;
        float hfc134a_GM(expt) ;
        float hfc134aeq_GM(expt) ;
        float hfc143a_GM(expt) ;
        float hfc152a_GM(expt) ;
        float hfc227ea_GM(expt) ;
        float hfc236fa_GM(expt) ;
        float hfc23_GM(expt) ;
        float hfc245fa_GM(expt) ;
        float hfc32_GM(expt) ;
        float hfc365mfc_GM(expt) ;
        float hfc4310mee_GM(expt) ;
        float methane_GM(expt) ;
        float methyl_bromide_GM(expt) ;
        float methyl_chloride_GM(expt) ;
        float nf3_GM(expt) ;
        float nitrous_oxide_GM(expt) ;
        float sf6_GM(expt) ;
        float so2f2_GM(expt) ;
[durack1ml:16/150128_CMIP6/RFMIP] durack1%
```


CMDI

Problematic data:

- Non-CF compliant (non-gridded) netcdf (RFMIP)
- Multiple vars limits ESGF functionality, no openDAP..
- Publisher breaking.. Will be fixed by CDMS2

```
[durack1ml:16/150128_CMIP6/RFM
FMIP_UColorado-RFMIP-20161122_
netcdf \161122_RobertPincus_mu
dimensions:
        expt = 18 ;
        level = 61 ;
        layer = 60 ;
        site = 100 ;
variables:
        float lon(site) ;
                lon:long_name
                lon:units = "d
                lon:standard_n
        float lat(site) ;
                lat:long_name
                lat:units = "d
                lat:standard_n
        float time(site) ;
                time:long_name
                time:units = "
                time:standard_
                time:calendar
        float sst(site) ;
                sst:title = "s
                sst:units = "K
                sst:long_name
                sst:standard_n
                sst:missing_va
                sst:FillValue
                sst:coordinate
        string expt_label(expt
                expt_label:lon
        float pres_layer(site,
```

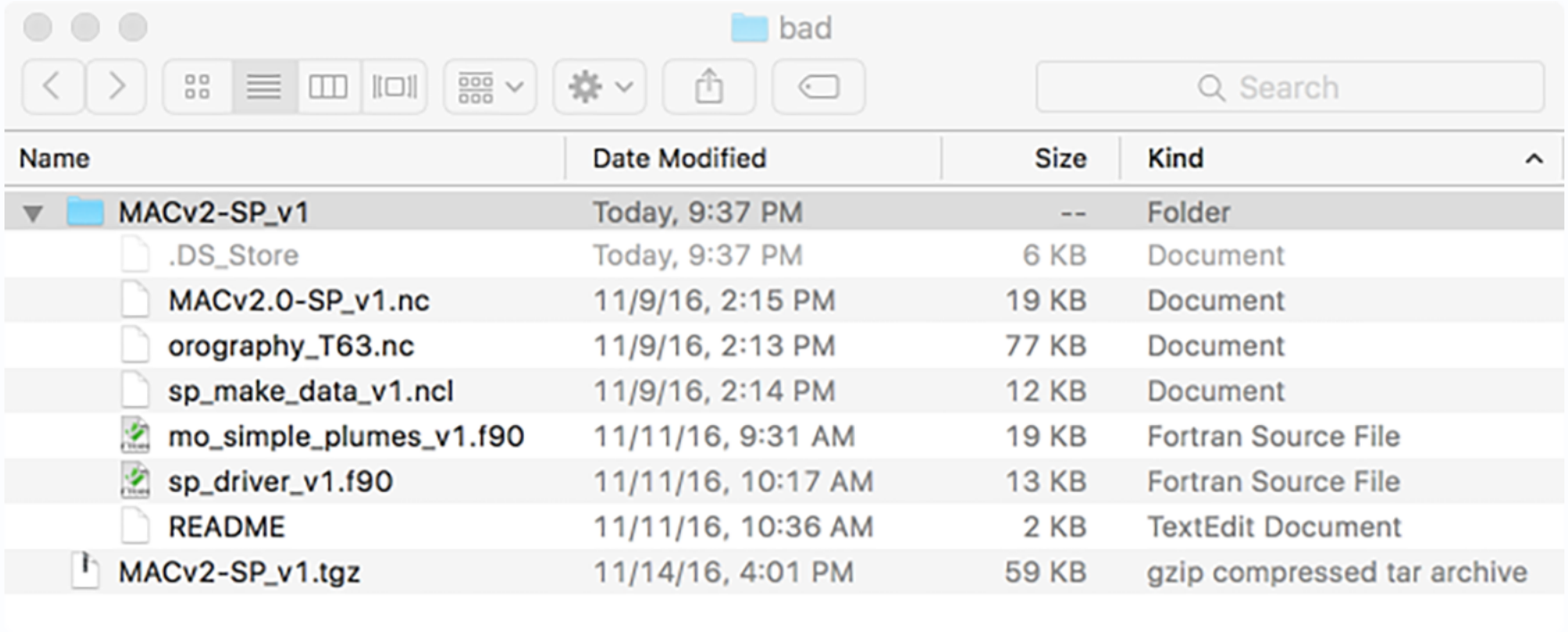
```
[root@esg-idx2 ~]# esgpublish --project input4MIPs --map test.map
INFO       2016-11-24 23:06:43,569 Creating dataset: input4MIPs.RFMIP2….
INFO       2016-11-24 23:06:43,571 Scanning /esg/data/test/
multiple_input4MIPs_radiation_RFMIP_UColorado-RFMIP-20161122_none.nc
Traceback (most recent call last):
  File "/usr/local/uvcdat/2.2.0/bin/esgpublish", line 4, in <module>
    __import__('pkg_resources').run_script('esgcet==2.14.6', 'esgpublish')
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/
setuptools-19.1.1-py2.7.egg/pkg_resources/__init__.py", line 745, in
run_script
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/
setuptools-19.1.1-py2.7.egg/pkg_resources/__init__.py", line 1670, in
run_script
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgcet-2.14.6-
py2.7.egg/EGG-INFO/scripts/esgpublish", line 526, in <module>
…….
    main(sys.argv[1:])
handler, cfHandler, configOptions,
aggregateDimensionName=aggregateDimensionName, offline=offline,
progressCallback=progressCallback, stopEvent=stopEvent,
extraFields=extraFields, masterGateway=masterGateway,
useVersion=useVersion, **context)
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgcet-2.14.6-
py2.7.egg/esgcet/publish/extract.py", line 307, in createDataset
    extractFromFile(dset, f, file, session, cfHandler,
aggdimName=aggregateDimensionName, varlocate=varlocate, **context)
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgcet-2.14.6-
py2.7.egg/esgcet/publish/extract.py", line 656, in extractFromFile
    var0 = openfile.getVariable(varname, index=0)
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgcet-2.14.6-
py2.7.egg/esgcet/config/netcdf_handler.py", line 141, in getVariable
    result = variable[index]
ValueError: data type must provide an itemsize
```

```
[durack1ml:16/150128_CMIP6/RFMIP] durack1% ncdump
 -h 161122_RobertPincus_multiple_input4MIPs_radia
tion_RFMIP_UColorado-RFMIP-20161122_none.nc | gre
p float
        float lon(site) ;
        float lat(site) ;
        float time(site) ;
        float sst(site) ;
        float pres_layer(site, layer) ;
        float pres_level(site, level) ;
        float surface_emissivity(site) ;
        float surface_albedo(site) ;
        float solar_zenith_angle(site) ;
        float total_solar_irradiance(site) ;
        float profile_weight(site) ;
        float oxygen_GM(expt) ;
        float nitrogen_GM(expt) ;
                                    p_layer(expt, site, layer) ;
                                    p_level(expt, site, level) ;
                                    face_temperature(expt, site) ;
                                    er_vapor(expt, site, layer) ;
                                    ne(expt, site, layer) ;
                                    bon_monoxide_GM(expt) ;
                                    6_GM(expt) ;
                                    8_GM(expt) ;
                                    10_GM(expt) ;
                                    12_GM(expt) ;
                                    14_GM(expt) ;
                                    16_GM(expt) ;
                                    18_GM(expt) ;
                                    4f8_GM(expt) ;
                                    bon_dioxide_GM(expt) ;
                                    bon_tetrachloride_GM(expt) ;
                                    _GM(expt) ;
                                    113_GM(expt) ;
                                    114_GM(expt) ;
                                    115_GM(expt) ;
                                    11_GM(expt) ;
                                    11eq_GM(expt) ;
                                    12_GM(expt) ;
                                    12eq_GM(expt) ;
                                    cl2_GM(expt) ;
                                    ccl3_GM(expt) ;
                                    l3_GM(expt) ;
                                    on1211_GM(expt) ;
                                    on1301_GM(expt) ;
                                    on2402_GM(expt) ;
                                    c141b_GM(expt) ;
                                    c142b_GM(expt) ;
                                    c22_GM(expt) ;
                                    125_GM(expt) ;
                                    134a_GM(expt) ;
                                    134aeq_GM(expt) ;
                                    143a_GM(expt) ;
                                    152a_GM(expt) ;
                                    227ea_GM(expt) ;
                                    236fa_GM(expt) ;
                                    23_GM(expt) ;
                                    245fa_GM(expt) ;
                                    32_GM(expt) ;
                                    365mfc_GM(expt) ;
                                    4310mee_GM(expt) ;
                                    hane_GM(expt) ;
                                    hyl_bromide_GM(expt) ;
                                    hyl_chloride_GM(expt) ;
                                    _GM(expt) ;
                                    rous_oxide_GM(expt) ;
        float sf6_GM(expt) ;
        float so2f2_GM(expt) ;
[durack1ml:16/150128_CMIP6/RFMIP] durack1%
```


CMDI

Problematic data:

- **Non-netcdf – tar gzipped, code snippets, text files (MPI)**



Section 2:
Adapting DRS
Making documentation, code and ancillary data a first class ESGF citizen

PCMDI

Adapting DRS for documentation/code:

Work toward allowing “documentation” to be published alongside the netcdf data, for this we would be using a standard directory structure as below (for the RFMIP and UMD data we plan to host and the PCMDI data we currently already have):

Datasets of “standard/CMIP” one variable per file format:

CMIP6 = <mip_era>/<activity_id>/<institution_id>/<source_id>/<experiment_id>/<member_id>/<table_id>/<variable_id>/<grid_label>/<version>

input4MIPs = CMIP6/input4MIPs/<institution_id>/<dataset_category>/<target_mip>/<frequency>/<realm>/<source_id>/<variable_id>/<grid_label>/<version>/files.nc

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/areacello/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sic/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sicbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tos/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tosbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/**documentation**/**none**/v20161020/*.pdf or *.txt or *.docx etc

Adapting DRS for documentation/code:

Work toward allowing “documentation” to be published alongside the netcdf data, for this we would be using a standard directory structure as below (for the RFMIP and UMD data we plan to host and the PCMDI data we currently already have):

Datasets of “standard/CMIP” one variable per file format:

CMIP6 = <mip_era>/<activity_id>/<institution_id>/<source_id>/<experiment_id>/<member_id>/<table_id>/<variable_id>/<grid_label>/<version>

input4MIPs = CMIP6/input4MIPs/<institution_id>/<dataset_category>/<target_mip>/<frequency>/<realm>/<source_id>/<variable_id>/<grid_label>/<version>/files.nc

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/areacello/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sic/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sicbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tos/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tosbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/**documentation**/**none**/v20161020/*.pdf or *.txt or *.docx etc

Dataset with multiple variables in single files (not gridded):

CMIP6/input4MIPs/UColorado/RFMIP/RFMIP/invariant/atmos/UColorado-RFMIP-0-2/**multiple**/**none**/v20161101/RFMIP-IRF-Inputs.nc

CMIP6/input4MIPs/UColorado/RFMIP/RFMIP/invariant/atmos/UColorado-RFMIP-0-2/**documentation**/**none**/v20161101/RFMIP.pdf or *.txt or *.docx etc

Adapting DRS for documentation/code:

Work toward allowing "documentation" to be published alongside the netcdf data, for this we would be using a standard directory structure as below (for the RFMIP and UMD data we plan to host and the PCMDI data we currently already have):

Datasets of "standard/CMIP" one variable per file format:

CMIP6 = <mip_era>/<activity_id>/<institution_id>/<source_id>/<experiment_id>/<member_id>/<table_id>/<variable_id>/<grid_label>/<version>

input4MIPs = CMIP6/input4MIPs/<institution_id>/<dataset_category>/<target_mip>/<frequency>/<realm>/<source_id>/<variable_id>/<grid_label>/<version>/files.nc

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/areacello/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sic/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/sicbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tos/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/tosbcs/gn/v20161020/

CMIP6/input4MIPs/PCMDI/SSTsAndSeaIce/CMIP/mon/ocean/PCMDI-AMIP-1-1-1/**documentation**/**none**/v20161020/*.pdf or *.txt or *.docx etc

Dataset with multiple variables in single files (not gridded):

CMIP6/input4MIPs/UColorado/RFMIP/RFMIP/invariant/atmos/UColorado-RFMIP-0-2/**multiple**/**none**/v20161101/RFMIP-IRF-Inputs.nc

CMIP6/input4MIPs/UColorado/RFMIP/RFMIP/invariant/atmos/UColorado-RFMIP-0-2/**documentation**/**none**/v20161101/RFMIP.pdf or *.txt or *.docx etc

Dataset with multiple variables in multiple files (gridded):

CMIP6/input4MIPs/UMD/LandUse/CMIP/mon/land/UMD-2-0/**multiple**/**gn**/v20161101/*a.nc, *b.nc, c.nc

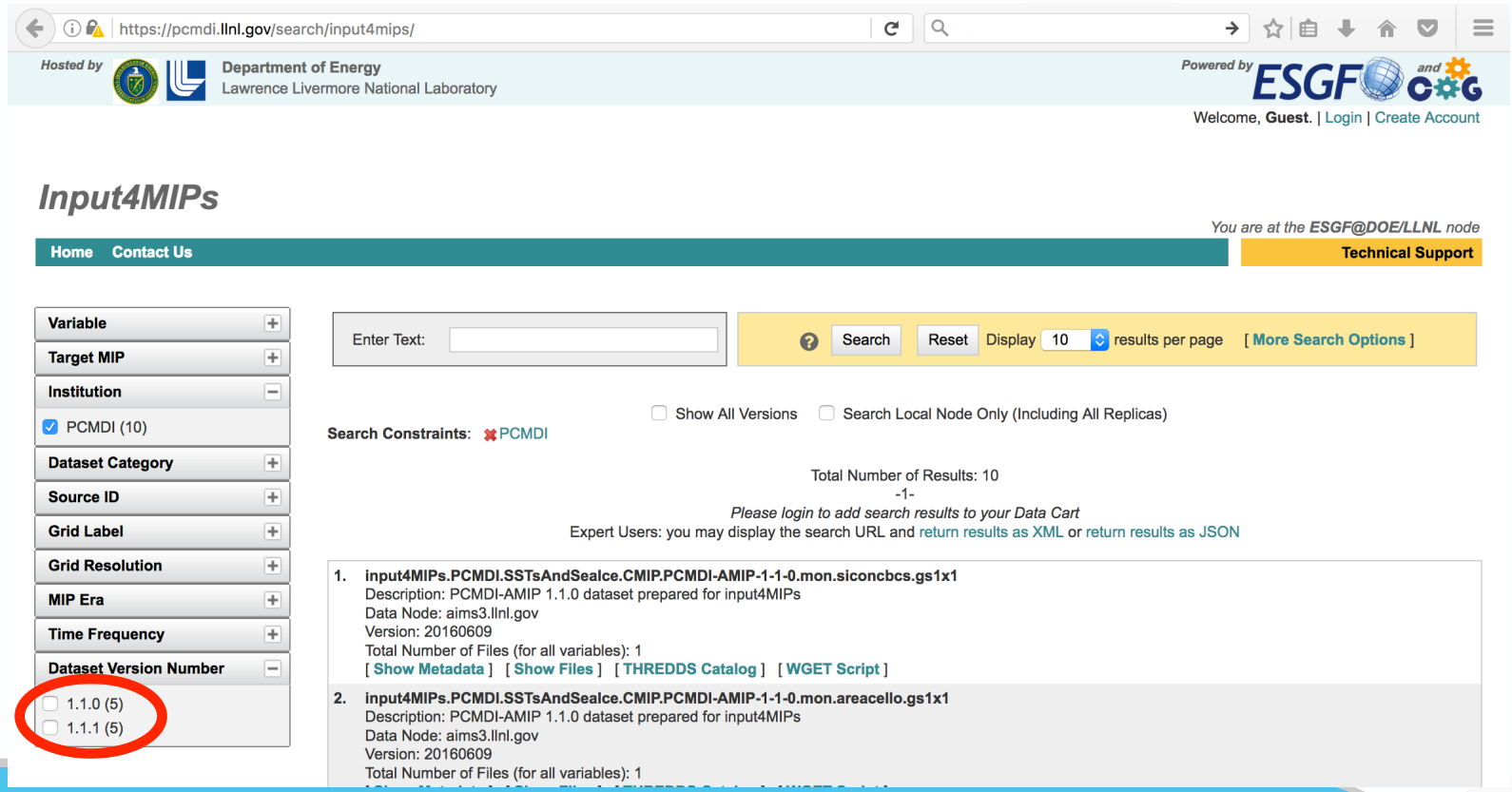
CMIP6/input4MIPs/UMD/LandUse/CMIP/mon/land/UMD-2-0/**documentation**/**none**/v20161101/*.pdf or *.txt or *.docx etc

Expanding DRS would be a great new addition, and useable across obs4MIPs data (where there will likely also be other data formats in addition to pdf documentation).

Section 3:
Deprecating datasets
How to gracefully “hide” obsolete datasets, but leave them available for trace-ability

Deprecating datasets (gracefully):

- "Hiding" datasets that are not the latest would be useful
- Useful for datasets that are not problematic, but are deprecated (rather than unpublishing)
- PCMDI AMIP dataset is updated every 6 months – `1.1.0` vs `1.1.1` vs `1.1.2`..



Section 4:
Breaking SYNDA
Multiple variable files do not conform to CMIP-standards and..

PCMDI

Breaking SYNDA:

- **SYNDA has hit problems** downloading input4MIPs data with more than a single variable per file
- **Contributed data passed the CF-check**, but isn't CMIP "format"
- As **input4MIPs data isn't derived from models**, the **DRS structure differs** from the CMIP5 data
- Many of the "obs" **products contributed have interacting variables – to use the data you need more than a single variable in memory**
- The path forward is not clear..

Thank you

Paul J. Durack
t +1 925 422 5208
e pauldurack@llnl.gov
w www-pcmdi.llnl.gov

PCMDI