



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)

What is input4MIPs:

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

The screenshot shows the input4MIPs website. The browser address bar displays <https://pcmdi.llnl.gov/projects/input4mips/>. The page header includes logos for the Department of Energy, Lawrence Livermore National Laboratory, and ESGF, along with a 'Welcome, Guest' message and links for 'Login' and 'Create Account'. The main heading is 'Input4MIPs'. A navigation bar contains 'Home' and 'Contact Us'. On the left, a sidebar lists 'input4MIPs Home', 'Visitors' (with links to 'List All News' and 'List ESGF Data Groups'), and 'Technical Support'. The main content area is titled 'input4MIPs: Boundary Condition and Forcing Datasets for CMIP6'. It explains that input4MIPs provides boundary condition and forcing datasets for CMIP6, including pre-industrial control (piControl), AMIP, and historical simulations. It notes that earlier versions of these datasets were used in the 5th Coupled Model Intercomparison Project (CMIP5). A blue link states 'To get input4MIPs data via ESGF: Please click on the "Search with options" link to the right.' Below this, it mentions that additional datasets may be available through links in a summary document, with a link to 'input4MIPs summary'. It further states that this document provides information about available and in-preparation datasets and points to documentation for all registered data providers. For more information, it directs users to contact Paul J. Durack at pcmdi-cmip@llnl.gov. A note indicates the last update was on Aug. 25, 2016, at 9:16 a.m. by Paul J. Durack. There are 'SHARE' and 'Tweet' buttons. On the right, a 'Federated ESGF-CoG Nodes' list includes CoG-CU, ESGF@CEDA, ESGF@DKRZ, ESGF@IPSL, ESGF@NASA/JPL, ESGF@NASA/NCCS, ESGF@NOAA/ESRL, and ESGF@NOAA/GFDL. Below this is a 'Search & Download Data' section with a 'Simple Text Search' box and a 'Go' button, and a 'Search with options' link. A 'Browse Projects' section shows 'Parent projects (1)' (ESGF-LLNL), 'Peer projects (4)' (ACME-LLNL, CMIP5, Obs4MIPs, WIP), and 'Child projects (0)'. There is also a 'Tags' section with a search box and a 'Go' button. The footer contains information about ESGF sponsors and partners (DoE Office of Science, IS-ENES, NASA, NOAA, NCI, NSF), the CoG version (3.7.0), the ESGF P2P Version (v2.3.8 master), Earth System CoG sponsors and partners (NOAA, NASA, NSF, DoE Office of Science, IS-ENES), the URL <https://pcmdi.llnl.gov>, a 'Privacy & Legal Notice' link, and the PCMDI logo.

Hosted by Department of Energy
Lawrence Livermore National Laboratory

Powered by and ESGF and CoG
Welcome, Guest. | Login | Create Account

Input4MIPs

You are at the [ESGF@DOE/LLNL](#) node

[Home](#) [Contact Us](#) [Technical Support](#)

input4MIPs: Boundary Condition and Forcing Datasets for CMIP6

input4MIPs (input datasets for Model Intercomparison Projects) is an activity to make available via ESGF the boundary condition and forcing datasets needed for CMIP6. Various datasets are needed for the pre-industrial control (piControl), AMIP, and historical simulations, and additional datasets are needed for many of the CMIP6-endorsed model intercomparison projects (MIPs) experiments. Earlier versions of many of these datasets were used in the 5th Coupled Model Intercomparison Project (CMIP5).

To get input4MIPs data via ESGF:
Please click on the "Search with options" link to the right.

Additional datasets may be available through links provided in the following summary document:
[input4MIPs summary](#)

This document provides information about available and in-preparation datasets and points to documentation for all registered data providers.

For more information, contact Paul J. Durack (pcmdi-cmip@llnl.gov)

Last Update: Aug. 25, 2016, 9:16 a.m. by Paul J. Durack

[SHARE](#) [Tweet](#)

Federated ESGF-CoG Nodes

- CoG-CU
- ESGF@CEDA
- ESGF@DKRZ
- ESGF@IPSL
- ESGF@NASA/JPL
- ESGF@NASA/NCCS
- ESGF@NOAA/ESRL
- ESGF@NOAA/GFDL
- ESGF@NSC/ILL

Search & Download Data ?

Simple Text Search

[Search with options](#)

Browse Projects

[This](#) [All](#) [My](#) [Tags](#)

Parent projects (1)

- ESGF-LLNL

Peer projects (4)

- ACME-LLNL
- CMIP5
- Obs4MIPs
- WIP

Child projects (0)

Enter Tag

Start typing, or use the 'Delete' key to show all available tags.

Input4MIPs Tags: None

ESGF sponsors and partners
DoE Office of Science | IS-ENES | NASA | NOAA | NCI | NSF

CoG version 3.7.0
ESGF P2P Version v2.3.8 master

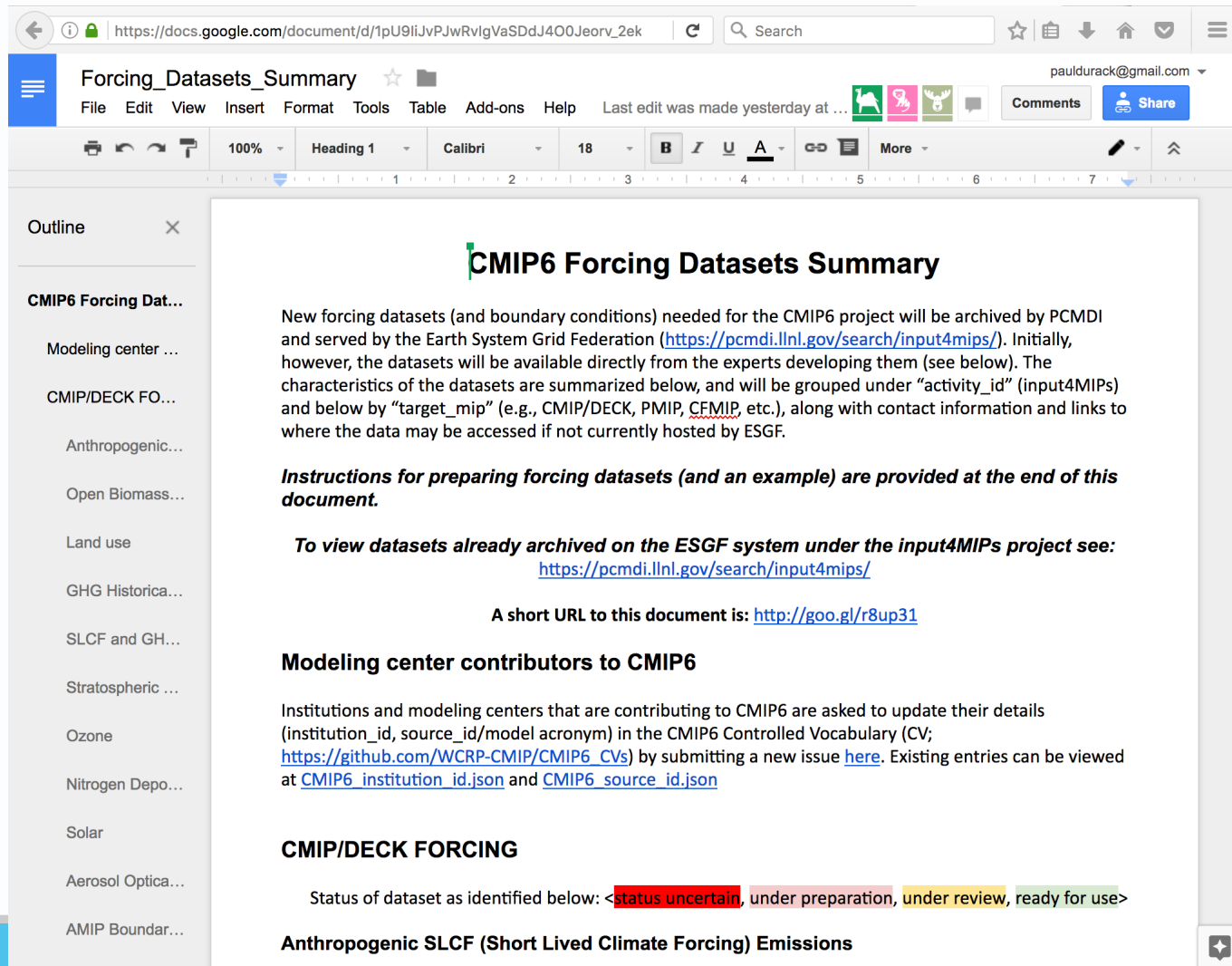
Earth System CoG sponsors and partners
NOAA | NASA | NSF | DoE Office of Science | IS-ENES

<https://pcmdi.llnl.gov> [Privacy & Legal Notice](#)

input4MIPs |

What is input4MIPs:


- Moving target – using “live” google doc to keep track of changing datasets





The screenshot shows a Google Docs interface for a document titled "Forcing_Datasets_Summary". The document is edited by pauldurack@gmail.com. The main content area displays the title "CMIP6 Forcing Datasets Summary" in a large, bold font. Below the title, the text explains that new forcing datasets and boundary conditions for the CMIP6 project will be archived by PCMDI and served by the Earth System Grid Federation. It mentions that the datasets will be available directly from the experts developing them and will be grouped under "activity_id" (input4MIPs) and below by "target_mip" (e.g., CMIP/DECK, PMIP, CFMIP, etc.). It also notes that contact information and links to where the data may be accessed if not currently hosted by ESGF are provided. A section titled "Instructions for preparing forcing datasets (and an example) are provided at the end of this document." is followed by a link to view datasets already archived on the ESGF system under the input4MIPs project: <https://pcmdi.llnl.gov/search/input4mips/>. A short URL to the document is provided: <http://goo.gl/r8up31>. The document then lists "Modeling center contributors to CMIP6" and mentions that institutions and modeling centers contributing to CMIP6 are asked to update their details in the CMIP6 Controlled Vocabulary (CV) by submitting a new issue. Existing entries can be viewed at [CMIP6 institution_id.json](#) and [CMIP6 source_id.json](#). The next section is "CMIP/DECK FORCING", which includes a status of dataset as identified below: <status uncertain, under preparation, under review, ready for use>. The final section is "Anthropogenic SLCF (Short Lived Climate Forcing) Emissions". The left sidebar shows an outline of the document with sections like "CMIP6 Forcing Dat...", "Modeling center ...", "CMIP/DECK FO...", "Anthropogenic...", "Open Biomass...", "Land use", "GHG Historica...", "SLCF and GH...", "Stratospheric ...", "Ozone", "Nitrogen Depo...", "Solar", "Aerosol Optica...", and "AMIP Boundar...".

What is input4MIPs:

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


Hosted by  Department of Energy
Lawrence Livermore National Laboratory


Powered by  and 
Welcome, Guest. | Login | Create Account

Input4MIPs


You are at the [ESGF@DOE/LLNL node](#)

[Home](#) [Contact Us](#) [Technical Support](#)

Variable 

Target MIP 

☐ CMIP (331)

Institution 


☐ PCMDI (10)

☐ PNNL-JGCRI (88)

☐ UReading (5)

☐ UoM (184)

☐ VUA (44)

Dataset Category 


☐ GHGConcentrations (184)


☐ Ozone (1)


☐ SSTsAndSealce (10)


☐ emissions (132)

☐ surfaceFluxes (4)


Source ID 

Grid Label 

Grid Resolution 


MIP Era 

☐ CMIP6 (331)

Time Frequency 

☐ mon (285)

☐ yr (46)

Dataset Version Number 

☐ 1.0 (4)

☐ 1.1 (44)

☐ 1.1.0 (5)

☐ 1.1.1 (5)

☐ 1.2.0 (184)

☐ 2016-06-18 (7)


☐ 2016-06-18-sectorDimV2 (37)

☐ 2016-07-26 (7)

☐ 2016-07-26-sectorDim (37)

☐ v1.0 (1)

Enter Text:

 [Search](#) [Reset](#) Display results per page [\[More Search Options \]](#)


☐ Show All Versions ☐ Search Local Node Only (Including All Replicas)

The search returned 0 results.

What is input4MIPs:

- Search facets – have attempted to conform

IP6 as possible

Hosted by  Department of Energy
Lawrence Livermore National Laboratory

Input4MIPs

[Home](#) [Contact Us](#)

Enter Text: [?](#) [Search](#)

☐ Show All Versions ☐ Search L

The search results

Variable [+](#)

Target MIP [-](#)

☐ CMIP (331)

Institution [-](#)

☐ PCMDI (10)

☐ PNNL-JGCRI (88)

☐ UReading (5)

☐ UoM (184)

☐ VUA (44)

Dataset Category [-](#)

☐ GHGConcentrations (184)

☐ Ozone (1)

☐ SSTsAndSealce (10)

☐ emissions (132)

☐ surfaceFluxes (4)

Source ID [+](#)

Grid Label [+](#)

Grid Resolution [+](#)

MIP Era [-](#)

☐ CMIP6 (331)

Time Frequency [-](#)

☐ mon (285)

☐ yr (46)

Dataset Version Number [-](#)

☐ 1.0 (4)

☐ 1.1 (44)

☐ 1.1.0 (5)

☐ 1.1.1 (5)

☐ 1.2.0 (184)

☐ 2016-06-18 (7)

☐ 2016-06-18-sectorDimV2 (37)

☐ 2016-07-26 (7)

☐ 2016-07-26-sectorDim (37)

☐ v1.0 (1)

Variable [+](#)

Target MIP [-](#)

☐ CMIP (331)

Institution [-](#)

☐ PCMDI (10)

☐ PNNL-JGCRI (88)

☐ UReading (5)

☐ UoM (184)

☐ VUA (44)

Dataset Category [-](#)

☐ GHGConcentrations (184)

☐ Ozone (1)

☐ SSTsAndSealce (10)

☐ emissions (132)

☐ surfaceFluxes (4)

Source ID [+](#)

Grid Label [+](#)

Grid Resolution [+](#)

MIP Era [-](#)

☐ CMIP6 (331)

Time Frequency [-](#)

☐ mon (285)

☐ yr (46)

Dataset Version Number [-](#)

☐ 1.0 (4)

☐ 1.1 (44)

☐ 1.1.0 (5)

☐ 1.1.1 (5)

☐ 1.2.0 (184)

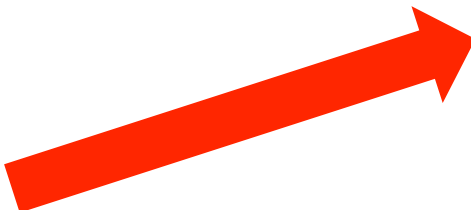
☐ 2016-06-18 (7)


☐ 2016-06-18-sectorDimV2 (37)

☐ 2016-07-26 (7)

☐ 2016-07-26-sectorDim (37)

☐ v1.0 (1)




in | [Create Account](#)

[@DOE/LLNL node](#)
[Technical Support](#)

[Options \]](#)

Section 1: Problem data

Expanding ESGF support for more data formats

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/GHGConc/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" ;
```

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/GHGConc/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;" ;
        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" ;
```

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 | 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) ;
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 hfc134a_gm_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) ;
float so2f2_GM(expt) ;
```

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/RFMIP] FMIP_UColorado-RFMIP-20161122_
netcdf \161122_RobertPincus_mu
dimensions:
```

```
expt = 18 ;
level = 61 ;
layer = 60 ;
site = 100 ;
```

variables:

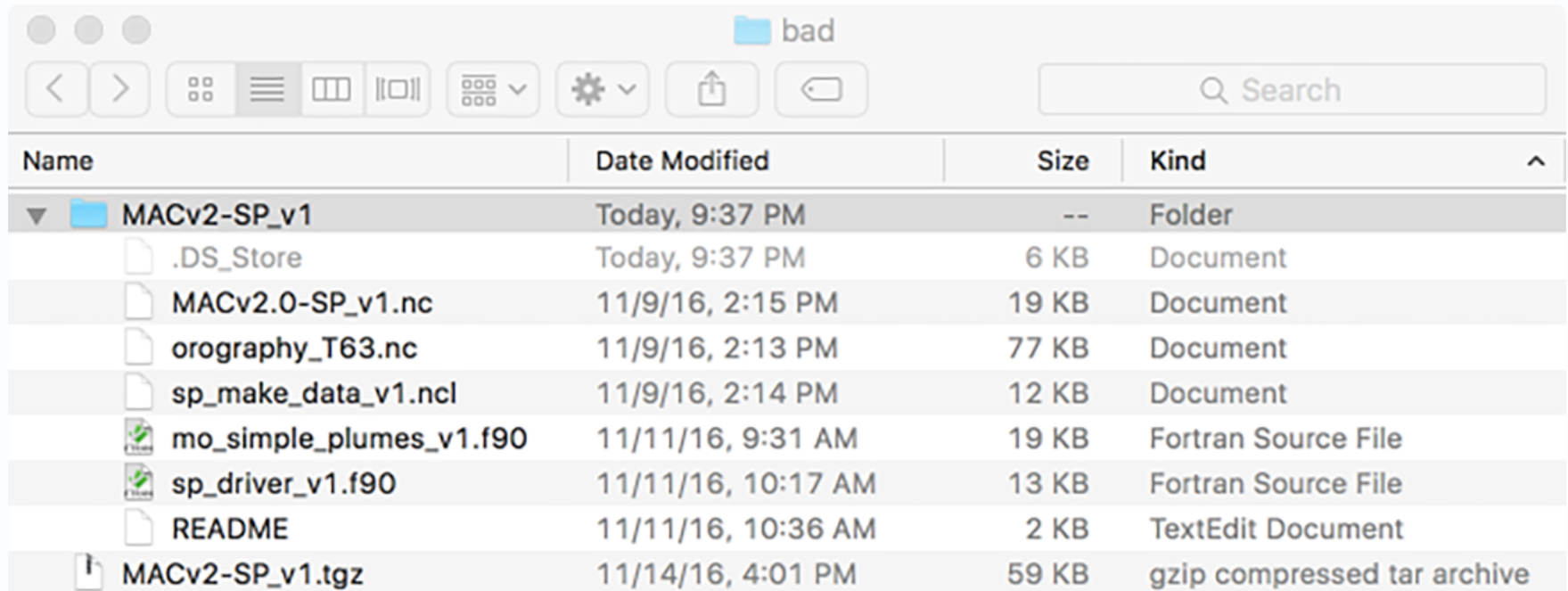
```
float lon(site) ;
lon:long_name
lon:units = "
lon:standard_n
float lat(site) ;
lat:long_name
lat:units = "d
lat:standard_n
float time(site) ;
time:long_name
time:units = "
time:standard
time:calendar
float sst(site) ;
sst:title = "
sst:units = "K
sst:long_name
sst:standard_n
sst:missing_va
sst:FillValue
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('esgset==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/esgset-2.14.6-
py2.7.egg/EGG-INF0/scripts/esgpublish", line 526, in <module>
    main(sys.argv[1:])
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgset-2.14.6-
py2.7.egg/esgset/publish/extract.py", line 307, in createDataset
    extractFromFile(dset, f, file, session, cfHandler,
aggddimName=aggregateDimensionName, varlocate=varlocate, **context)
  File "/usr/local/uvcdat/2.2.0/lib/python2.7/site-packages/esgset-2.14.6-
py2.7.egg/esgset/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/esgset-2.14.6-
py2.7.egg/esgset/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) ;
float temp_layer(expt, site, layer) ;
float p_level(expt, site, level) ;
float surface_temperature(expt, site) ;
float mer_vapor(expt, site, layer) ;
float mer_monoxide_GM(expt) ;
float s_GM(expt) ;
float B_GM(expt) ;
float I0_GM(expt) ;
float I2_GM(expt) ;
float I4_GM(expt) ;
float I6_GM(expt) ;
float I8_GM(expt) ;
float 4f8_GM(expt) ;
float mon_dioxide_GM(expt) ;
float mon_tetrachloride_GM(expt) ;
float I_GM(expt) ;
float I13_GM(expt) ;
float I14_GM(expt) ;
float I15_GM(expt) ;
float I1_GM(expt) ;
float I1eq_GM(expt) ;
float I2_GM(expt) ;
float I2eq_GM(expt) ;
float c12_GM(expt) ;
float ccl3_GM(expt) ;
float I3_GM(expt) ;
float on1211_GM(expt) ;
float on1301_GM(expt) ;
float on2402_GM(expt) ;
float c141b_GM(expt) ;
float c142b_GM(expt) ;
float c22_GM(expt) ;
float I25_GM(expt) ;
float I34a_GM(expt) ;
float I34aeq_GM(expt) ;
float I43a_GM(expt) ;
float I52a_GM(expt) ;
float 227ea_GM(expt) ;
float 236fa_GM(expt) ;
float 23_GM(expt) ;
float 245fa_GM(expt) ;
float B2_GM(expt) ;
float B65mfc_GM(expt) ;
float 4310mee_GM(expt) ;
float hane_GM(expt) ;
float hyl_bromide_GM(expt) ;
float hyl_chloride_GM(expt) ;
float I_GM(expt) ;
float rous_oxide_GM(expt) ;
float st6_GM(expt) ;
float so2f2_GM(expt) ;
[durack1ml:16/150128_CMIP6/RFMIP] durack1%
```

Problematic data:

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



Name	Date Modified	Size	Kind	
▼ MACv2-SP_v1	Today, 9:37 PM	--	Folder	
.DS_Store	Today, 9:37 PM	6 KB	Document	
MACv2.0-SP_v1.nc	11/9/16, 2:15 PM	19 KB	Document	
orography_T63.nc	11/9/16, 2:13 PM	77 KB	Document	
sp_make_data_v1.ncl	11/9/16, 2:14 PM	12 KB	Document	
mo_simple_plumes_v1.f90	11/11/16, 9:31 AM	19 KB	Fortran Source File	
sp_driver_v1.f90	11/11/16, 10:17 AM	13 KB	Fortran Source File	
README	11/11/16, 10:36 AM	2 KB	TextEdit Document	
MACv2-SP_v1.tgz	11/14/16, 4:01 PM	59 KB	gzip compressed tar archive	

Section 2: Adapting DRS

Making documentation, code and ancillary data a first class ESGF citizen

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..

The screenshot shows the input4MIPs search interface. The browser address bar displays <https://pcmdi.llnl.gov/search/input4mips/>. The page header includes logos for the Department of Energy, Lawrence Livermore National Laboratory, and ESGF. The main title is "Input4MIPs". A navigation bar shows "Home" and "Contact Us". A sidebar on the left lists search filters: Variable, Target MIP, Institution, Dataset Category, Source ID, Grid Label, Grid Resolution, MIP Era, Time Frequency, and Dataset Version Number. The "Dataset Version Number" filter is expanded, showing two options: "1.1.0 (5)" and "1.1.1 (5)". The "1.1.0 (5)" option is selected and circled in red. The search results area shows a search bar with "Enter Text:" and buttons for "Search", "Reset", and "Display 10 results per page". Below the search bar, there are checkboxes for "Show All Versions" and "Search Local Node Only (Including All Replicas)". The search constraints are listed as "PCMDI". The total number of results is 10. The first result is "input4MIPs.PCMDI.SSTsAndSealce.CMIP.PCMDI-AMIP-1-1-0.mon.siconcbcs.gs1x1". The second result is "input4MIPs.PCMDI.SSTsAndSealce.CMIP.PCMDI-AMIP-1-1-0.mon.areacello.gs1x1".

Section 4: Breaking SYNDA

Multiple variable files do not conform to CMIP-standards and..

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.llnl.gov/home

