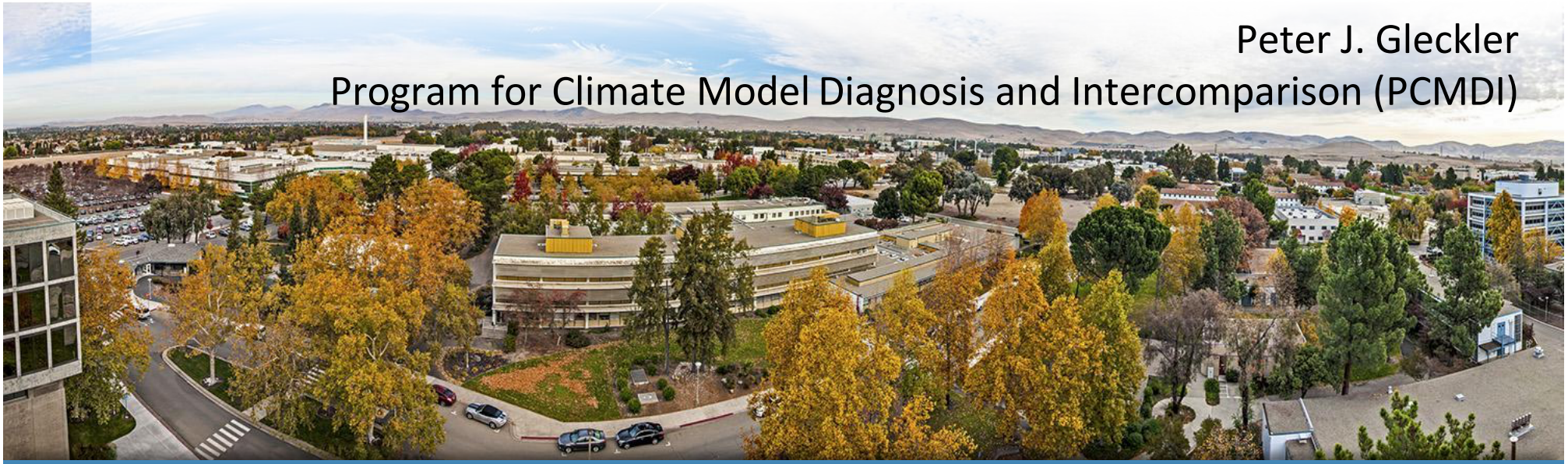
An update on Earth System Grid Federation use in CMIP6, model performance metrics, input4MIPs and **obs4MIPs**

WDAC8, Marrakesh, 20-21 March 2019

Peter J. Gleckler
Program for Climate Model Diagnosis and Intercomparison (PCMDI)



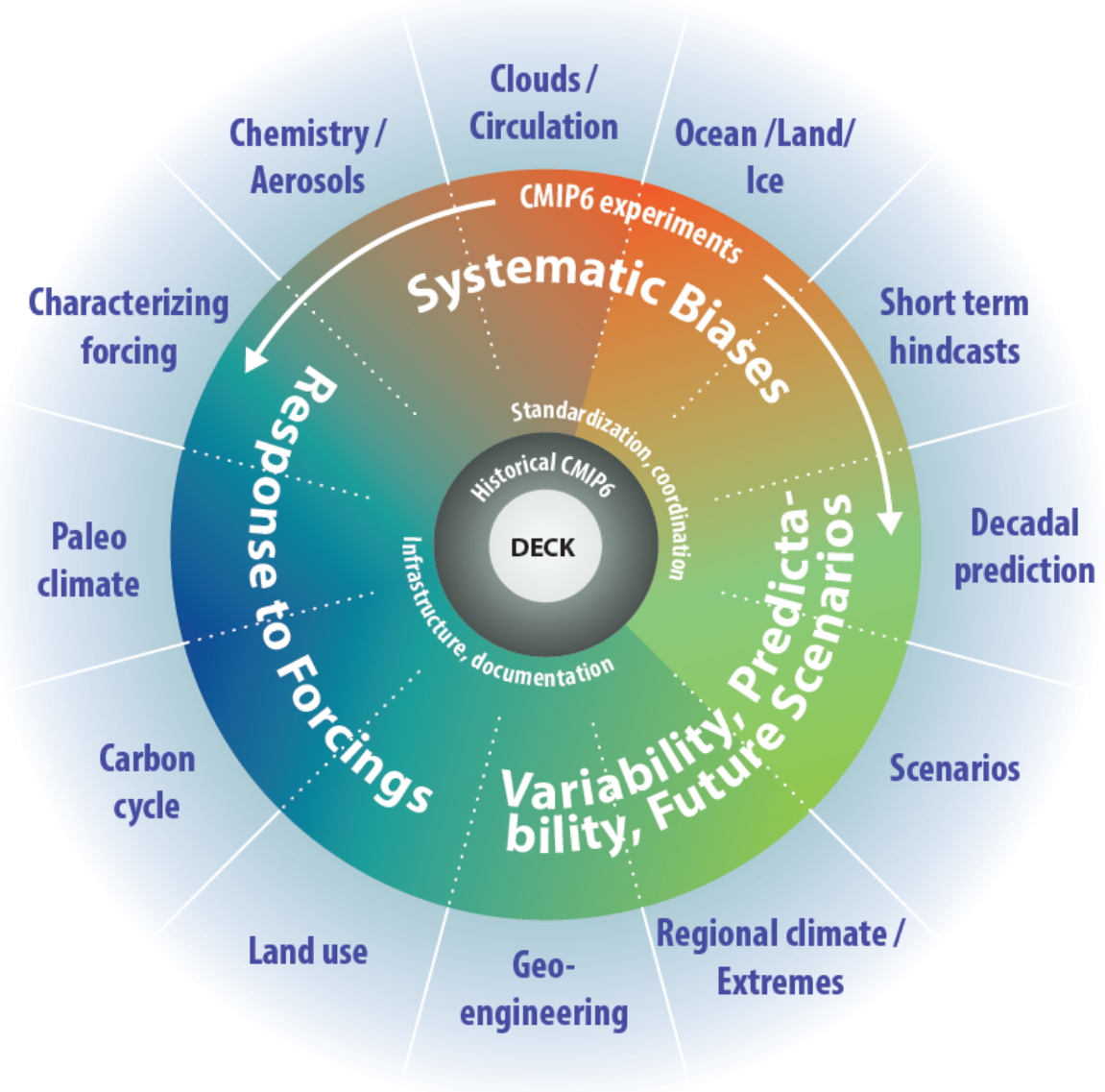
LLNL-PRES-954476
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

CMIP and related activities support a large body of climate research

- Agencies may find it difficult to justify use of “research funds” to support the essential CMIP infrastructure: coordination, leadership, model output post-processing and publishing.
- The value of CMIP is that it *enables* fundamental research
  - Users now expect easy access to multi-model simulation output
  - 100’s of research papers relied on CMIP3 output
  - 1000’s of research papers rely on CMIP5 output
- We should ensure sufficient resources are invested in CMIP infrastructure to continue this record of scientific impact.

CMIP6 design overview:



DECK

- Small set of benchmark runs
- To evolve only slowly (e.g. OMIP, LUMIP)

Historical CMIPx

- Forcing to be updated for each new phase

CMIP6-endorsed MIPs

- An evolving collection to address specific scientific issues

CMIP5/6 evolution: More institutions, more models, more experiments, more data

Thanks to Karl Taylor and Paul Durack for input on CMIP6

- 44 institutions/consortia have registered (CMIP5: 31 inst.)
- 101 models are registered (CMIP5: 59 models)
- 287 experiments defined; 102 tier 1 (CMIP5: 33; 14 tier 1 exps.)
- order **20 PB** of model output expected (CMIP5: ~2 PB)

CMIP6_CVs

https://github.com/WCRP-CMIP/CMIP6_CVs

Core Controlled Vocabularies (CVs) for use in CMIP6

Registering Institutions, Models, or requesting changes to CVs:

To register your institution or model or to request changes to a CV, please submit an issue/ticket following the instructions on the CMIP6_CVs issue page.

Some support for CMIP participating modeling groups is available: pcmdi-cmip@llnl.gov

To view the current `experiment_id` entries point your browser to CMIP6_experiment_id.html

To view the current `institution_id` entries point your browser to CMIP6_institution_id.html

To view the current `source_id` entries point your browser to CMIP6_source_id.html

The CVs build on logic that is described in the CMIP6 Global Attributes, DRS, Filenames, Directory Structure, and CV's document

Lists of registered:

- experiments
- institutions
- models

IPCC timeline

2019	
January 7	Second Lead Author Meeting
April 29	First order draft expert review
August 26	Third Lead Author Meeting
2020	
March 2	Second order draft expert review
June 1	Fourth Lead Author Meeting
October 18	Submission of final draft
2021	
April 16	IPCC acceptance/adoption/approval

31 December 2019:
Journal articles submitted

30 September 2020:
Journal articles accepted

https://www.wcrp-climate.org/news/wcrp-news/1360-cmip6-ma-workshop

Search

WCRP World Climate Research Programme

Search Ok

Advanced search

About WCRP | Core Projects | Unifying Themes | Grand Challenges | Initiatives & Activities | Events | News | Resources

CMIP6 Model Analysis Workshop

25-28 March 2019, Barcelona Supercomputing Center (BSC), Barcelona (Spain)

The Coupled Model Intercomparison Project Phase 6 (CMIP6) Model Analysis workshop is jointly organized by the WCRP Working Group on Coupled Modelling (WGCM) CMIP Panel and the European Commission Horizon 2020 projects PRIMAVERA (PRocess-based climate sIMulation: AdVances in high-resolution modelling and European climate Risk Assessment) and EUCP (EUropean Climate Prediction system). Following the format of the WCRP CMIP5 model analysis workshop held in 2012, the workshop focus will be on:

- Multi-model CMIP6 analyses and evaluation that take advantage of the large suite of CMIP6 experiments
- Efforts to connect model development and analysis to identify Earth system model improvements that help reduce systematic biases and/or increase the realism of models
- Methods for multi-model analysis
- Climate change impacts

The workshop will be structured around the three scientific questions:

1. How does the Earth system respond to forcing?
2. What are the origins and consequences of systematic model biases?
3. How can we assess future climate change given climate variability, predictability, and uncertainty in scenarios?

Workshop approach: Short-presentation/poster format

The workshop will consist of a series of seven half-day sessions of three hours each. Each session will begin with 20-25 presenters given a 3-minute time slot to show no more than one slide summarizing the main conclusions of their poster. The rest of the half-day session will consist of viewing posters of that session. In addition, there will be a plenary talk each day.

Participation is limited by the size of the venue (~200 people) and format of the workshop. Abstracts will be accepted based on relevance to the workshop focus.

WCRP News

- WCRP News
- WCRP Highlights
- WCRP Newsletter
  - WCRP News Articles
  - Subscribe to WCRP News
- Main RSS feed

WCRP News Archive

- December, 2018
- November, 2018
- October, 2018
- September, 2018
- August, 2018
- July, 2018
- June, 2018
- May, 2018
- April, 2018
- March, 2018
- February, 2018
- January, 2018

Next week!

The Earth System Grid Federation (ESGF) is being used for CMIP6 and other WCRP projects

ESGF data is distributed across 22 nodes

CMIP6 status: data availability (as of 15 March)

https://pcmdi.llnl.gov/CMIP6/ ArchiveStatistics/esgf_data_holdings/

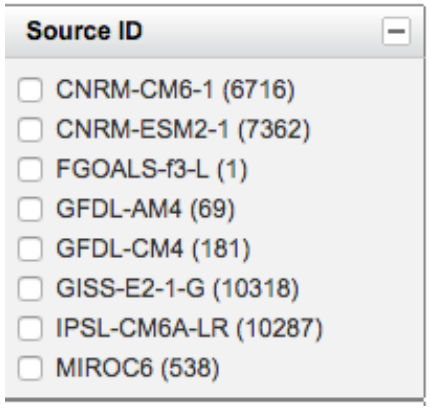
- Model output now being served by ESGF from 10 institutions (15 models) and 21 experiments
- Much more output to be made available over the next year

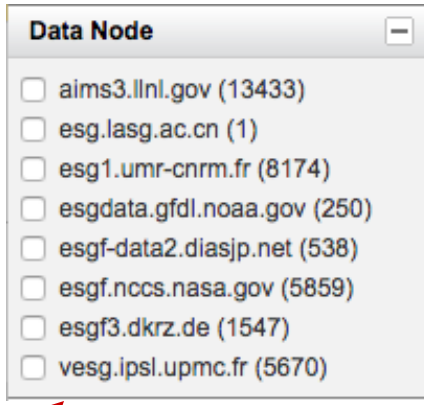
# of datasets archived

source_id	# of expts	historical	piControl	1pctCO2	amip	abrupt-4xCO2
# of models	44	10	10	7	10	7
BCC-CSM2-MR	4	578	147	148		147
BCC-ESM1	2	542	137			
CESM2	2	8760			421	
CESM2-WACCM	2	2514			1409	
CNRM-CM6-1	5	4079	302	387	500	1818
CNRM-ESM2-1	5	2951	440	1515	564	1462
E3SM-1-0	1		17			
FGOALS-f3-L	1				1	
GFDL-AM4	1				69	
GFDL-CM4	1		356			
GISS-E2-1-G	5	1711	176	166	665	166
IPSL-CM6A-ATM-HR	0					
IPSL-CM6A-LR	5	22648	2490	835	3373	9441
MIROC6	5	1090	109	109	930	109
MRI-ESM2-0	5	270	54	54	162	702

All available CMIP6 data is exposed at 3 data portals

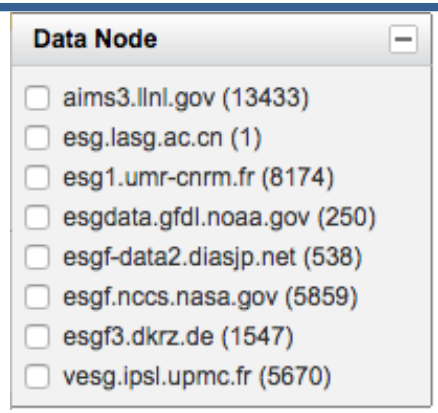
- Each portal currently federates the 8 data nodes:
  - PCMDI (USA) https://esgf-node.llnl.gov/search/cmip6/
  - CEDA (UK) https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/
  - IPSL (France) https://esgf-node.ipsl.upmc.fr/search/cmip6-ipsl/
  - coming soon: DKRZ (Germany), GFDL (USA)
- The search interface provides up-to-date listing of available models and experiments

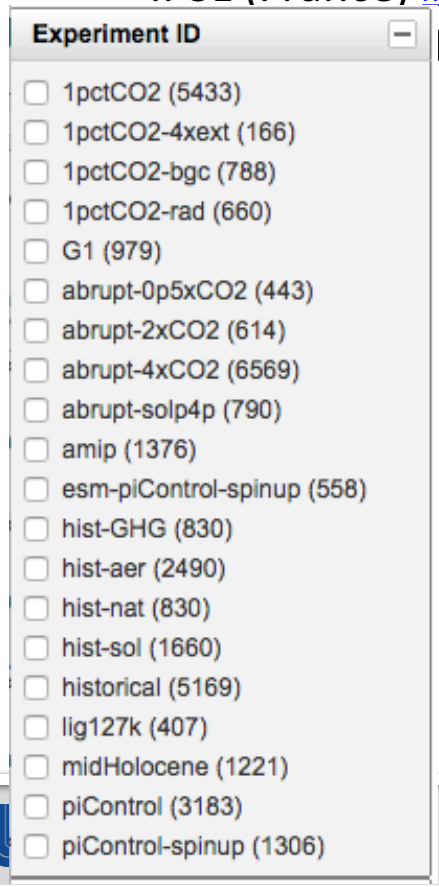




All available CMIP6 data is exposed at 3 data portals

- Each portal currently federates the 8 data nodes:
  - PCMDI (USA) https://esgf-node.llnl.gov/search/cmip6/
  - CEDA (UK) https://esgf-index1.ceda.ac.uk/search/cmip6-ceda/
  - IPSL (France) https://esgf-node.ipsl.upmc.fr/search/cmip6-ipsl/
  - DKRZ (Germany), GFDL (USA)

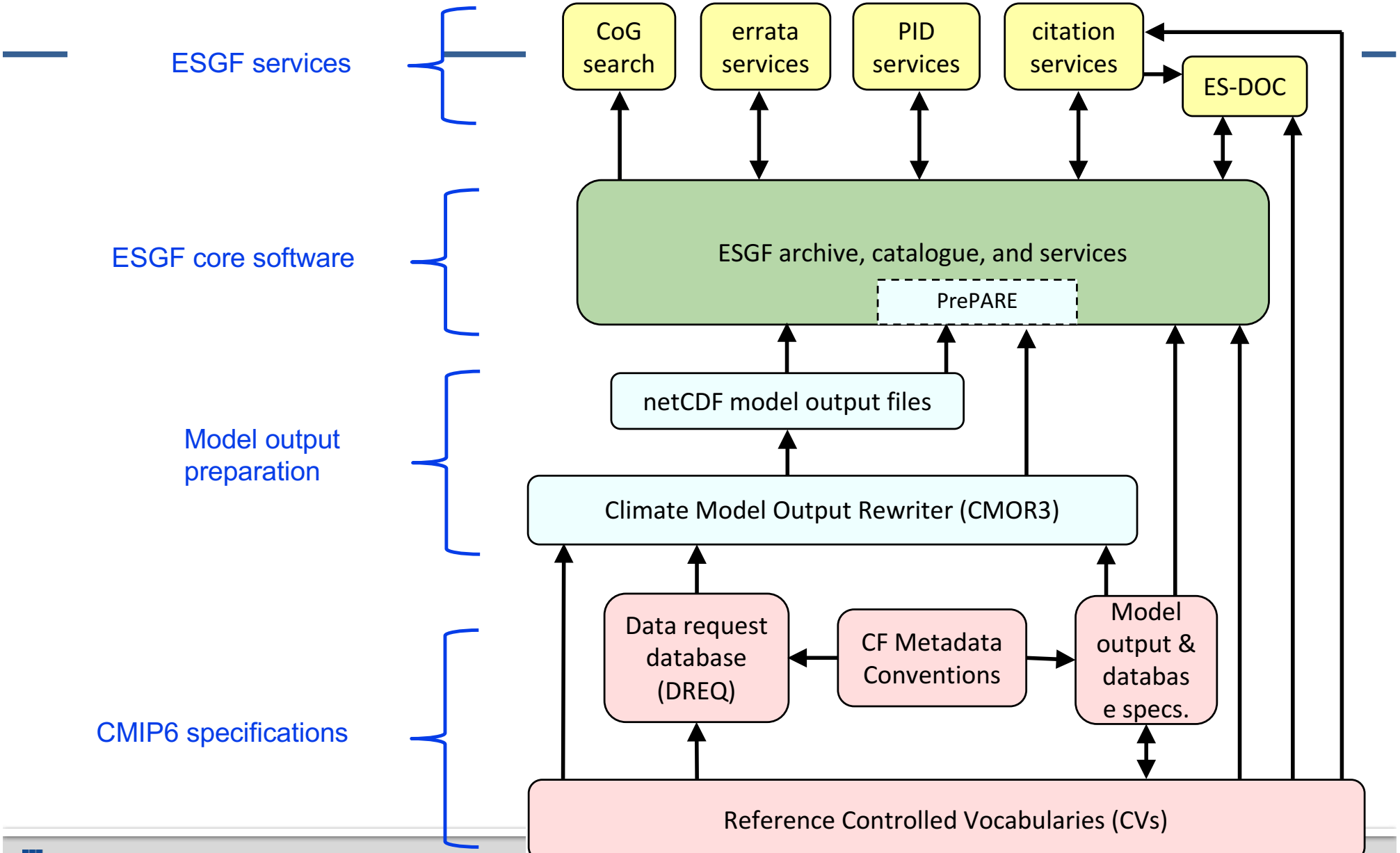




ESGF core software stack

- Supports a federated data archive hosting the CMIP6 data
- Status: In place and operational!
- Replication procedure is working
  - As of March 15 2019 PCMDI has replicated ~85% of available datasets

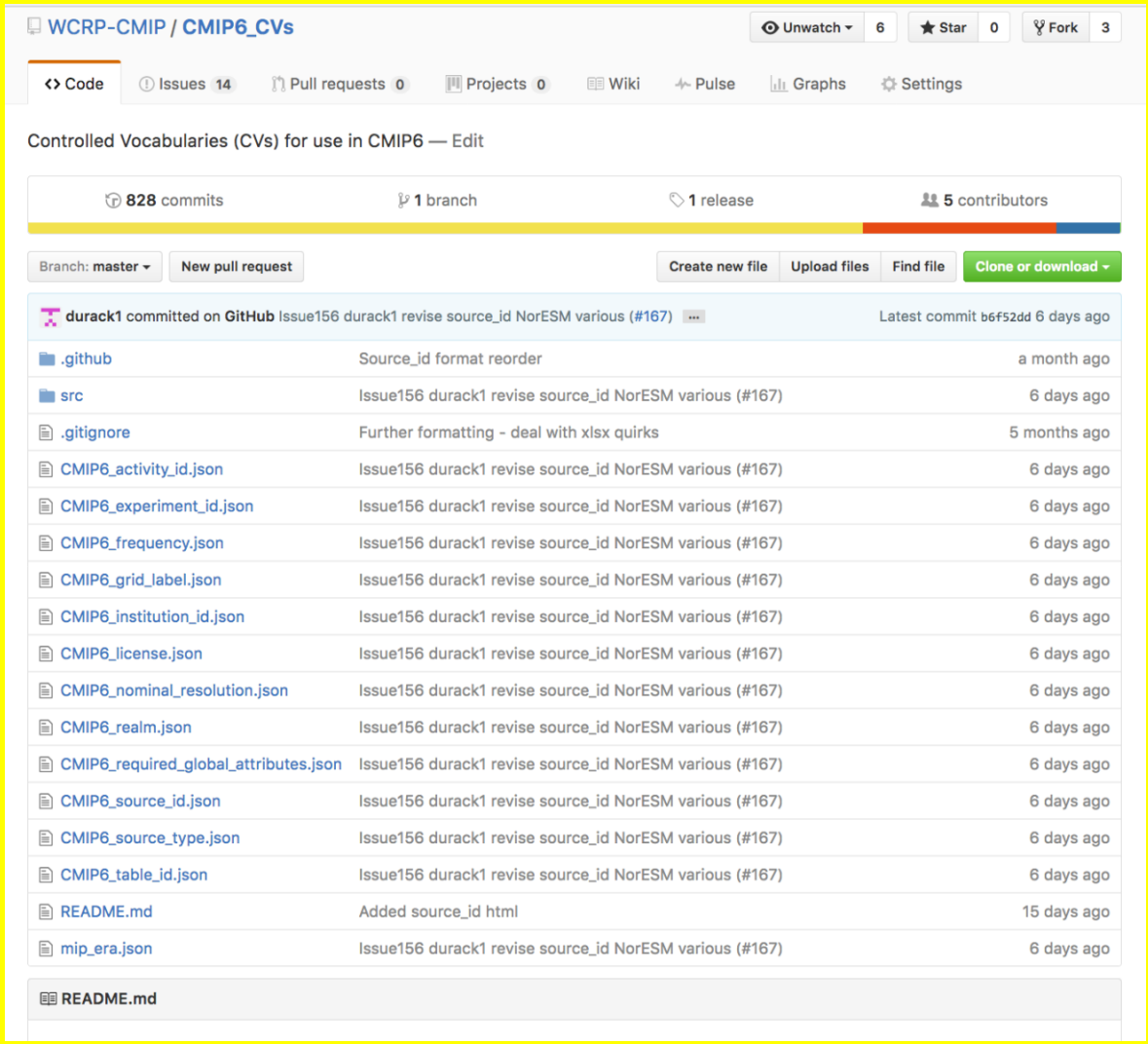
Infrastructure components and dependencies



A major advance for CMIP6: “controlled vocabularies” (CVs) now machine readable for use across infrastructure components

- CVs allow users and individual infrastructure elements to communicate.
- Recorded in JSON files
- Status: all needed CV’s defined, including activity, Institution, model, experiment, sub-experiment, realm, frequency, …..

https://github.com/WCRP-CMIP/CMIP6_CVs

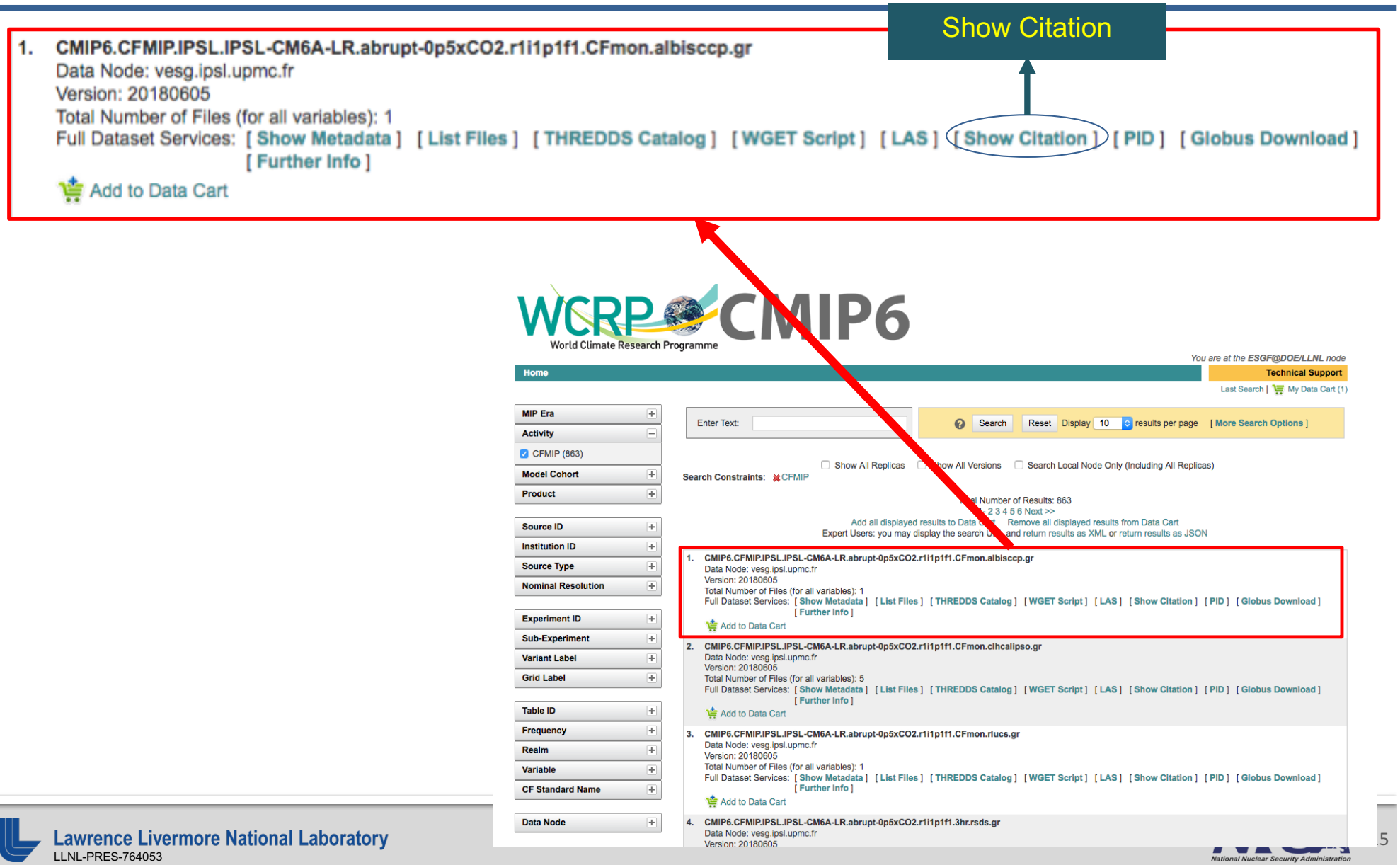


Climate Model Output Rewriter (CMOR3) released with new module for verifying metadata

- CMOR is used by many modeling groups to ensure their model output meets CMIP6 requirements
- PrePARE is a new module that
  - checks that CMIP6 output *not* processed by CMOR meets CMIP6 requirements
  - is executed by the ESGF publisher to ensure only CMIP6 compliant files are published.
- Status: in place and in use!
  - Code available at https://github.com/PCMDI/cmor
  - Documentation available at http://cmor.llnl.gov/
  - Development phase is complete
  - Bugs corrected when discovered

Remember this one for obs4MIPs discussion

Data citation services are linked to CoG search interface



Citation page

Metadata for 'CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2'

General Information

General Information

Name CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2

Abstract Coupled Model Intercomparison Project Phase 6 (CMIP6) data sets.
These data includes all datasets published for 'CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2' according to the Data Reference Syntax defined as
'mip_era.activity_id.institution_id.source_id.experiment_id.member_id.table_id.variable_id.grid_label.version'.

The Earth System Model IPSL-CM6A-LR, released in 2017, includes the components:
atmos: LMDZ (NPv6, N96; 144 x 143 longitude/latitude; 79 levels; top level 40000 m), land: ORCHIDEE (v2.0, Water/Carbon/Energy mode), ocean: NEMO-OPA (eORCA1.3, tripolar primarily 1deg; 362 x 332 longitude/latitude; 75 levels; top grid cell 0-2 m), ocnBgchem: NEMO-PISCES, seaIce: NEMO-LIM3.
The model was run by the Institut Pierre Simon Laplace, Paris 75252, France (IPSL) in native nominal resolutions: atmos: 250 km, land: 250 km, ocean: 100 km, ocnBgchem: 100 km, seaIce: 100 km.

Project: These data have been generated as part of the internationally-coordinated Coupled Model Intercomparison Project Phase 6 (CMIP6; see also CMD Special Issue: http://www.geosci-model-dev.net/special_issue590.html). The

Cite this data

Citation **(2018)**. *IPSL IPSL-CM6A-LR model output prepared for CMIP6 CFMIP abrupt-0p5xCO2*. Earth System Grid Federation.
http://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2

BibTeX RIS

Model and experiment documentation linked to CoG search interface

1. **CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2.r1i1p1f1.CFmon.albisccp.gr**
   Data Node: vesg.ipsl.upmc.fr
   Version: 20180605
   Total Number of Files (for all variables): 1
   Full Dataset Services: [Show Metadata] [List Files] [THREDDS Catalog] [WGET Script] [LAS] [Show Citation] [PID] [Globus Download]
   [Further Info]
   Add to Data Cart

Further Info

Model and experiment documentation by es-doc

es-doc
Earth System Documentation

CMIP6 Further Information v0.5.0.0 Support Help

Further Info URL: https://furtherinfo.es-doc.org/CMIP6.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2.none.r1i1p1f1

ES-DOC Documentation

MIP Era	CMIP6
Institution	IPSL
Model	IPSL-CM6A-LR
Experiment	abrupt-0p5xCO2
Ensemble Description	N/A
Machine Performance	N/A

Dataset Documentation

Dataset ESGF Search	N/A
Dataset Errata	N/A
Dataset Citation(s)	https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CFMIP.IPSL.IPSL-CM6A-LR.abrupt-0p5xCO2

Other Documentation

WCRP CMIP6 Homepage	https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6
ES-DOC CMIP6 Homepage	https://es-doc.org/cmip6

Summary:

- The major CMIP6 infrastructure elements are in place and working satisfactorily.
- ESGF expected to serve data through CMIP6 research cycle (>5 yrs)
- High priorities:
  - Enable globus gridftp at all data nodes
  - Provide server-side computation capability at some portals
  - Encourage modeling groups to provide documentation through ES-DOCS
- Major issue: No funding available for user-support
- Consider establishing a distributed “facility”, supported by multiple agencies and internationally, to fund CMIP infrastructure and operations.

input4MIPs:
CMIP infrastructure enabling science

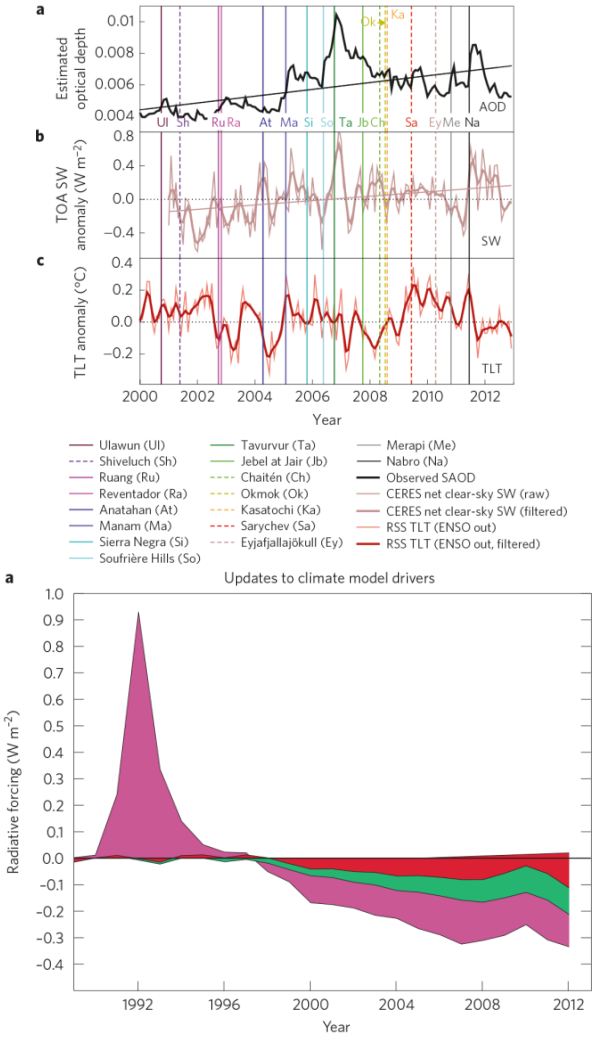
Planning for future forcing data to drive next generation CMIP simulations

Paul J. Durack and Karl E. Taylor
Program for Climate Model Diagnosis and Intercomparison (PCMDI)

LLNL-PRES-954476
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

input4MIPs: motivation

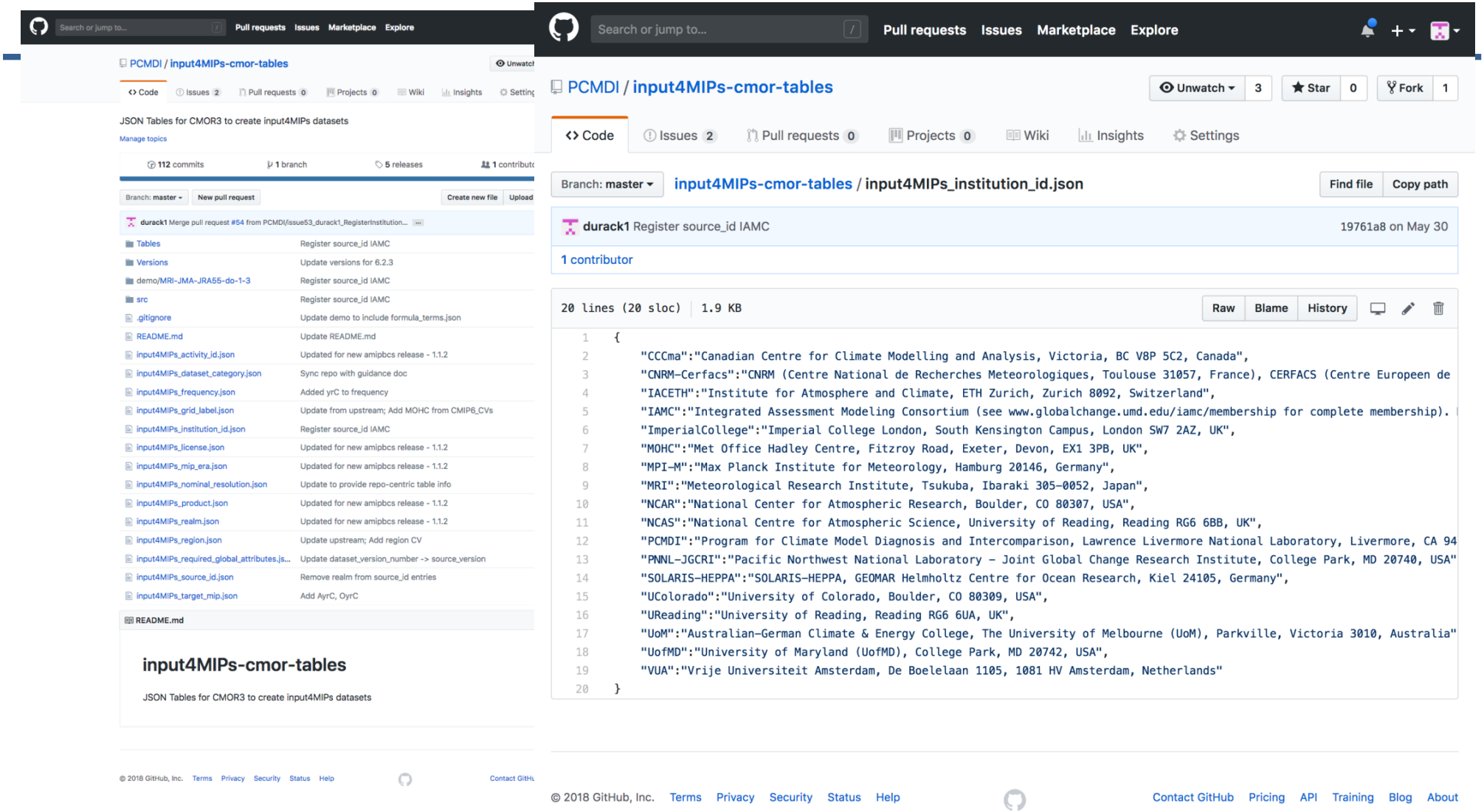


- Improving forcing datasets and their standard usage for CMIPx has been actively discussed for sometime
- After CMIP5 was complete issues were found with the forcing datasets
- Numerous assessments noted discrepancies e.g. with the volcanic forcing data used with a number of forcing-relevant volcanoes omitted from the post-2000 time history (e.g Santer *et al*., 2014 NATGeo; Schmidt *et al*., 2014 NATGeo)
- Problems also known for future scenarios
- input4MIPs provides more comprehensive versioning, collating/archiving and documentation of the datasets used for simulations, in particular the *historical* simulations

Project support

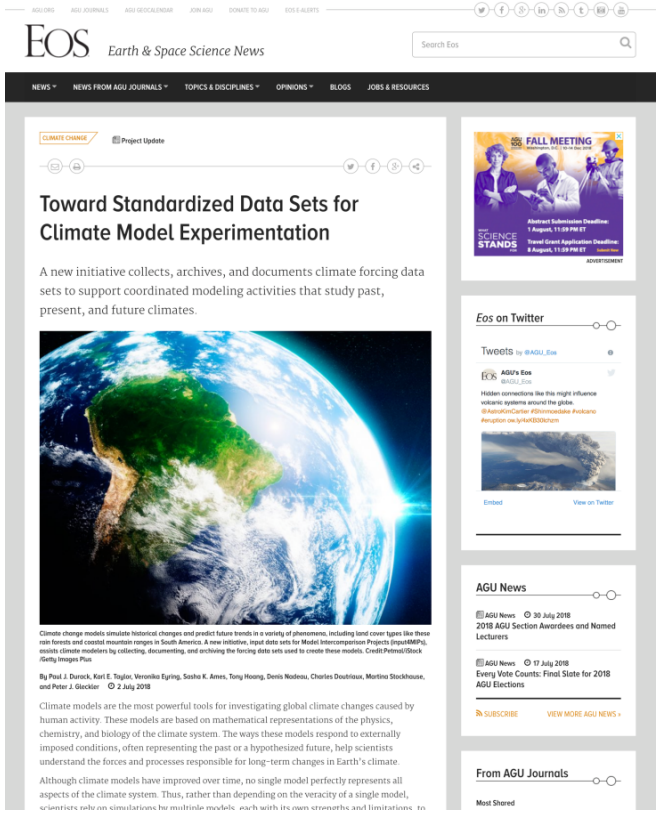
- LLNL contributions are supported by the US Dept. of Energy, Office of Science
  - PCMDI (Science leadership)
  - ESGF/AIMS (infrastructure support)
- DKRZ, Germany provides citation service
- Numerous individuals and their supporting agencies to establish the project
- 18 contributing organizations and their supporting agencies
- 10 contributing countries
- Supporting CMIP6

CMIP6: input4MIPs-cmor-tables



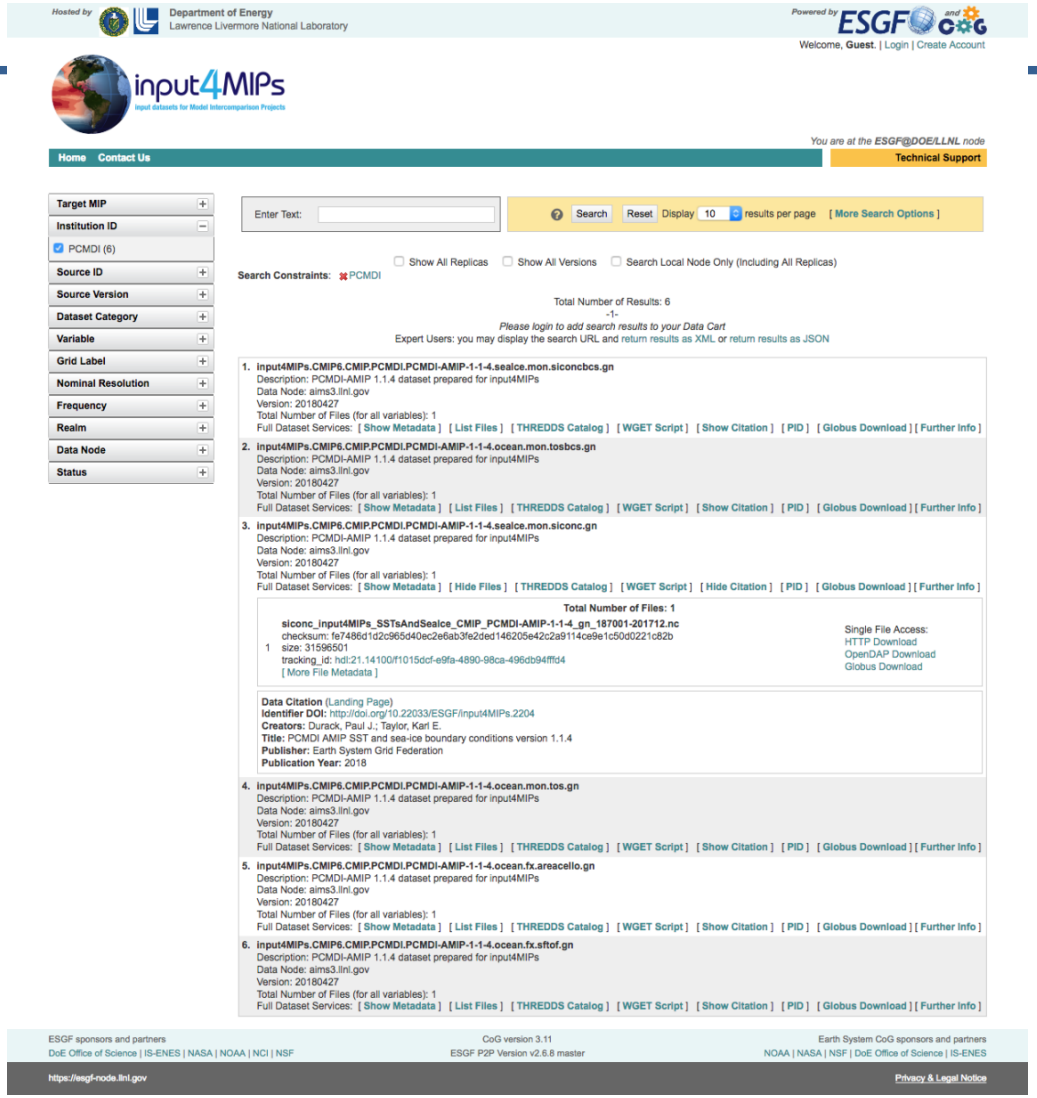
https://github.com/PCMDI/input4MIPs-cmor-tables

input4MIPs: current status



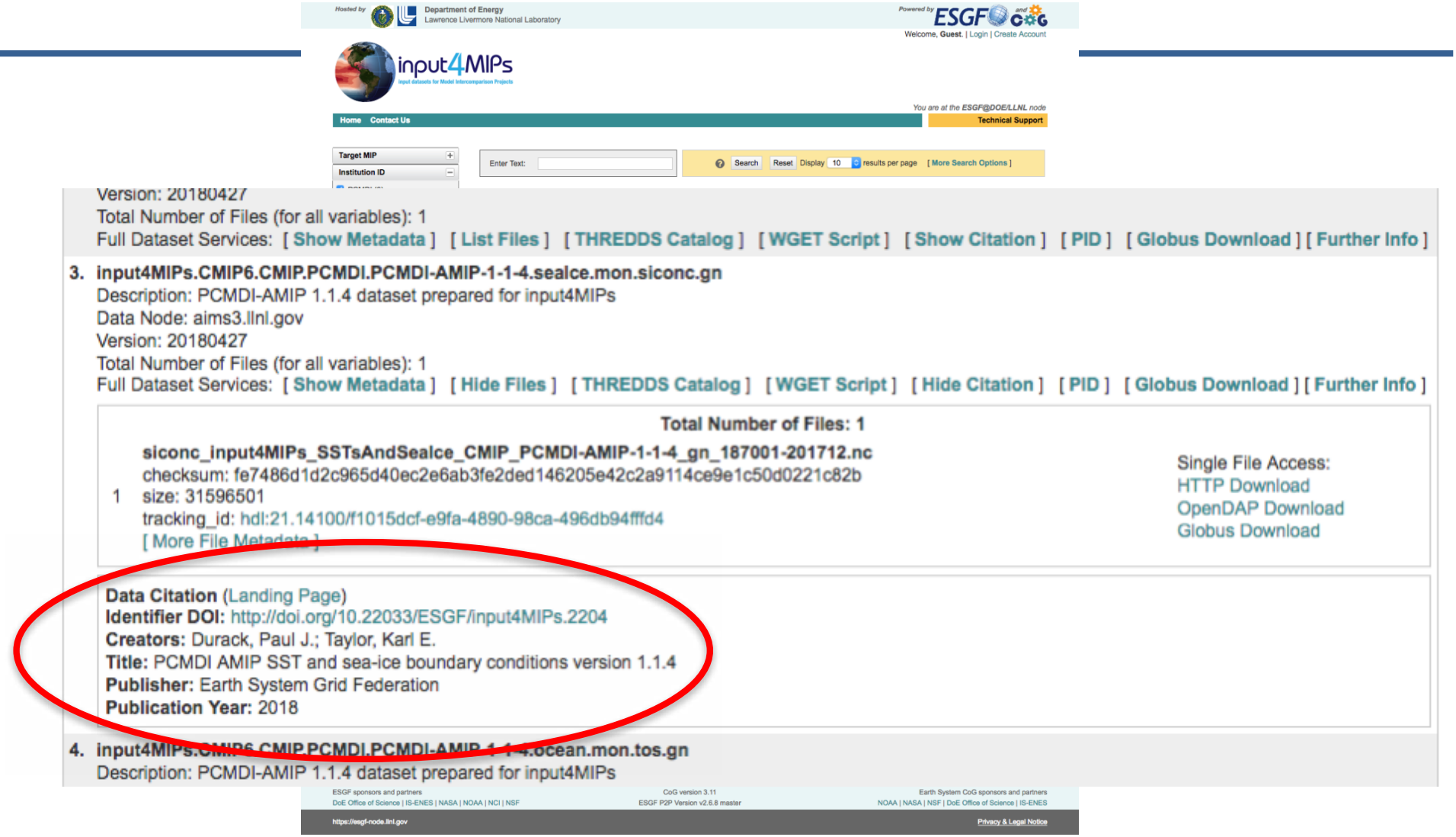
- Current dataset collection ver 6.2.18
- 7304 netcdf files
  - 1.5Tb total storage footprint
  - 13 MIPs served
  - Latest information http://goo.gl/r8up31
- Datasets being updated/added
  - 6.2.19 (~December 2018) – ScenarioMIP – Future Land Use v2.1f – added_tree_cover
  - 6.2.20 (~December 2018) – CMIP – AMIP Boundary forcing v1.1.5
  - 6.2.21 (~January 2019) – CMIP – Stratospheric aerosol v4.0.0
  - 6.2.22 (~March 2019) – ScenarioMIP – ozone (additional experiments)
  - Dataset collection version history http://goo.gl/r8up31

CMIP6: input4MIPs ESGF project



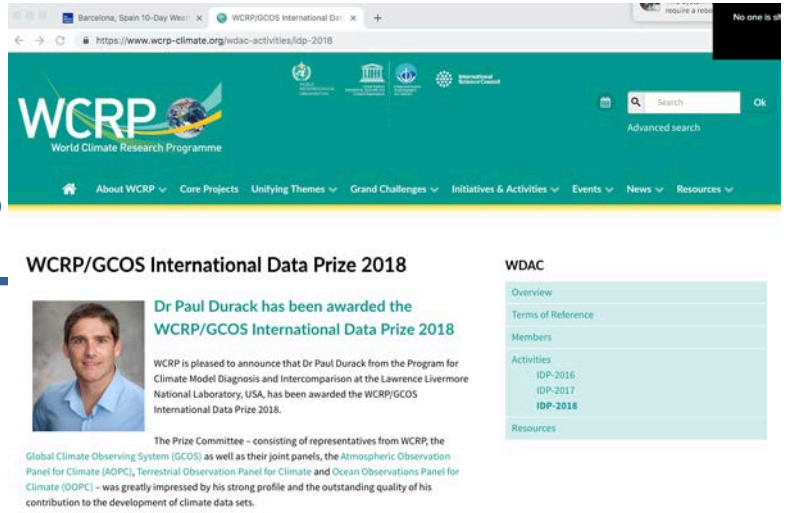
https://esgf-node.llnl.gov/search/input4mips/

CMIP6: input4MIPs ESGF project



https://esgf-node.llnl.gov/search/input4mips/

input4MIPs
CMIP infrastructure connections



- Datasets used by simulations are curated (with version info)
- As datasets are trace-able they can be documented
  - tracking_id connects to DKRZ citation service, providing DOI
- ES-DOCs will provide input4MIPs dataset collection info to modeling groups so
  - Datasets and the versions used in simulations can be accurately recorded
  - ES-DOCs picks up tracking_id/DOI to provide connection to data citation in CMIP6 model documentation
  - DOIs (and tracking_ids) provide digital connectivity for documentation to leverage

An update on obs4MIPs

WDAC Observations for Model Evaluation Task Team

Peter Gleckler, co-chair, PCMDI and **Duane Waliser**, co-chair, JPL/NASA
Mike Bosilovich, GSFC/NASA
Helene Chepfer, IPSL
Carol Anne Clayson, WHOI
Veronika Erying, DLR
Robert Ferraro, JPL/NASA
Pierre-Phillipe Mathieu, ESA
Jerry Potter, GSFC
Roger Saunders, UKMO
Jörg Schulz, EUMETSAT
Karl Taylor, PCMDI
Jean-Noël Thépaut, ECMWF

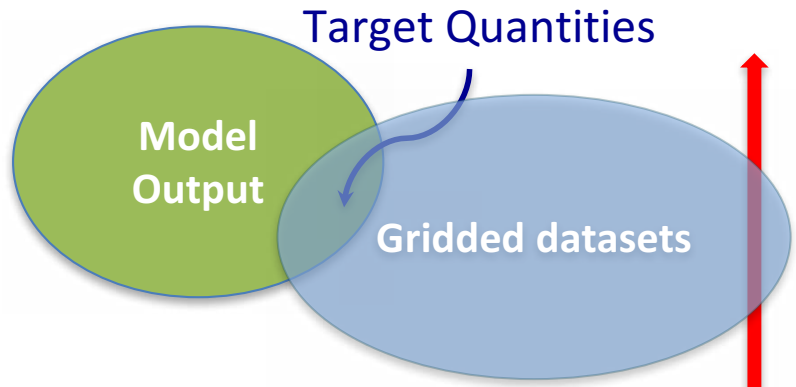
Additional regular contributors: Otis Brown, Michel Rixen
Tsengdar Lee (NASA) and Renu Joseph (DOE)
Luca Cinquini (JPL) – CoG technical support
Denis Nadeu (PCMDI) – CMOR development
Paul Durack (PCMDI) – Data specifications
Sophie Cloché (IPSL) – CFMIP archive
Jim Biard (NCEI) and Matthias Tuma (WCRP) – beta testers
… and many others

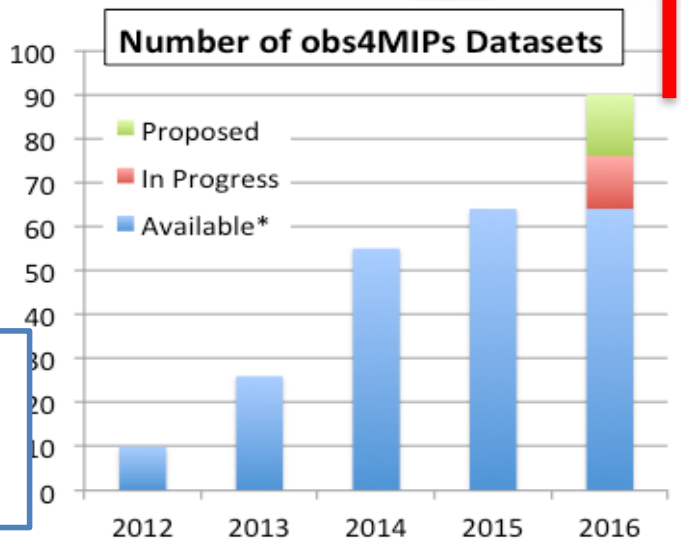
WDAC8, Marrakesh, 20-21 March 2019

obs4MIPs

https://www.earthsystemcog.org/projects/obs4mips

- A project for identifying, documenting and disseminating observations for climate model evaluation in WCRP model intercomparisons, notably CMIP.
- Data (and tech notes) accessible with the distributed CMIP model output via ESGF, adhering to same conventions
- Guided by the WCRP Data Advisory Council obs4MIPs Task Team





Complete (~125)
In Progress* (~15)
Proposals from Data Call (~100)

…. and growing!

MOTIVATION

Initially,

- obs4MIPs was envisioned as a vehicle to provide 1000s of CMIP analysts access to satellite/gridded products that are technically aligned with model output along with added value information. This is still the primary motivation, however…

More recently,

- As efforts to systematically and routinely evaluate models advances, it is clear that the obs4MIPs protocols will be invaluable

Towards systematic evaluation of the CMIP DECK and historical simulations

Peer-reviewed CMIP based publications will remain the primary way the research is documented. However, there are pressing reasons to more efficiently produce, summarize, and make well-established model evaluation results available:

- Inform national assessments, the IPCC process, stakeholders, and public
- More directly contribute to model development (via useful quick feedback)
- Advance science more efficiently (provide routine summaries, less re-inventing)

Community-based model evaluation capabilities are becoming a reality, thanks to the design target provided by the CMIP/obs4MIPs standards. Examples include CMEC (Coordinated Model Evaluation Capabilities) and ESMValTool, both which entrain multiple analysis tools.

obs4MIPs planning meeting for CMIP6

April 2014, NASA HQ but still relevant!

Selected consensus recommendations

- **Expand the inventory**
- **Include more higher frequency** data (a “golden period”?)
- Reliable and defendable **error characterization**/estimation of observations
- Include datasets in support of **off-line simulators** (prime example: COSP—[CFMIP] Observation Simulator Package)
- **Collocated observations, including in-situ** for processes level diagnostics
- **Precise definitions of data products** (what’s actually being reported), including biases, and precise definitions of the model output variables are required

Ferraro et al. (2015) BAMS and full meeting report on CoG website

Progress since WDAC7

- obs4MIPs data specifications (ODS2.1) now being used!
- Primary CoG site migrated (with content updates) to PCMDI
- Prototyping ESGF publication transition with new standards
- New search facets implemented (March 01 2019)
- Contributions are actively being augmented and expanded by:

  NASA JPL, NASA Goddard, NOAA NCEI, ESA and others

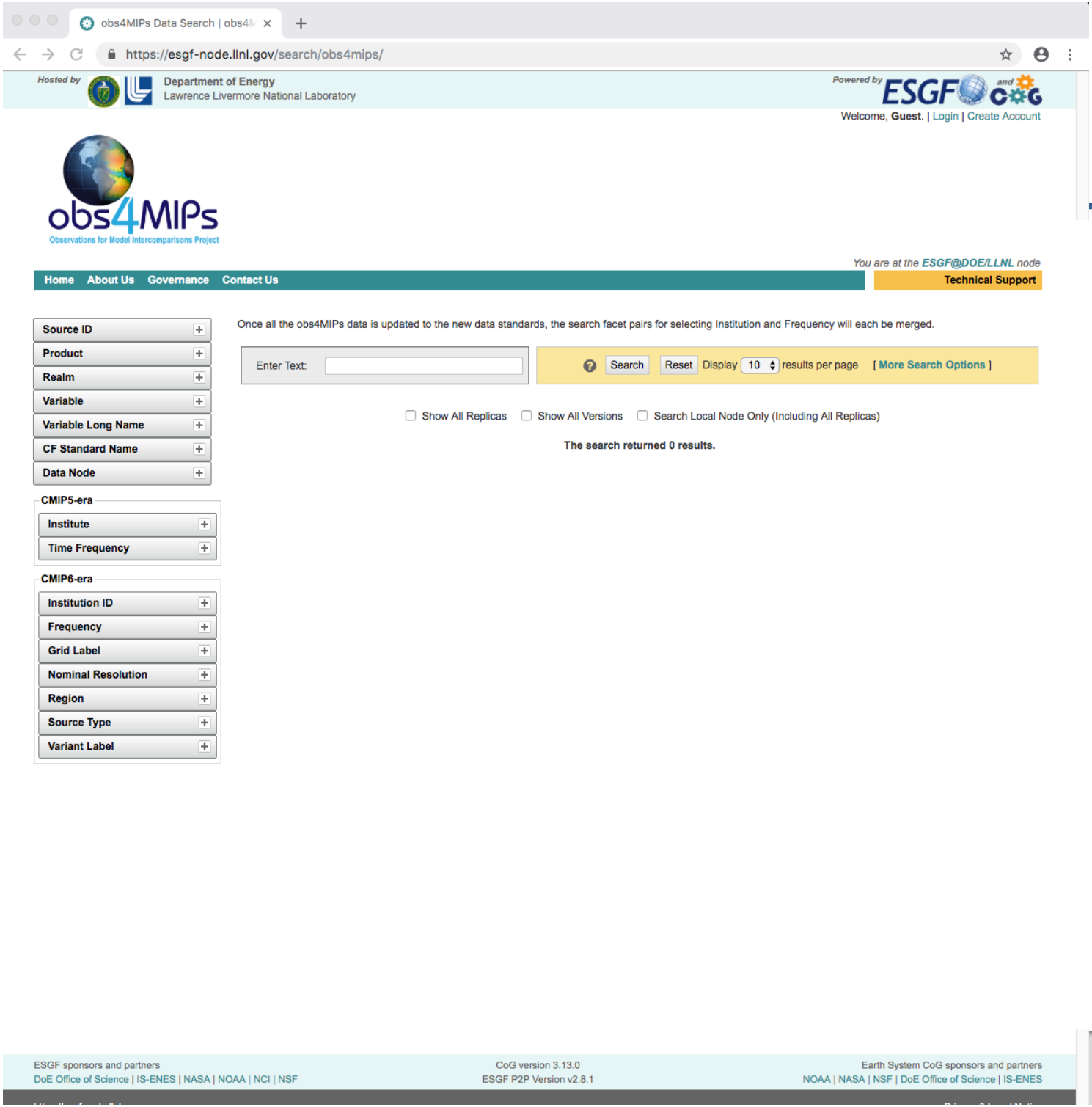
- Two papers about ready to submit (Overview and data standards)
- 4 smaller task team telecons

Coordinated CMIP/obs4MIPs global attributes, controlled vocabulary (CV), Registered Content (RC), and Data Reference Syntax (DRS)

Some predefined global attributes (there are many others)

Variable ID #	**'sfcWind'**
Source_id +	**'NOAA-NCEI-SeaWinds-1-2'**
Institution_id +	**'NOAA-NCEI'**
Region #	**'global ocean'**
Nominal_resolution #	**'1x1 degree'**

+ Registered Content (RC)
# CV with pre-defined options maintained on github



Expanded search capabilities

- Source ID
- Product
- Realm
- Variable
- Variable Long Name
- CF Standard Name
- Data Node

CMIP5-era

- Institute
- Time Frequency

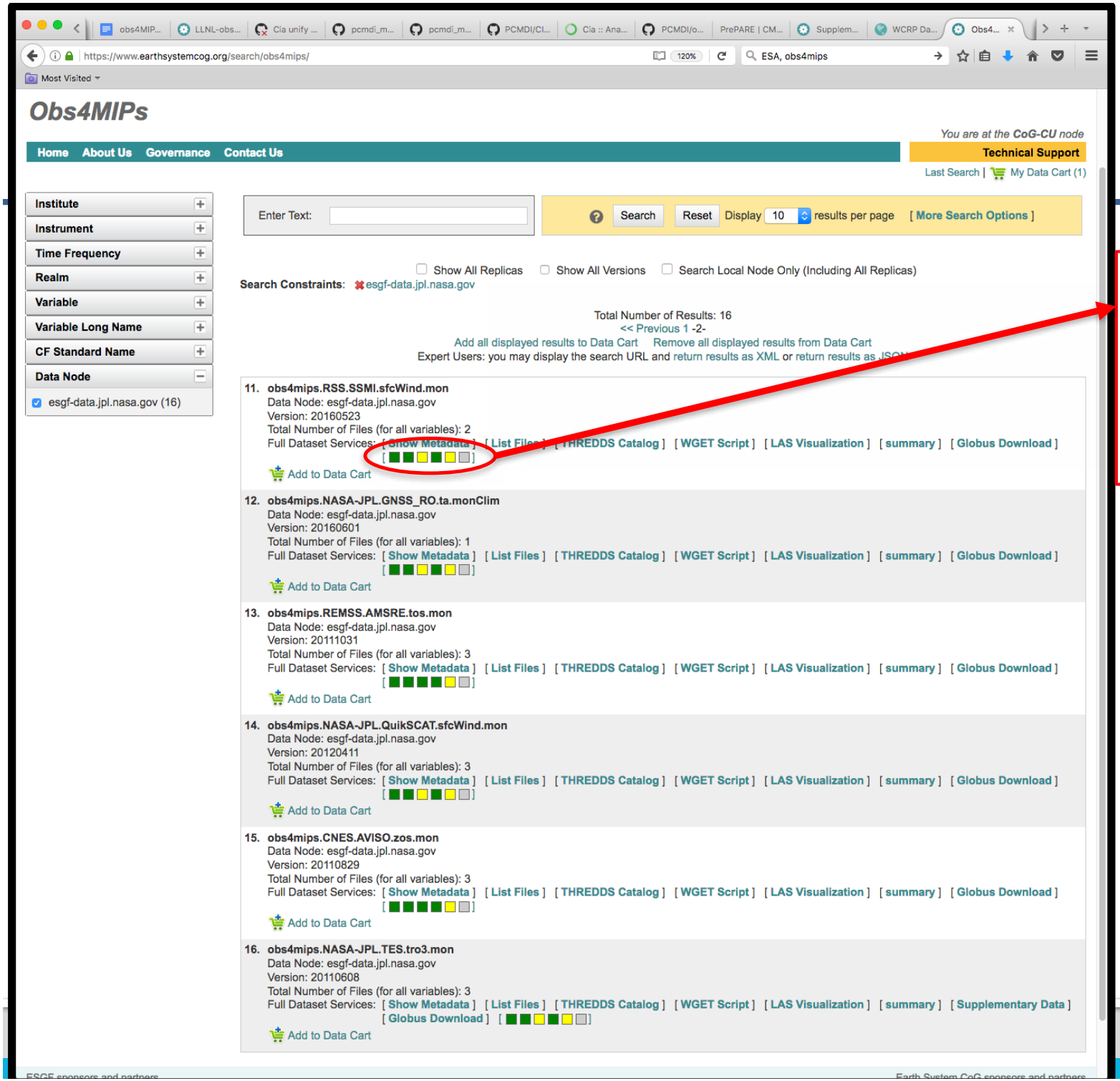
CMIP6-era

- Institution ID
- Frequency
- Grid Label
- Nominal Resolution
- Region
- Source Type
- Variant Label

obs4MIPs Dataset Suitability & Maturity Indicators

Enables us to expand what data gets included in obs4MIPs

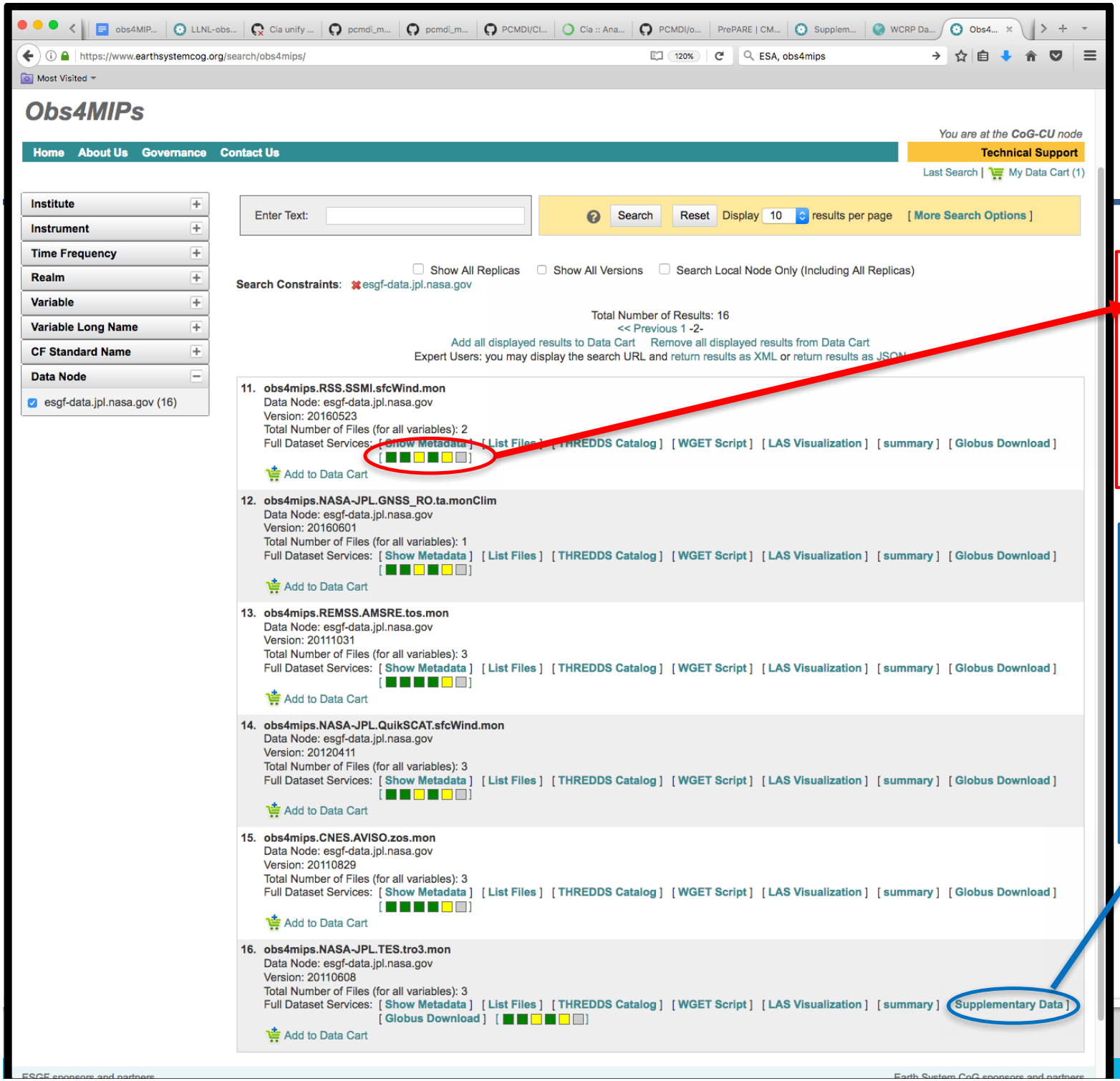
Technical Requirements		***Dataset Suitability and Maturity***			***Comparison Complexity***
Meets obs4MIPs data technical requirements	Includes obs4MIPs technical note information	Closeness or robustness of measurement to observed reference quantity	Maturity with respect to climate model evaluation	Provision for robust uncertainty information	Complexity of Model Observation Comparison
Data suitably processed with CMOR and/or consistent with obs4MIPs standards	Complete technical note information provided	Measurement approach provides a very close relationship to observation quantity	Multiple peer-reviewed examples of application to climate model evaluation	Uncertainty information provided per retrieval/grid point	Comparison can be made directly with CMIP model output variable
Largely complete with minor metadata inconsistencies	Technical note information incomplete and/or could be improved	Measurement approach requires complex and/or non-linear retrieval methods and/or subjective inferences/definitions	One peer-reviewed example of application to climate or component model evaluation.	General uncertainty information given relative to the methodology and dataset as a whole - backed by actual field/in-situ validation exercises	Comparison requires some simple post processing of CMIP output variable(s) (e.g. vertical integral or ratio of two variables)
Non-compliant. Should be removed from database!	Technical note not provided	Measurement approach requires significant use/influence from complex or weakly constrained model and/or has significant ambiguity in definition(s)	No peer-reviewed xamples of application to model evlauation	No uncertainty information provided	**Comparison requires complex processing of CMIP output (e.g. "simulator", budget calculation)**



Color coded **suitability indicators** To be monitored by the task team

Prototyped with JPL data

37



Color coded **suitability indicators** To be monitored by the task team

Supplemental Information is "free form", accessible from "best estimate" but not searchable independently

Prototyped with JPL data

38

Just a few more enhancements needed…

- Upgrade the daily scan of inventory to include suitability indicators (color codes), supplemental information, and tech notes on a unified page – this should be accomplished within one month
- CMOR “Prepare” utility needs to be generalized to accommodate observations - this will require some investment but we are targeting for this to be done before 2020.

Neither will slow down the inclusion of new datasets

Summary and Perspective

- obs4MIPs is positioned to have substantial impact on CMIP6 and other activities
- Task team has been able to address many recommendations (e.g., enable more data and information to be included)
- With infrastructure now in place, the goal for the coming year is to substantially expand the obs4MIPs archive, exploiting the new obs4MIPs data specifications (ODS2.1), supplemental information and data indicators
- Ongoing challenge: further enabling datasets to efficiently be made to meet the obs4MIPs data specifications and published on ESGF
- For each dataset, someone managing an ESGF node has to commit to dataset publication
- Efforts to advance obs4MIPs remain focused on gridded datasets. Expanding the scope to include in-situ data will required new contributors prepared to do substantial work

Summary and Perspective ii

- Project is entering a new phase – implementing what the tasked team has envisioned, implemented, and developed with the help of computer scientists
- As two manuscripts are completed, it will be an appropriate time to revisit the make-up of task team with more emphasis on those who can enable datasets to get published to ESGF; but scientific expertise still needed!
- While a great deal of infrastructure is now in place, a broader obs4MIPs would require more people doing hands on work.
- Irrespective of the data delivery system (ESGF), the obs4MIPs data specifications solidify a technical link with the modeling community

Links to related material

- CMIP6 data specifications (google docs)
- obs4MIPs data specifications
- obs4MIPs tables controlled vocabulary (github)
- Climate Model/obs Output Rewritter, CMOR (website)

EXTRAS