# CMIP infrastructure: What's needed?

Karl E. Taylor

Presented to the WGCM's 17th Session
Victoria, Canada
2 October 2013

# Outline

- Current infrastructure overview
  - What's in place and working?
  - What's not and is being worked on?
- Anticipating the future needs
  - How might a new CMIP design framework affect things?
  - How will data volumes and user requirements change?
- Proposal for a WGCM Infrastructure Panel (WIP)

# What's in place? An extensive list of model output, but we need to

- Analyze statistics on usage to prioritize future request:
  - Which variables have been most popular or are essential for addressing important science issues?
  - What time-periods should be sampled at high frequency?
- Reconsider variable tables newly introduced for CMIP5:
  - Should we add/delete fields from the land, surface ice, ocean, biogeochemistry, and aerosol-related tables?
- Coordination of output lists across MIPs
  - uniformity vs. specialization, data volume vs. ease of preparation

# A standardized vocabulary and metadata were defined

- This enables
  - unique identification of output
  - sharing of data across the network.
- The so-called "Data Reference Syntax" (DRS) was agreed upon for
  - Experiment names ("historical", "rcp85", etc.)
  - Output sampling frequency ("mon", "da", etc.)
  - Model names
  - Identification of ensemble members ("rip" designator)
  - Time-range format specification
  - Etc.
- But modifications will be needed, which could be overseen by a WGCM Software Infrastructure Panel

# Output requirement details were specified

- netCDF files ("classic" data model)
- Standardized file names
- Standardized global, variable, and coordinate attributes make the files fully self-describing
- Accommodation of non-lat/lon native grids
- Full description of vertical grid coordinates
- But, any adaptations of this to other MIPs should be overseen by a WGCM Infrastructure Panel

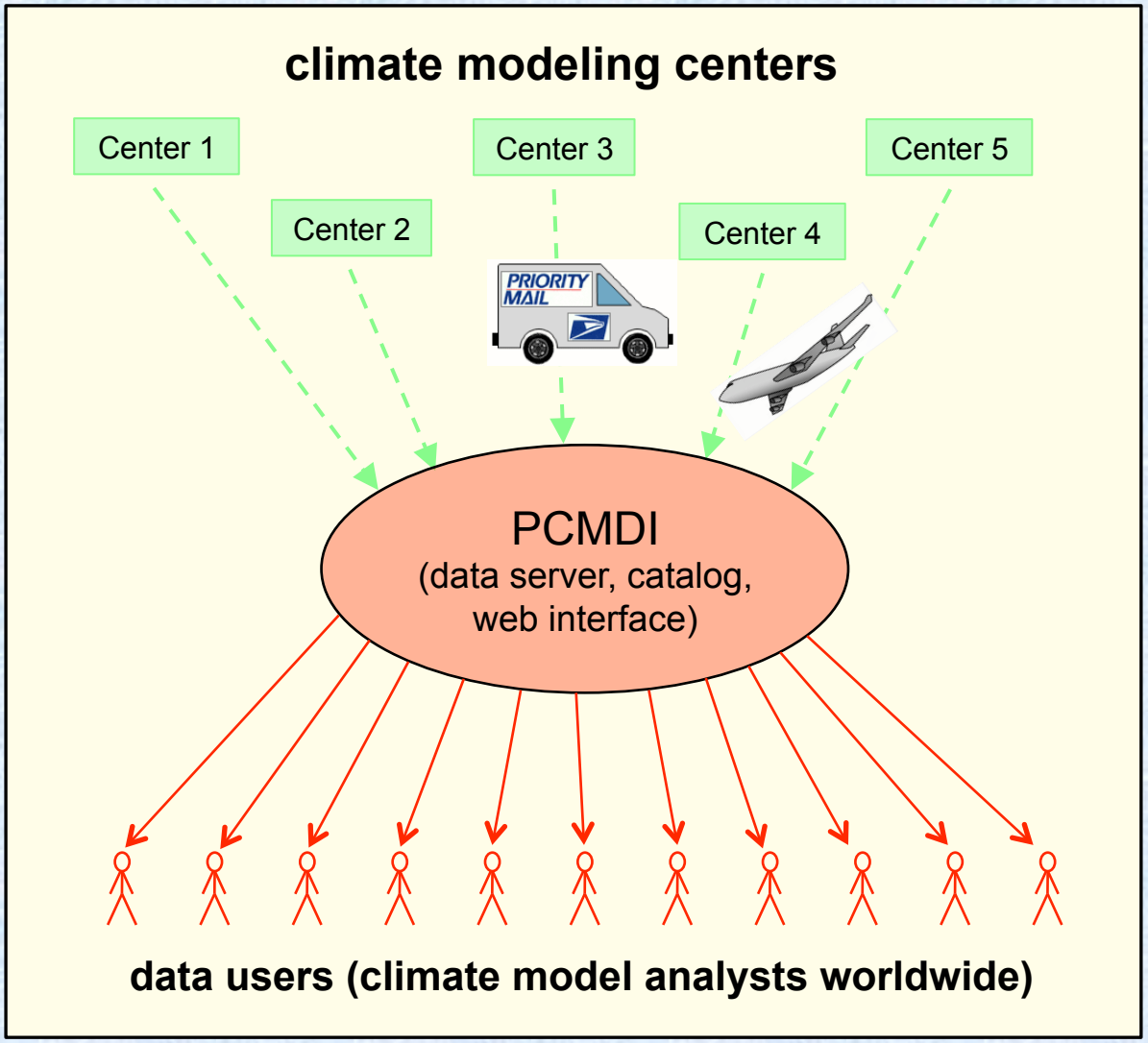
# The CMOR2 code aided in producing model output that met the CMIP5 requirements

- Much of the required metadata was supplied via "CMOR tables"
- Error checks ensured a much "cleaner" data archive
- But although the code is adaptable to other MIPs, additional work is needed to fully generalize it for additional datasets (e.g., obs4MIPs, reanalyses, downscaling)

# The Earth System Grid Federation (ESGF) software matured

- ESGF is a partnership of international collaborators, led by PCMDI.
- ESGF is responsible for providing access to CMIP model output
  - Also other MIPs (e.g., PMIP, CORDEX, geoMIP)
  - And other projects
  - 1.5 petabytes of CMIP5 output now being served.
- ESGF has evolved from a PCMDI hosted archive to a distributed archive.
  - nb. The huge challenge of moving to a distributed archive was not as smooth or rapid as we would have liked.
  - The system has been working well for 12 months now.

# CMIP3 data handling: ESG* central archive at PCMDI

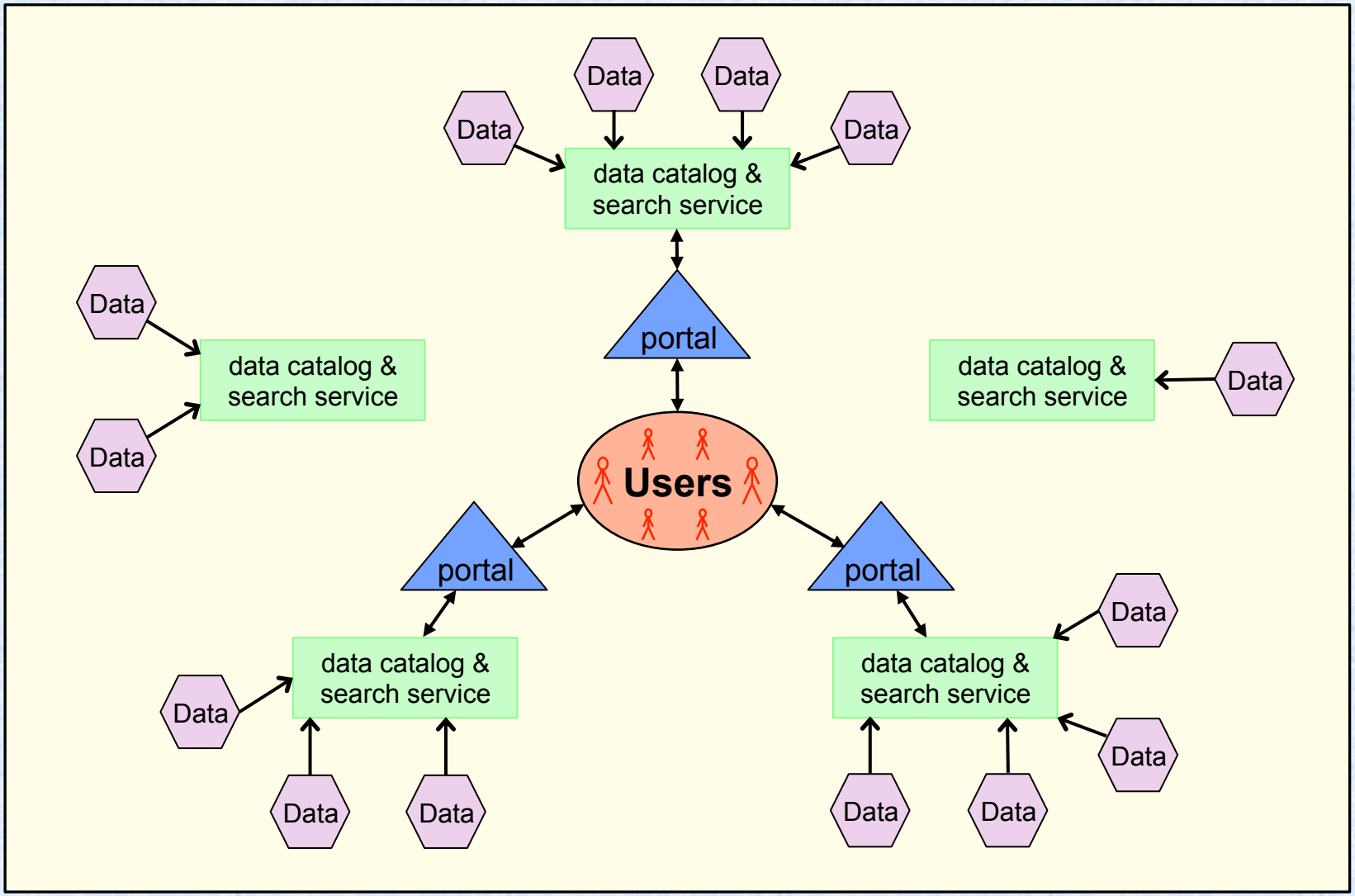



- Data shipped to PCMDI on hard disks
- Delayed availability
- Hindered corrections

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

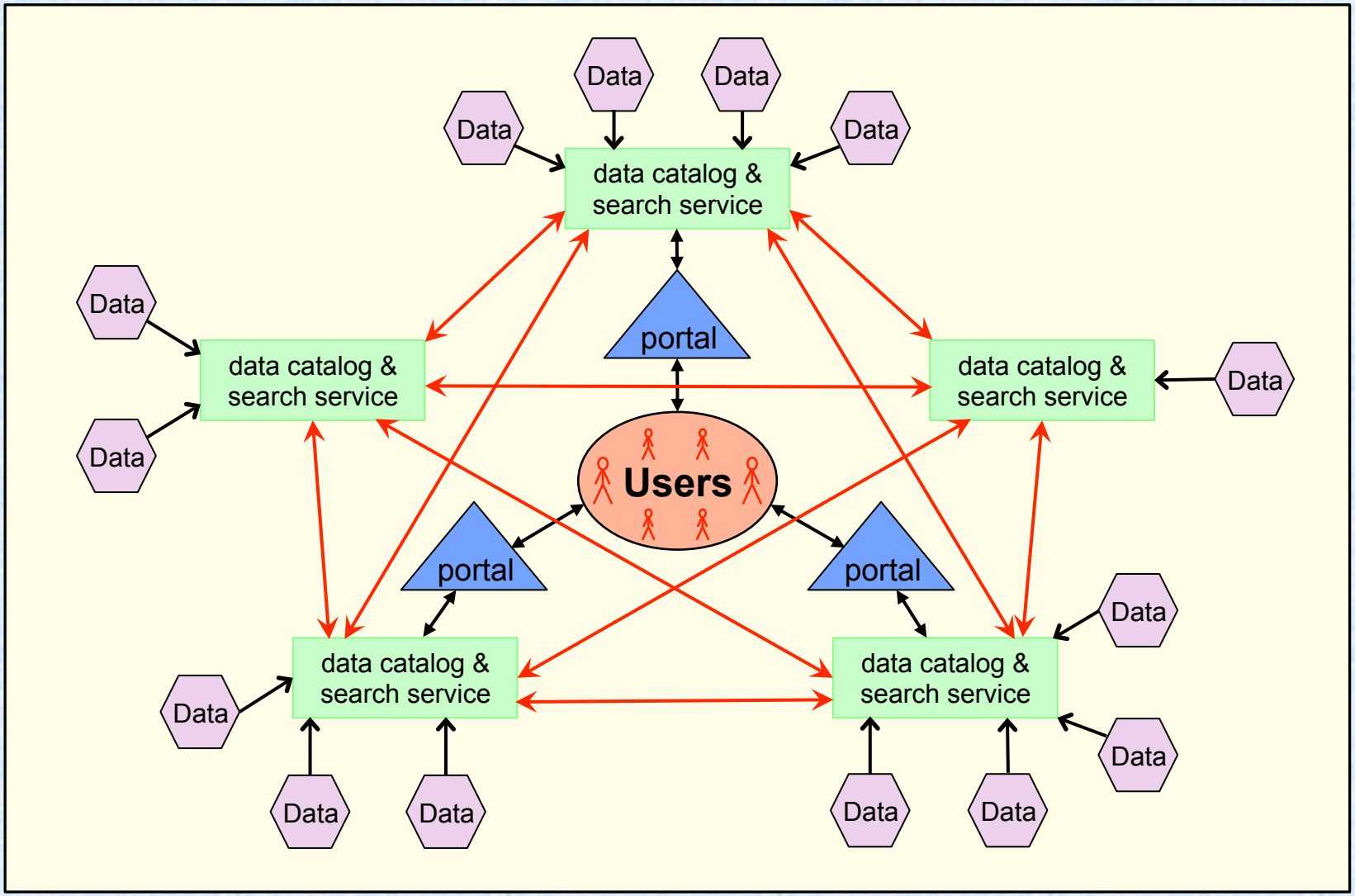
*ESG = Earth System Grid

# CMIP5 new approach: Distributed data archive (ESGF*)

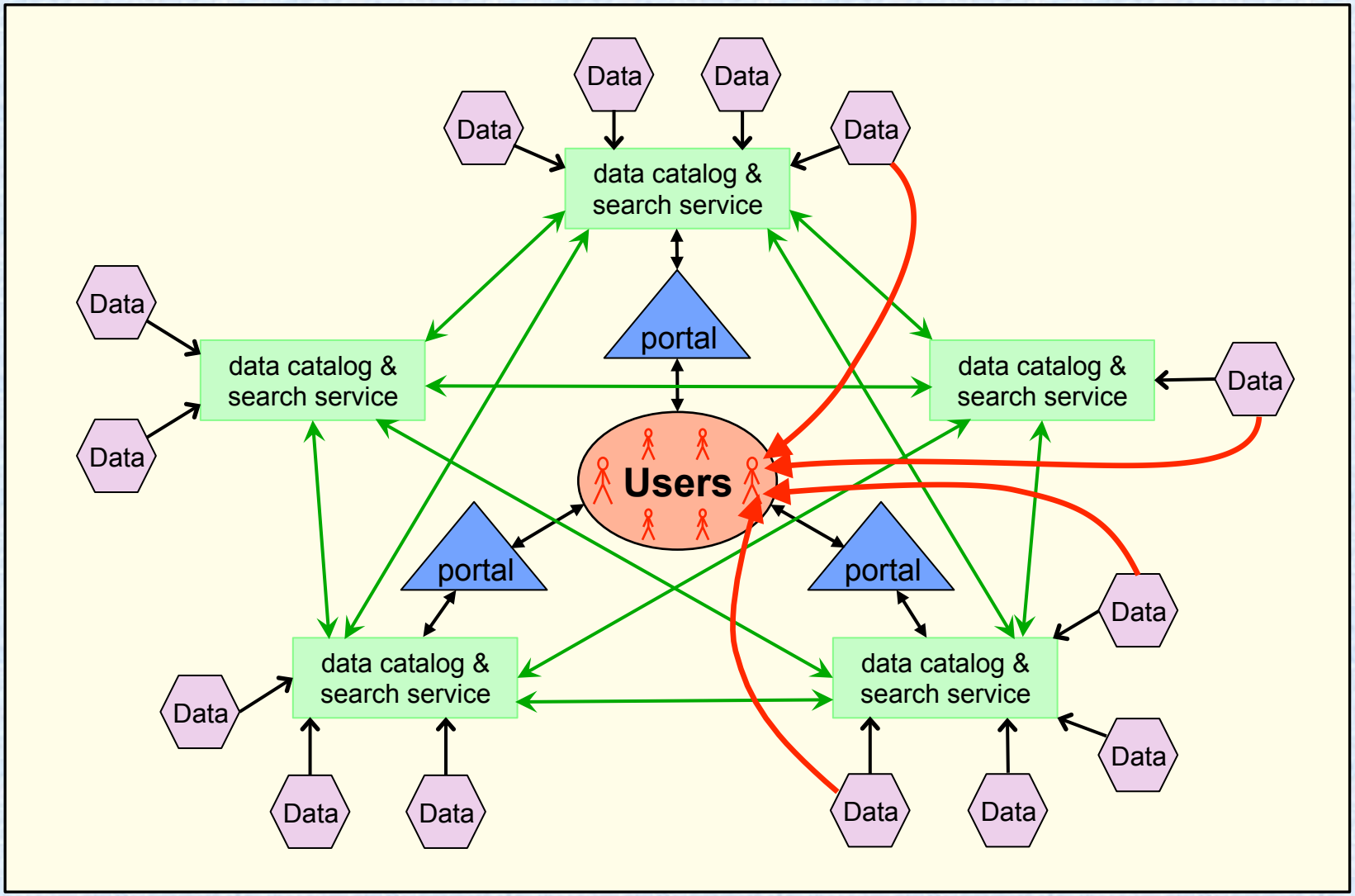



*ESGF = Earth System Grid Federation

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

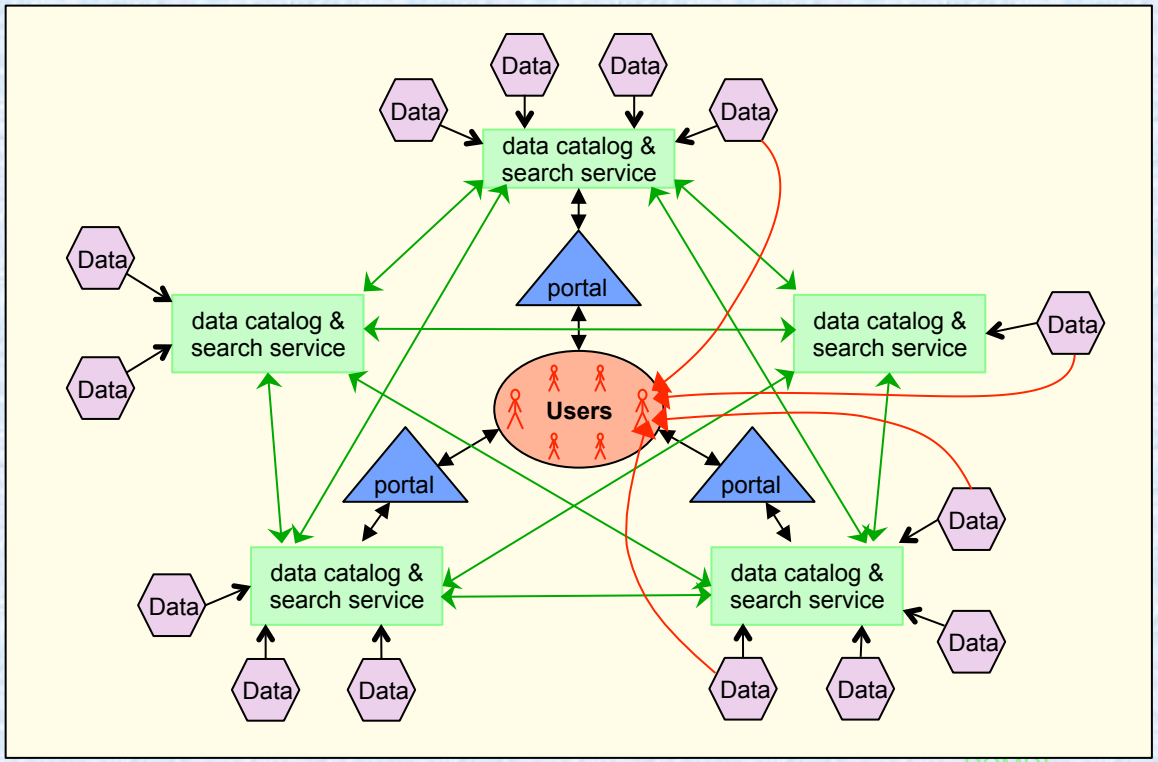


# Through a single data portal, users seek and harvest data directly from multiple data nodes.



# 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**

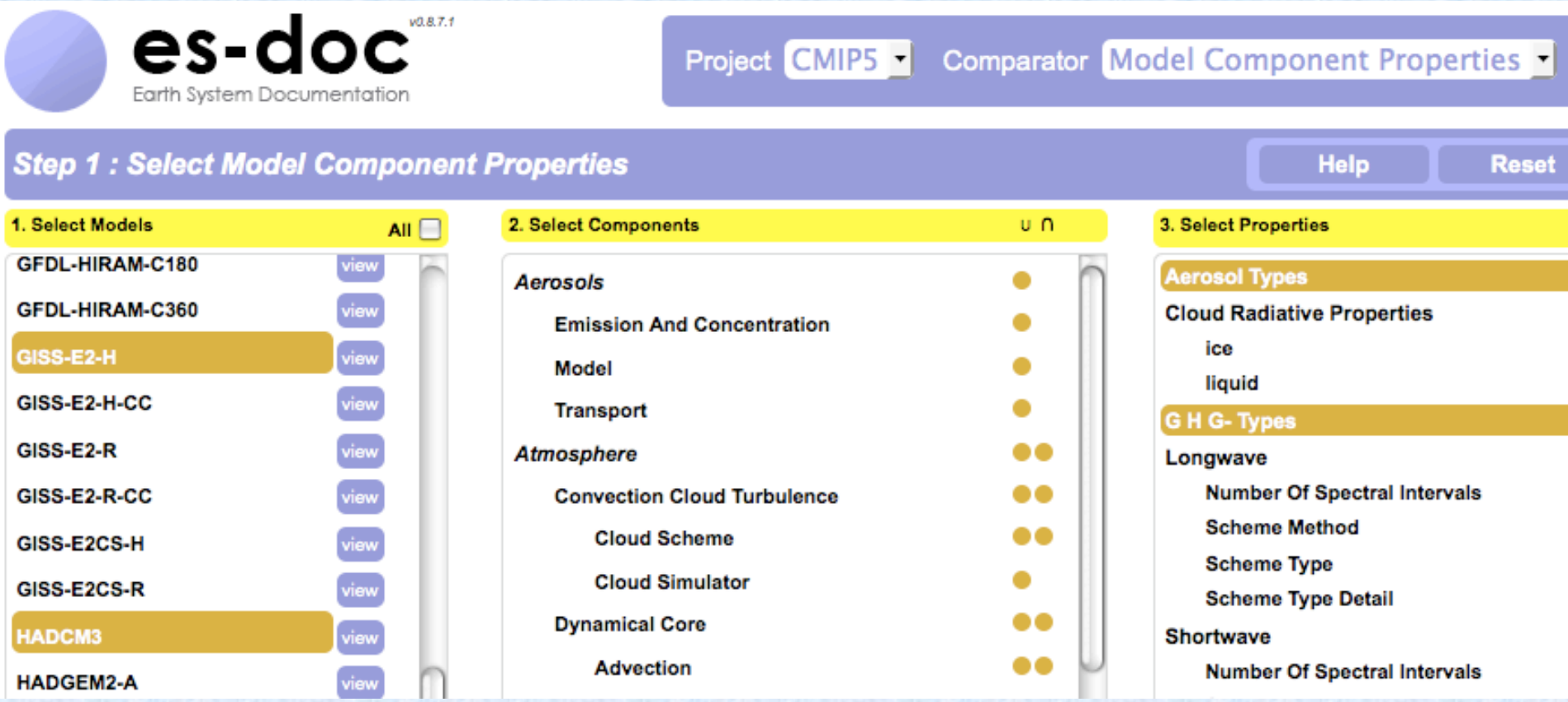
# ESGF will be the cornerstone for the data archive, but

- The proposed ESGF governance board should be formed to
  - Ensure input from modeling groups, major climate data centers, funding agencies, and users guides future development, accounting for all implications.
  - Establish a more formal agreement between the data nodes and the ESGF software managers so that to all parties clearly understand what resources commitments must be made and what procedures need to be followed.
- Many improvements are in the works or envisioned
  - Server-side processing (to reduce data download volume)
  - Automated "replication" (mirroring)
  - Better identification and communication of flawed data
  - More understandable "version" information
  - Alternatives to WGET scripts (e.g, gridftp scripts)

# CMIP5 model documentation

- A first attempt to provide a structured archive of model documentation (METAFOR), but
- Intimidating questionnaire frustrated modeling groups
- Good tools to harvest documentation were unavailable until now
- An ES-DOC collaborative effort promises to be ready from CMIP6.
  - Has recently provided a newly developed tool for accessing CMIP5 model documentation: http://prod.static.esdoc.webfactional.com/js_client/demo/prod/viewer.html
  - Has developed a "comparator" for creating tables indicating differences in model characteristics

# ES-Doc (http://es-doc.org/) comparator tool



# Provenance: An ongoing effort assigns DOI's to CMIP5 datasets for citation (ensuring research reproducibility)

## BUT

- More difficult than anticipated because
  - Datasets continue to evolve (corrections, additions)
  - Challenging QC "requirements" challenged available resources
- Alternatives need to be explored: Perhaps could build on the current CMIP5 "publication" list (CMIP5 website)
  - 350 recorded as of October 1, 2013
  - Information available on
    - Models used
    - Experiments analyzed
    - Variables analyzed
    - Citation information
  - Could expand this to include list of files (tracking i.d.'s) analyzed in the study, providing definitive provenance information ensuring reproducibility

# CMIP Coupled Model Intercomparison Project
## World Climate Research Programme

U.S. DEPARTMENT OF ENERGY | Office of Science

Submit | Edit | View | Administration

## Publications analyzing model: ACCESS1.0

| Author | Article Title | Journal |
|---|---|---|
| Bathols J. , C. Heady, I. G. Watterson | Are climate models more skillful in their home continent?; (Citation) (More Information) | Bulletin of the American Meteorological Society |
| Bracegirdle T. J. , D. B. Stephenson | On the robustness of emergent constraints used in multi-model climate change projections of Arctic warming; (Citation) (More Information) | Journal of Climate |
| Brown J. R. , A. F. Moise, R. A. Colman | The South Pacific Convergence Zone in CMIP5 simulations of historical and future climate; (Citation) (More Information) | Climate Dynamics |
| Brutel-Vuilmet C. , M. Menegoz, G. Krinner | An analysis of present and future seasonal Northern Hemisphere land snow cover simulated by CMIP5 coupled climate models; (Citation) (More Information) | The Cryosphere |
| Ceppi P. , Y. Hwang, D. M. Frierson, D. L. Hartmann | Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing; (Citation) (More Information) | Geophysical Research Letters |
| Collier M. A. , L. D. Rotstayn, J. Kim, K. Kim | An assessment of central and eastern Pacific El Nino's in the CSIRO-Mk3.6, ACCESS1.0 and ACCESS1.3 CMIP5 coupled climate models and their impact on Australian Rainfall; (Citation) (More Information) | Journal of Climate |
| Dirmeyer P. A. , Y. Jin, B. Singh, X. Yan | Trends in land-atmosphere interactions from CMIP5 simulations; (Citation) (More Information) | Journal of Hydrometeorology |
| Dirmeyer P. A. , Y. Jin, B. Singh, X. Yan | Evolving land-atmosphere interactions over North America from CMIP5 simulations; (Citation) (More Information) | Journal of Climate |
| Du Y. , X. Shang-Ping, Y. Ya-Li, X. Zheng, L. Liu, ... | Indian Ocean variability in the CMIP5 multi-model ensemble: The basin mode; (Citation) (More Information) | Journal of Climate |
| DU Z. , R. HUANG, G. Huang | How Well can CMIP5 CGCMs Simulate the EAP/PJ Teleconnection Pattern and its Corresponding Summer Climate in the East Asian Monsoon Region; (Citation) (More Information) | Other |
| DU Z. , R. HUANG, G. Huang | How well can CMIP5 CGCMs simulate the Asian summer monsoon rainfall and its interannual variability and their future projections; (Citation) (More Information) | Advances in Atmospheric Sciences |
| Fettweis X. , B. Franco, M. | Estimating Greenland ice sheet surface mass balance contribution to future sea level rise using the regional atmospheric climate model MAR; (Citation) | The Cryosphere Discuss |

Total Publications Count: 350

| Model | Count |
|---|---|
| ACCESS1.0 | 45 |
| ACCESS1.3 | 28 |
| BCC-CSM1.1 | 94 |
| BCC-CSM1.1-m | 17 |
| BESM-OA2.3 | 6 |
| BNU-ESM | 23 |
| CanAM4 | 23 |
| CanCM4 | 34 |
| CanESM2 | 126 |
| CCSM4 | 104 |
| CCSM4-RSMAS | 13 |
| CESM-BGC | 20 |
| CESM1-CAM5 | 24 |
| CESM1-CAM5.1.FV2 | 15 |
| CESM1-FASTCHEM | 17 |
| CESM1-WACCM | 18 |
| CFSv2-2011 | 12 |
| CMCC-CESM | 15 |
| CMCC-CM | 25 |
| CMCC-CMS | 17 |
| CNRM-CM5 | 120 |

# Full information available on what CMIP5 data was used in each study (http://cmip.llnl.gov/cmip5/publications/allpublications)

Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models; (Citation)

**Andrews T. , J. M. Gregory M. J. Webb K. E. Taylor null : " Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models" , *Geophysical Research Letters* 39 , doi:10.1029/2012GL051607 , http://www.agu.org/pubs/crossref/2012/2012GL051607.shtml**

(More Information)

| Experiments | Models | Variables | Keywords |
|---|---|---|---|
| abrupt4xCO2 | CanESM2 | land area fraction | WG1 (physical climate system) |
| piControl | CNRM-CM5 | surface temperature | Abrupt change |
| sstClim | CSIRO-Mk3.6.0 | toa incoming shortwave flux | Globe |
| sstClim4xCO2 | GFDL-CM3 | toa outgoing longwave flux | Energy budget |
| | GFDL-ESM2G | toa outgoing longwave flux assuming clear sky | Radiatiave forcing |
| | GFDL-ESM2M | toa outgoing shortwave flux | Clouds |
| | HadGEM2-ES | toa outgoing shortwave flux assuming clear sky | Radiation |
| | INM-CM4 | | Feedbacks |
| | IPSL-CM5A-LR | | Climate sensitivity |
| | MIROC-ESM | | |
| | MIROC5 | | |
| | MPI-ESM-LR | | |
| | MPI-ESM-P | | |
| | MRI-CGCM3 | | |
| | NorESM1-M | | |

# What can WGCM do? Establish a WCRP Infrastructure Panel (WIP)

- Oversee standards that guarantee that users and different data distribution centers can discover, browse, catalog, archive and share climate datasets.

- What standards?
  - CF metadata standards
  - Specifications beyond CF guaranteeing fully self-describing and easy-to-use datasets (e.g., CMIP requirements for output)
  - Catalog and software interface standards ensuring remote access to data, independent of local format (e.g., OPeNDAP, THREDDS)
  - Node management and data publication protocols
  - Defined dataset description schemes and controlled vocabularies (e.g., the CMIP5 "Data Reference Syntax:")
  - Standards governing model and experiment documentation (e.g., CIM)

# Why not carry on as in the past?

- Heavy responsibility placed on a few individuals
  - Worked O.K. for CMIP5, but may fail as purview expands across WGCM projects
- A panel drawn from a broader group of experts may more nimbly respond to future needs
- Modeling groups invest considerable resources meeting the MIP requirements.
  - Anything done to ensure that standards are as unform as possible across all MIPs will reduce the burden.
- Membership on an official panel might help individual members to fund their work in this area.

# WGCM Infrastructure Panel: next steps

- WGCM approves concept
- Terms of reference for the panel are specified and a list of nominees is prepared by authors of the proposal
- The WGCM approves the terms of reference and appoints panel members.