# CMIP Data Handling: Past & Future

Karl E. Taylor

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

Presented to the AGCI CMIP6 Workshop
Aspen, CO
6 August 2013

# Scope

- Preparation of model output
- ESGF: Data archive and access
- Model documentation
- Provenance: assuring reproducibility and crediting modeling groups

# Preparation of model output

- Requested output list
  - Drew from previous CMIP lists (especially atmos.)
  - Relied on experts from various communities to augment original list
    - WGOMD: ocean
    - CFMIP: special satellite simulator diagnostics
    - Experts on biogeochemistry, carbon cycle, land surface, aerosols, cryosphere
    - TGICA: output of interest to impacts community
- Standardized description (defined vocabulary)
  - Data reference syntax (DRS)
  - Strict specifications of file attributes providing metadata needed to interpret model output fields.
- Conforming output to requirements (CMOR)

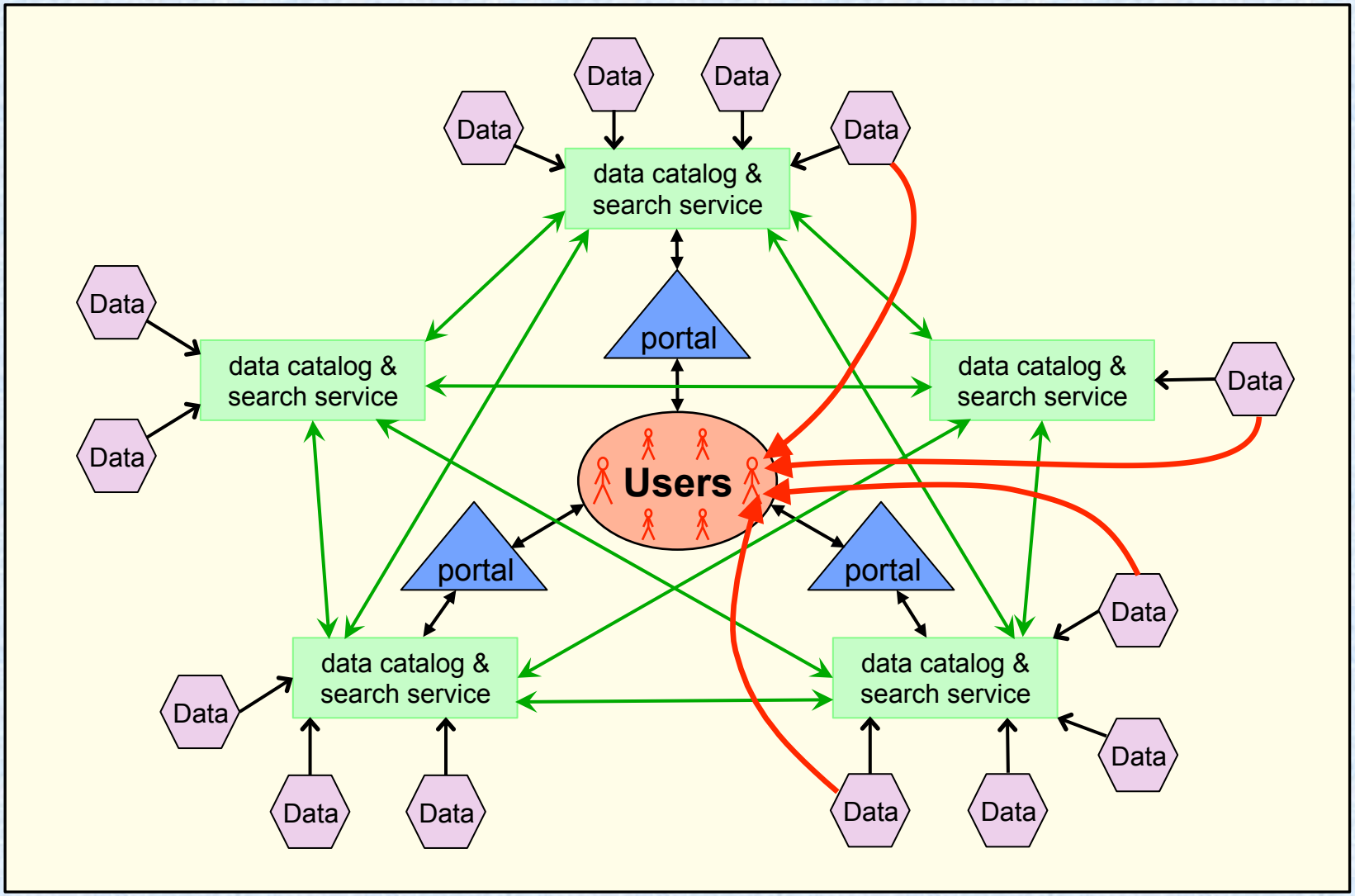
# Preparation of model output: next steps

- Improve description of model forcing
- Reduce volume of output requested???
- Seek further input on need for additional fields, especially from IAV and IA modeling communities

# ESGF: Data archive and access

- First attempt to federate multiple data nodes to provide uniform access to a distributed data archive.
- Modeling centers host their own data
  - No delay in making output available to users
  - Flawed data can be locally be corrected
- 1.5 petabytes now being served

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



# ESGF: Data archive and access

- First attempt to federate multiple data nodes to provide uniform access to a distributed data archive.
- Modeling centers host their own data
  - No delay in making output available to users
  - Flawed data can be locally be corrected
- 1.5 petabytes now being served

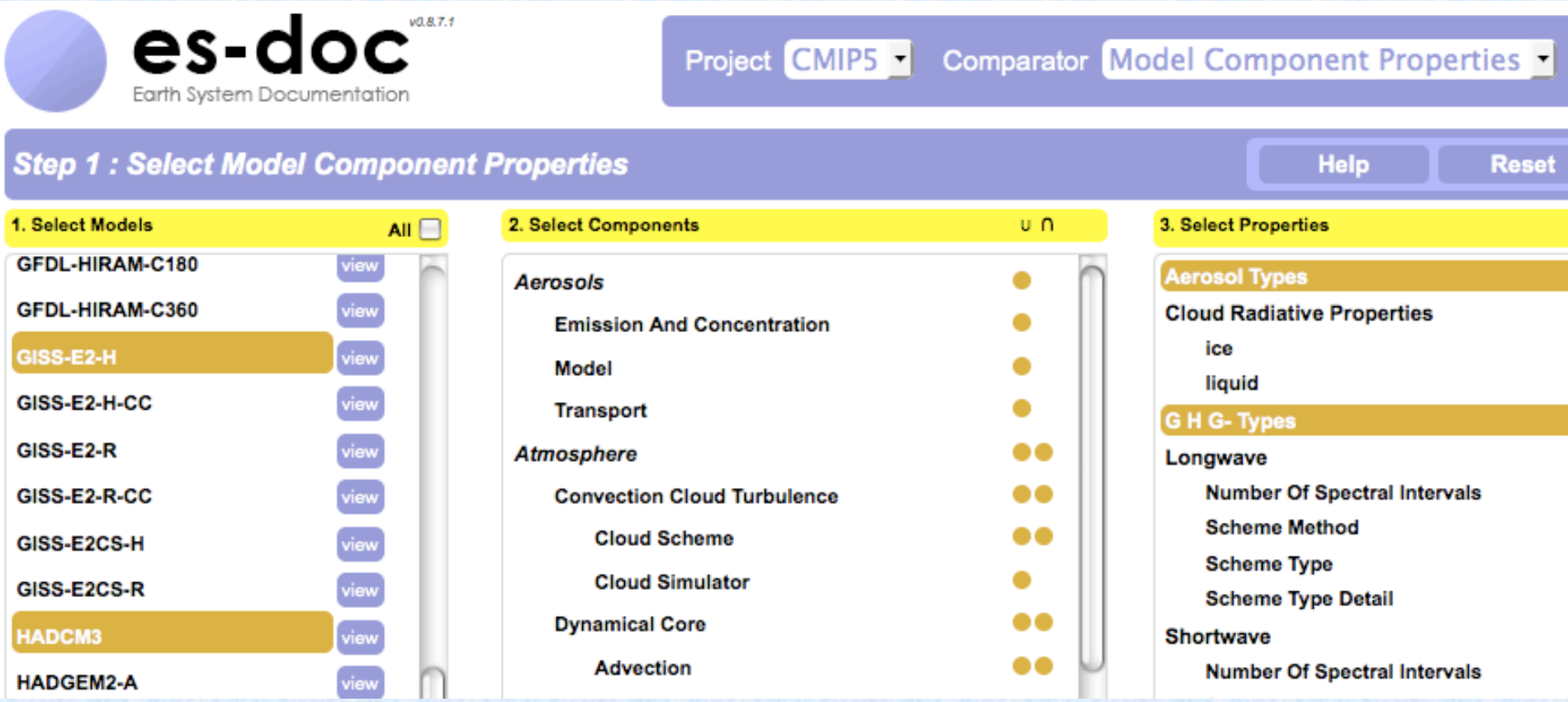
# ESGF: Problems and future

- Initially (before 9/12) the complexity of the distributed data archive caused extreme frustration for those downloading the data and delayed analysis
- ESGF has now been operating smoothly for 10 months
- Replication (mirroring) of data at a few major data centers ensures resiliency
- Improvements needed in error reporting/notification service to provides updates on flawed data and availability of new data of interest
- Implementation of server-side data processing to reduce volume of data downloaded

# Model documentation

- First attempt to provide a structured archive of model documentation (METAFOR)
- Intimidating questionnaire filled in by modeling groups
- Until now, users have had difficulties accessing the information.
- An ES-DOC collaborative effort promises to be ready from CMIP6.
  - Has recently provided a newly developed tool to provide users with 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: assuring reproducibility and crediting modeling groups

- Intent to assign DOI's to CMIP5 datasets
  - Datasets continue to evolve (corrections, additions)
  - Formal procedure has been established, but application is lagging
- Alternative: Web form enabling researchers to record publications based on CMIP5 output.
  - 322 recorded as of August 1, 2013
  - Information available on
    - Models used
    - Experiments analyzed
    - Variables analyzed
    - Citation information
- Could expand publication record to include list of files (tracking i.d.'s) analyzed in the study
  - Provide definitive provenance information ensuring reproducibility

# 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 | | |

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

U.S. DEPARTMENT OF ENERGY | Office of Science

Submit | Edit | View | Administration

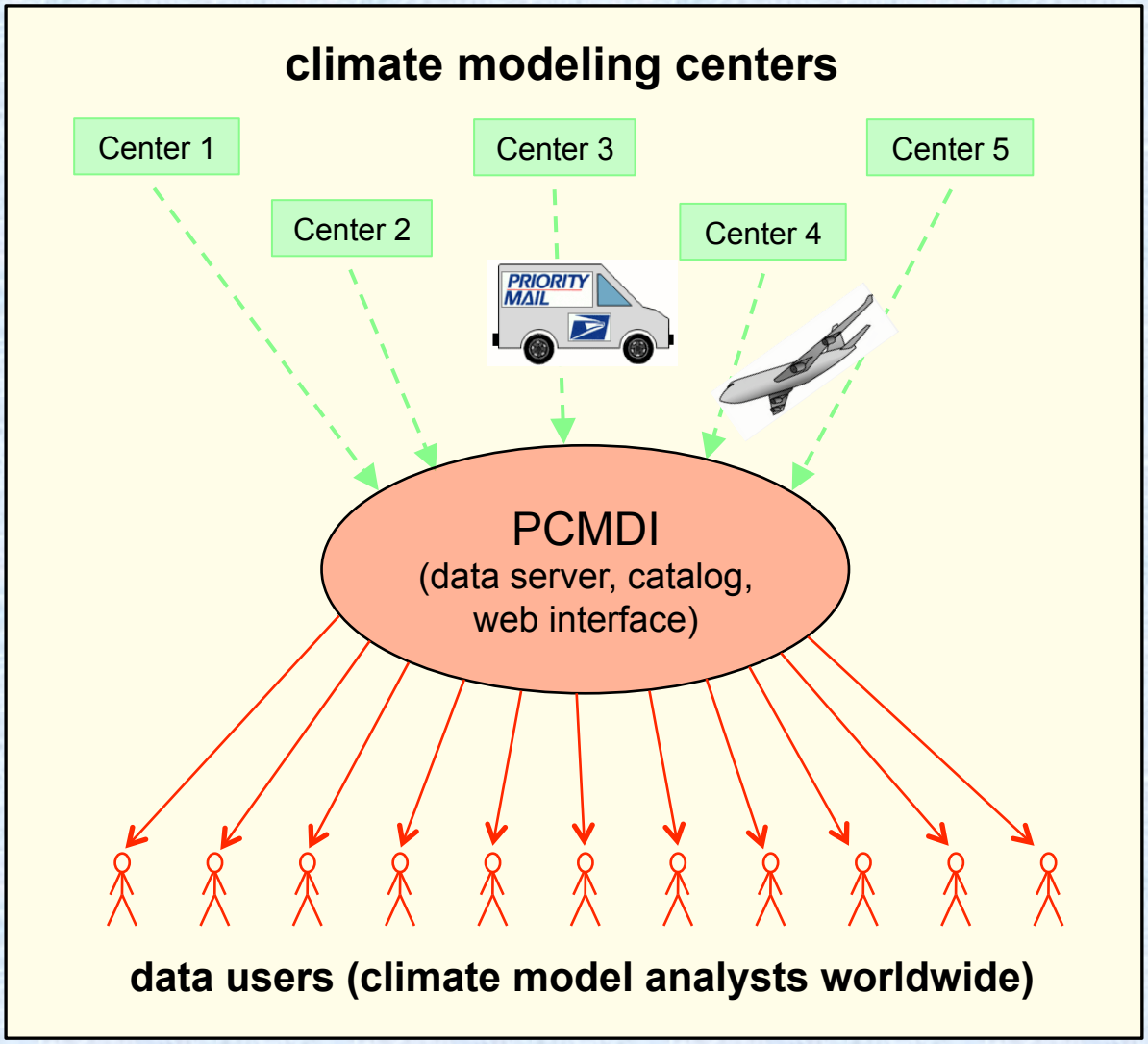
## Publications analyzing model: ACCESS1.0

| 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: 248

| 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 |

# CMIP3 data handling: ESG* central archive at PCMDI

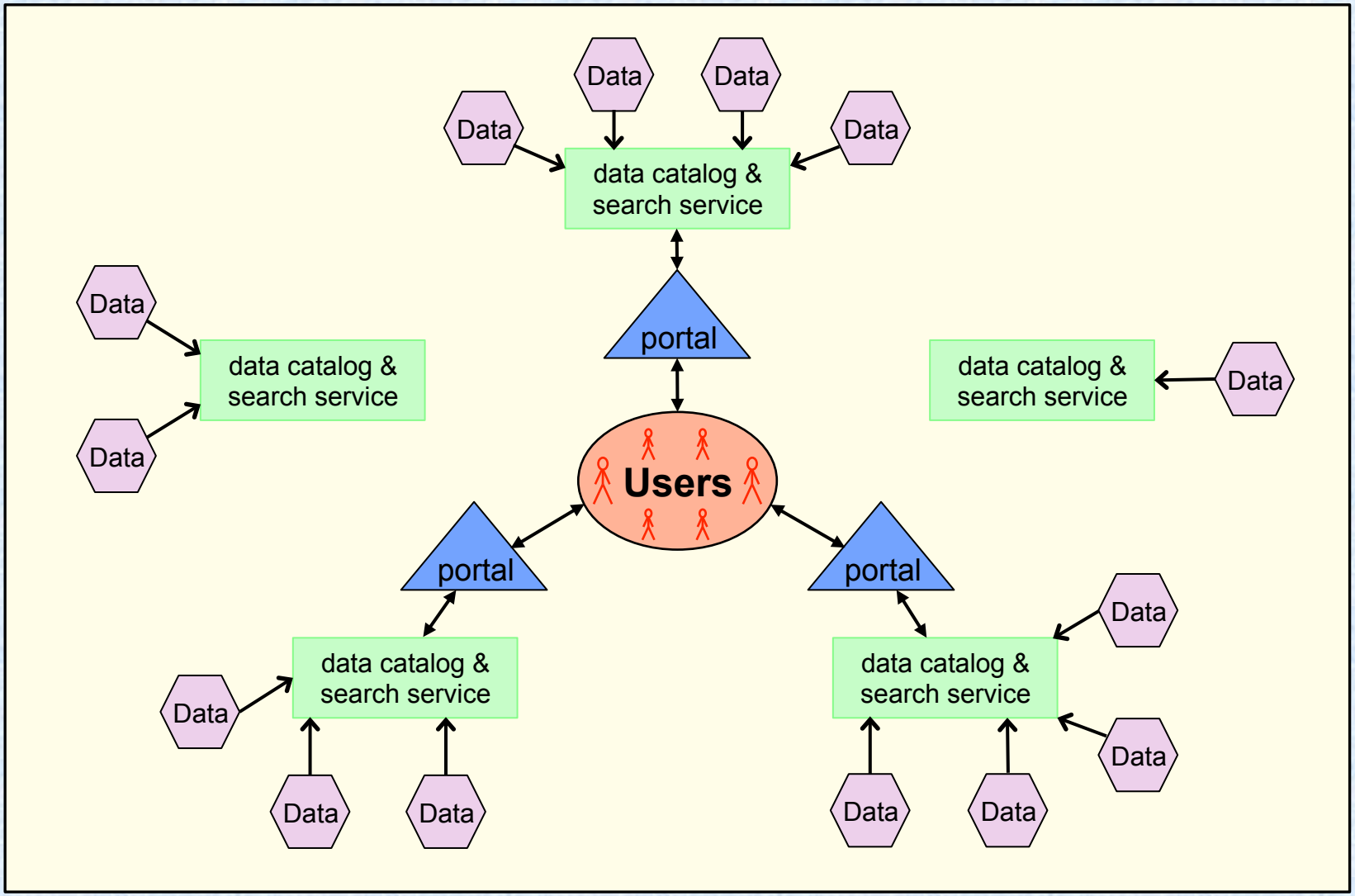



- Data shipped to PCMDI on hard disks
- Delayed availability
- Hindered corrections

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

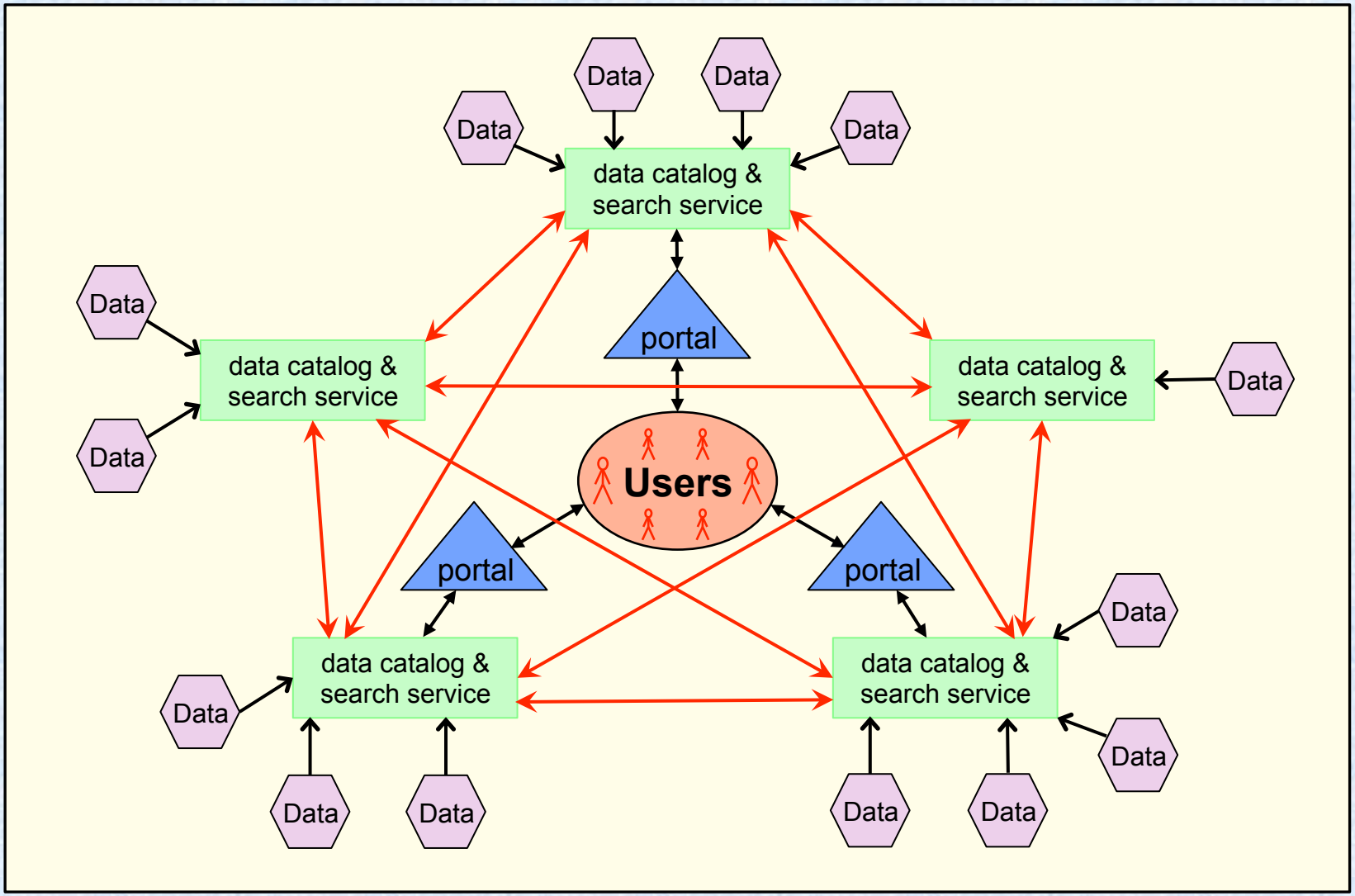
*ESG = Earth System Grid

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



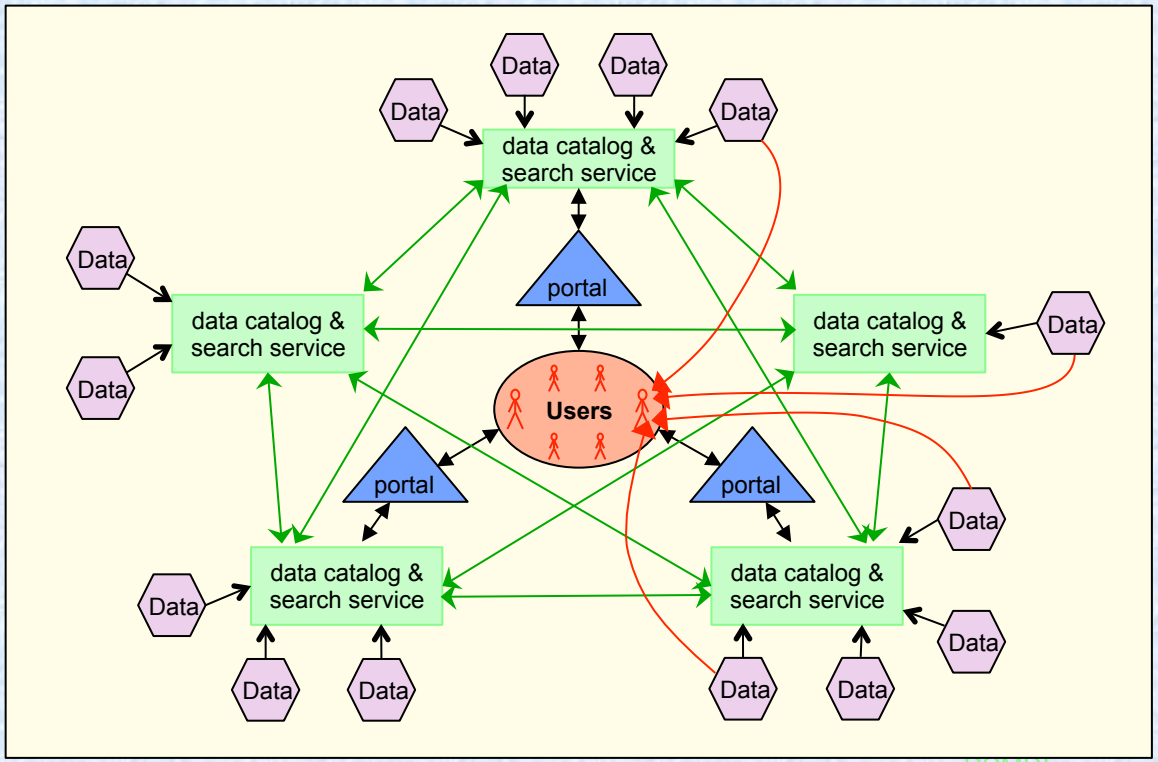

*ESGF = Earth System Grid Federation

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



# 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**