How Can Communities Less Familiar with Climate Models be Better Served Data?

Karl E. Taylor

Program for Climate Model Diagnosis and Intercomparison (PCMDI)
Lawrence Livermore National Laboratory

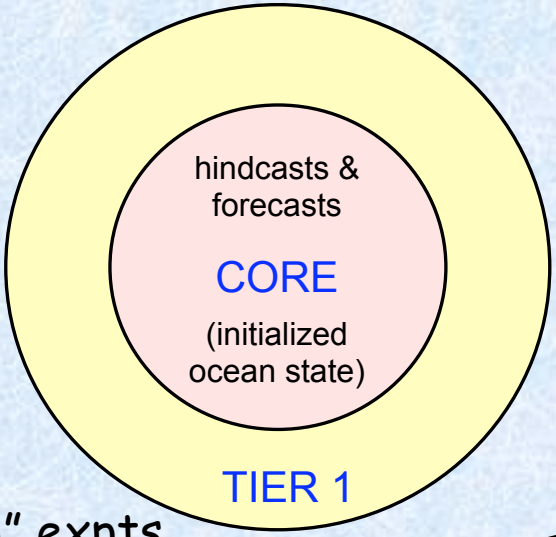
Presented to the EMF Workshop on Climate Change Impacts and Integrated Assessment

Snowmass, CO
1 August 2013

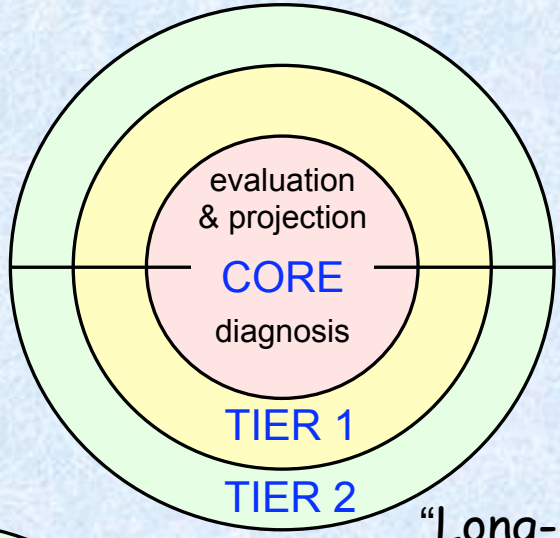
Outline

- How is CMIP model output currently archived and served?
- What challenges face communities less familiar with climate models?
- What can be done to facilitate access to model output by a wider community?

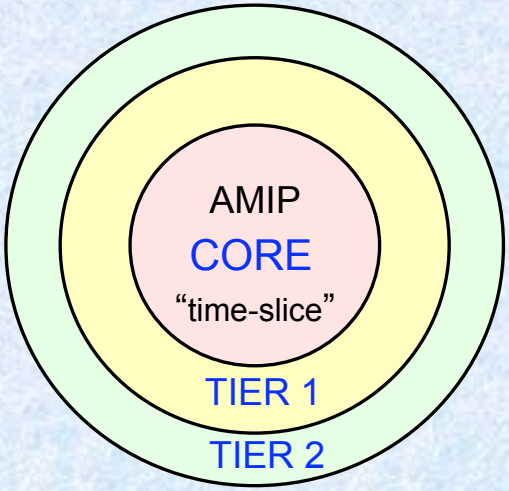
CMIP5 is organized around three types of simulations



"Near-Term" expts.
(decadal prediction)



"Long-Term"
traditional expts.
(century & longer)

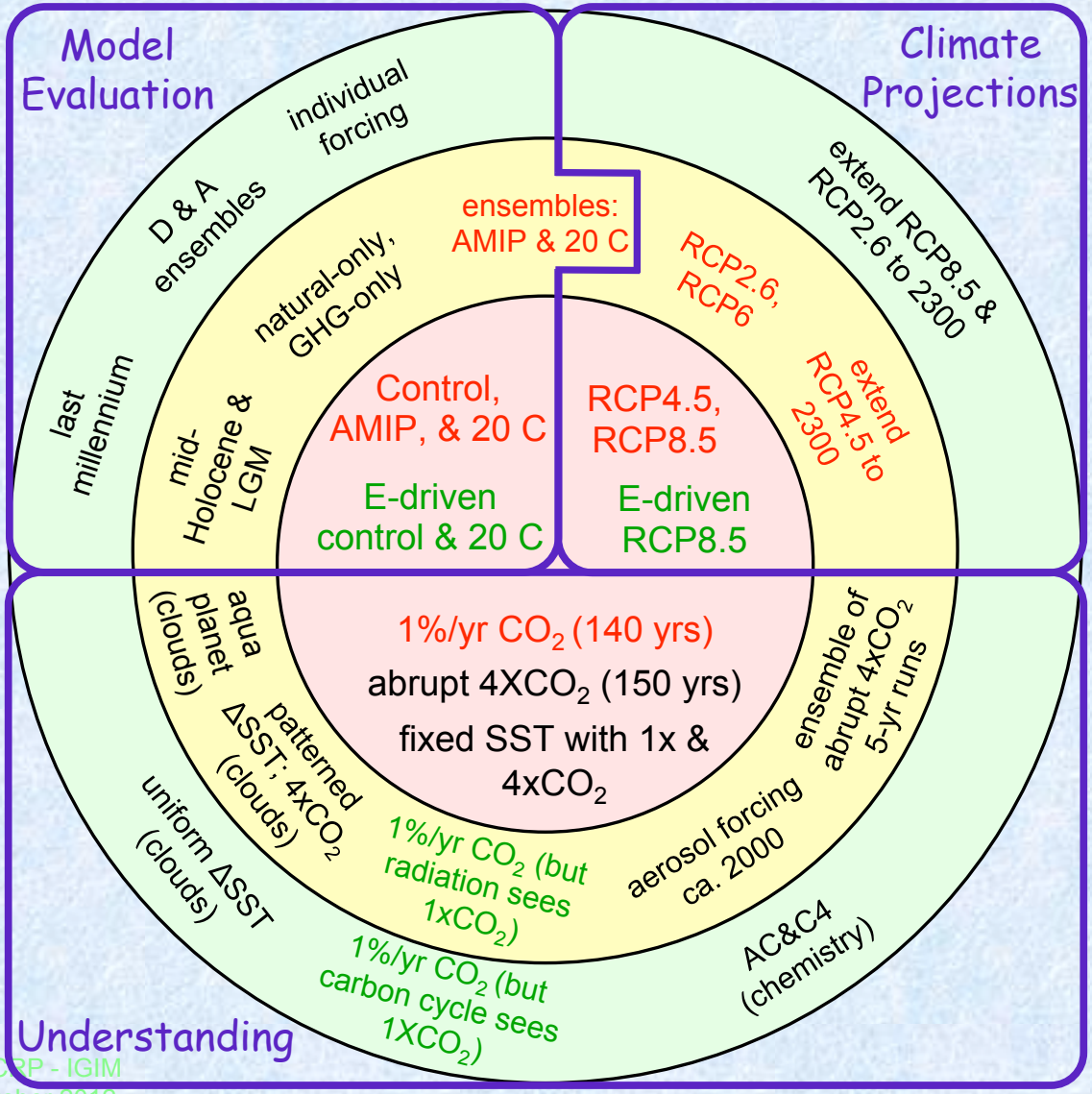


Atmosphere-Only Simulations

(for computationally demanding and NWP models)

See Taylor, Stouffer, & Meehl, *Bull. Amer. Meterol. Soc.*, 2012

A rich set of "long-term" experiments, drawn from several predecessor MIPs, focuses on model evaluation, projections, and understanding



Red subset matches the entire CMIP3 experimental suite

Green subset is for coupled carbon-cycle climate models only

K. E. Taylor
PCMDI

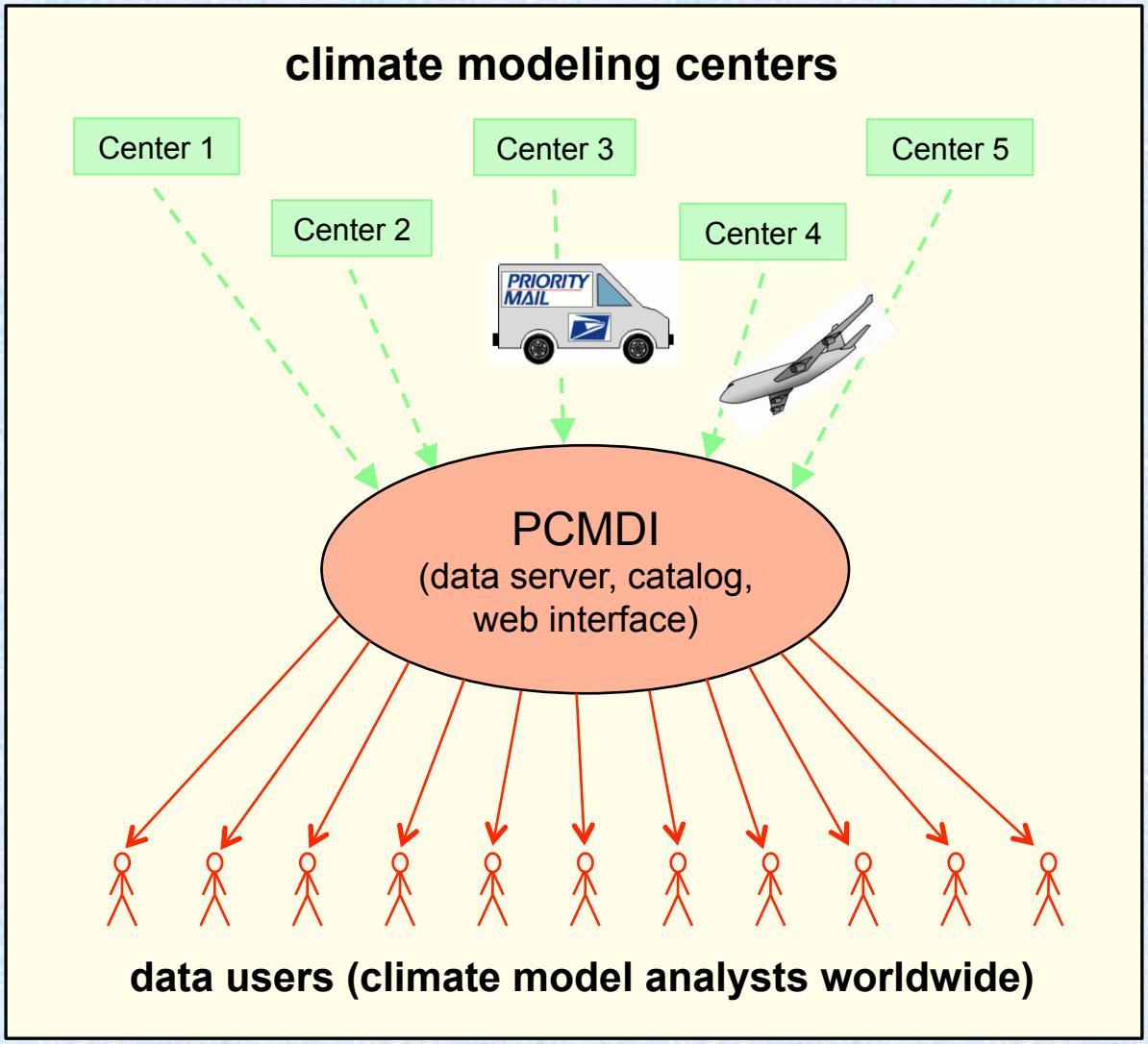
CMIP5 output fields requested (goes well beyond what was available from CMIP3)

- Domains (number of monthly variables*):
  - Atmosphere (60)
  - Aerosols (77)
  - Ocean (69)
  - Ocean biogechemistry (74)
  - Land surface & carbon cycle (58)
  - Sea ice (38)
  - Land ice (14)
  - CFMIP output (~100)
- Temporal sampling (number of variables*)
  - Climatology (22)
  - Annual (57)
  - Monthly (390)
  - Daily (53)
  - 6-hourly (6)
  - 3-hourly (23)

*Not all variables are saved for all experiments and time-periods

http://cmip-pcmdi.llnl.gov/cmip5/output_req.html

CMIP3 data handling: ESG* central archive at PCMDI

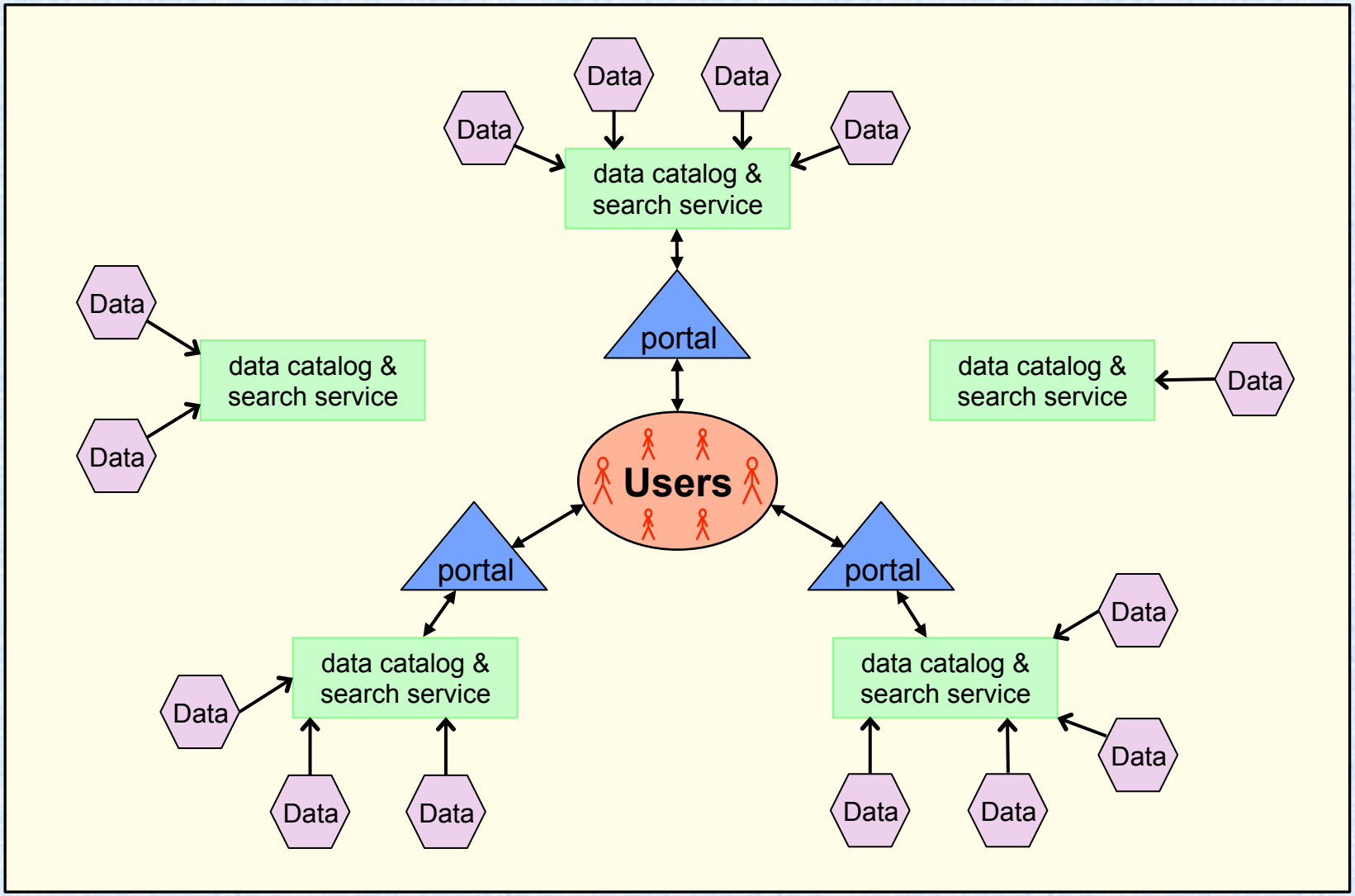


- Data shipped to PCMDI on hard disks
- Delayed availability
- Hindered corrections

- Search service via web gateway
- Download from single location (ftp, http)
- Fragile dependence on a single server.

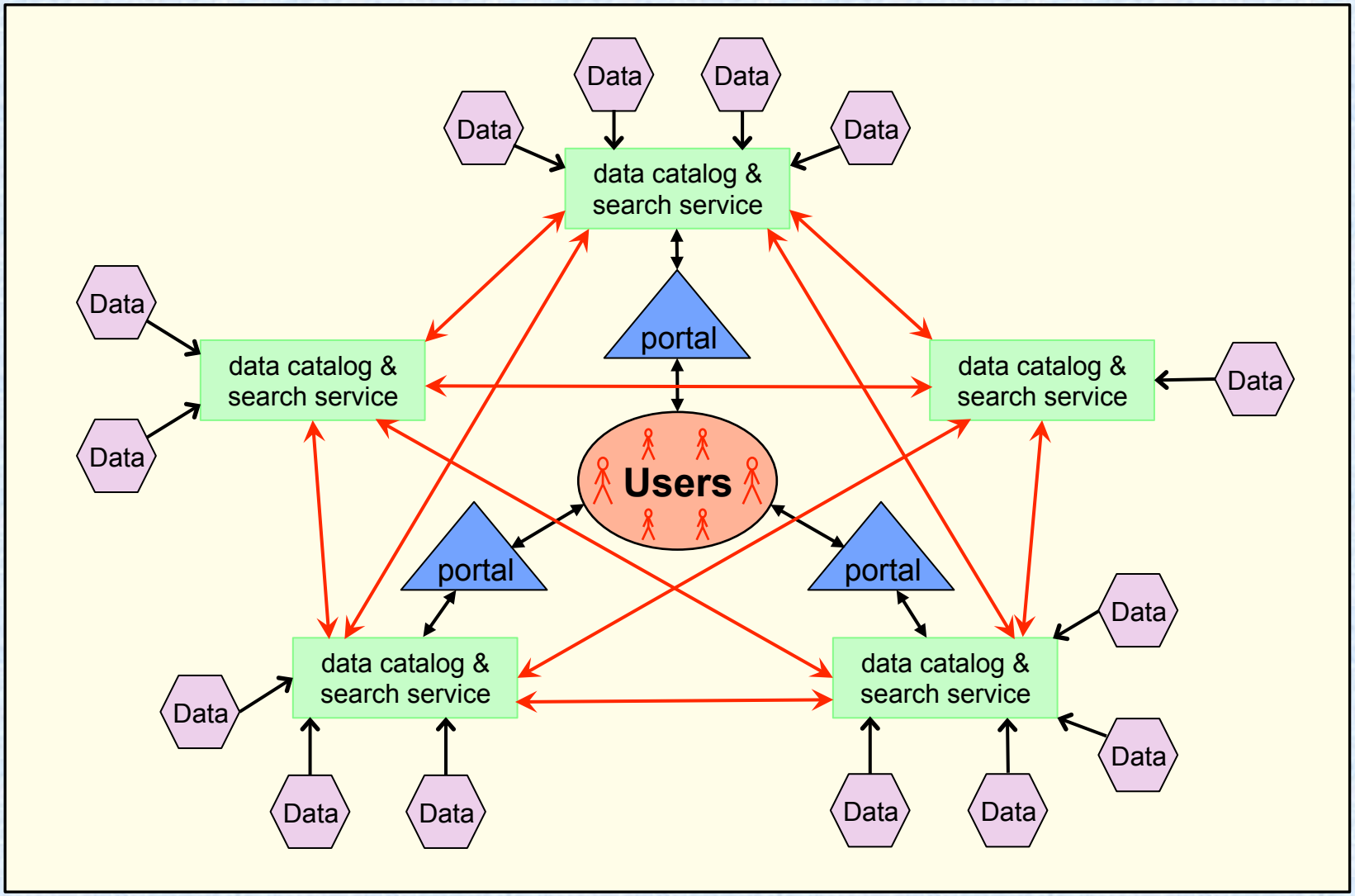
*ESG = Earth System Grid

CMIP5 new approach: Distributed data archive (ESGF*)

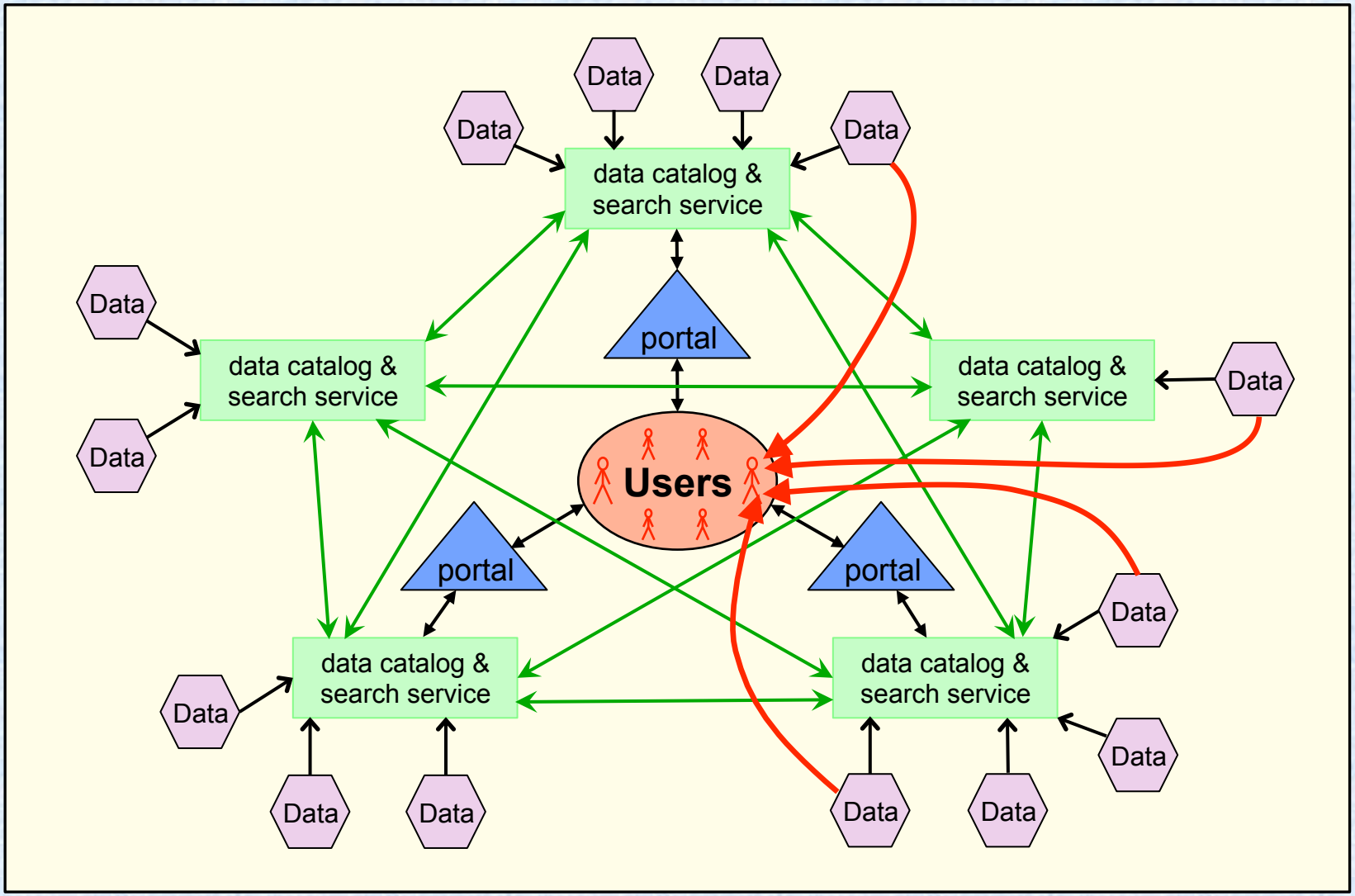


*ESGF = Earth System Grid Federation

All data can be browsed through a single portal because index nodes are federated.

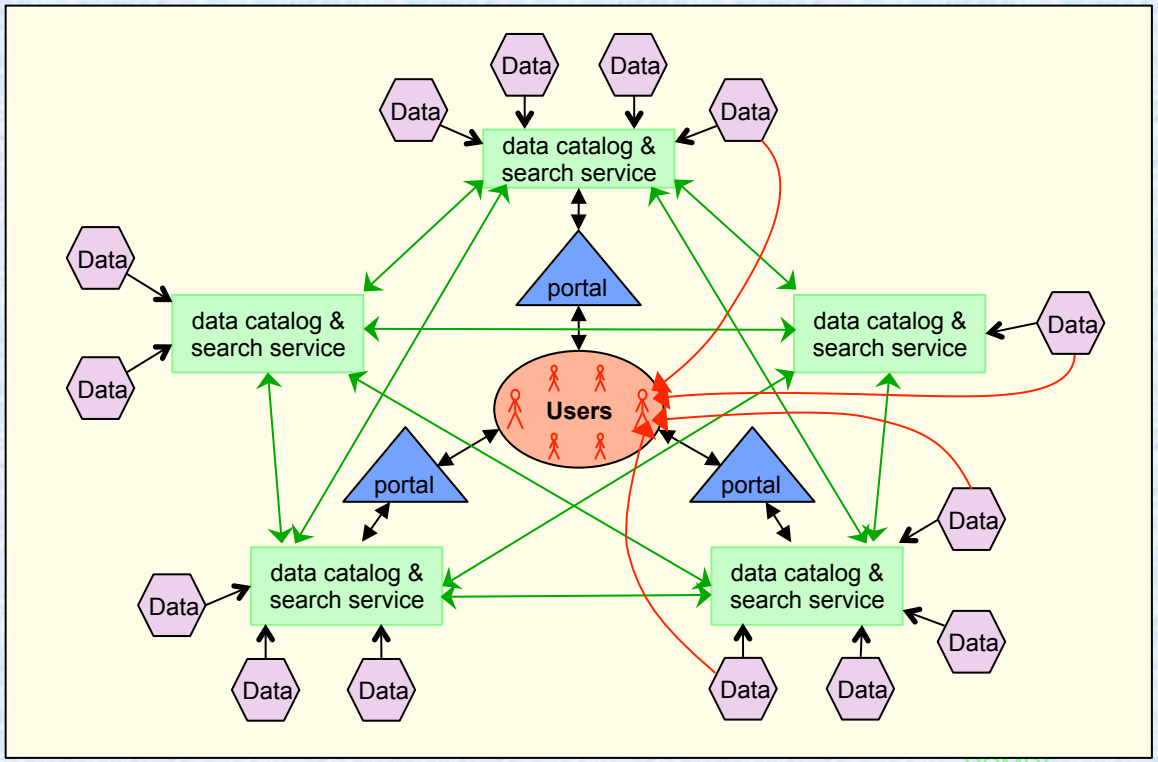


Once desired datasets have been found, user harvests data directly from the nodes.



ESGF is unparalleled in capabilities and complexity

- Diagram does not show:
  - Script-driven direct access route to data (bypassing portal)
  - Server-side computer services
  - Security & authentication layer
- Also:
  - PCMDI and other major data centers have replicated high-demand datasets.



CMIP5 output can be obtained at **http://pcmdi9.llnl.gov**

Users are rewarded with multi-model output conforming to strict standards

- A common set of experiments
- A single "format": netCDF files
- Conforming to CF metadata standards
  - Self-describing data
  - Full grid information
- With specific vocabulary defining
  - Model
  - Experiment name
  - Forcing factors
  - And more

Output from all models can be read and analyzed uniformly

Challenges dealing with CMIP5 data

- A complex array of experiments
- An unusual method for downloading data
- For many, an unfamiliar data format
- Data reported on various grids
- Ensembles of simulations with members differing in sometimes misunderstood ways
- Data reported at a resolution too coarse to be useful (requiring downscaling)
- And more

Not trivial for a novice to find out what data might be useful and to avoid misinterpretations

- What experiment?
  - Which RCP? What are differences?
  - What's the difference between emission-driven and concentration-forced?
- Which model(s)?
  - Are some superior?
  - For models with multiple realizations, which ones should be used?
- How can I correct for biases?
  - Drift in long-term experiments (drift removal)
  - Initial adjustments in decadal experiments (bias correction)

Extending CMIP data to non-specialists

- PCMDI was charged to make model output available to the climate research community.
- PCMDI has partnered with the IPCC's TGICA* to offer data service through its Data Distribution Centre (DDC), which serves communities less familiar with climate models.
- This partnership has successfully opened up model output to users everywhere, but
  - the data remain enigmatic to many
  - The DDC website is slow to evolve (requires cumbersome process of approval)

* TGICA= Task Group on Data and Scenario Support for Impact and Climate Analysis

With more resources additional services could be provided in support of CMIP data

- Tutorials providing guidance on use/limitations of CMIP data
- Better documentation of models, experiments, and methodology for forcing and boundary condition data sets (including the RCP process)
- Derived data products, e.g.,
  - All models on a common grid
  - Multi-model mean (and measure of spread)
  - Drift-corrected data
- Downscaled data
  - Using a variety of approaches
  - With guidance on strengths and limitations of each approach
- What else?

What else could be done?

- Seek more input from users outside the climate research community
- Additional output fields?
- Improvements to CMIP5 and DDC websites
- What else?

"Long-term" experiments: output available, as of (March 3 and) Sept. 22, 2012

Experiment(s)	# of models
* Control & historical	(29) 48
* AMIP	(18) 28
* RCP4.5 & 8.5	(24) 40
RCP2.6	(21) 29
RCP6	(15) 22
RCP's to year 2300	?
* 1% CO2 increase	(21) 32
* Fixed SST CO2 forcing diagnosis	(10) 12
* Abrupt 4XCO2 diagnostic	(20) 30

Experiment(s)	# of models
Fast adjustment diagnostic	?
Aerosol forcing	(6) 9
*ESM cntrl, hist. & RCP8.5	(8) 13
Carbon cycle feedback isolation	(7) 10
Mid-Holocene	(11) 13
LGM	(4) 7
Millenium	(6) 7
CFMIP runs	(6) 10
D & A runs	(16) 17

* *Core simulations*

K. E. Taylor
PCMDI

"Decadal" experiments: Output available, as of (March 3 and) Sept. 22, 2012

** Core simulations simulations*

Experiment(s)	Number of models
*Hindcasts and predictions	(14) 18
AMIP	(7) ?
Volcano-free hindcasts	(1) 2
2010 "Pinatubo-like" eruption	(1) 3
Initialization alternatives	?
Pre-industrial control	(12) ?
1% CO2 increase	(7) ?