



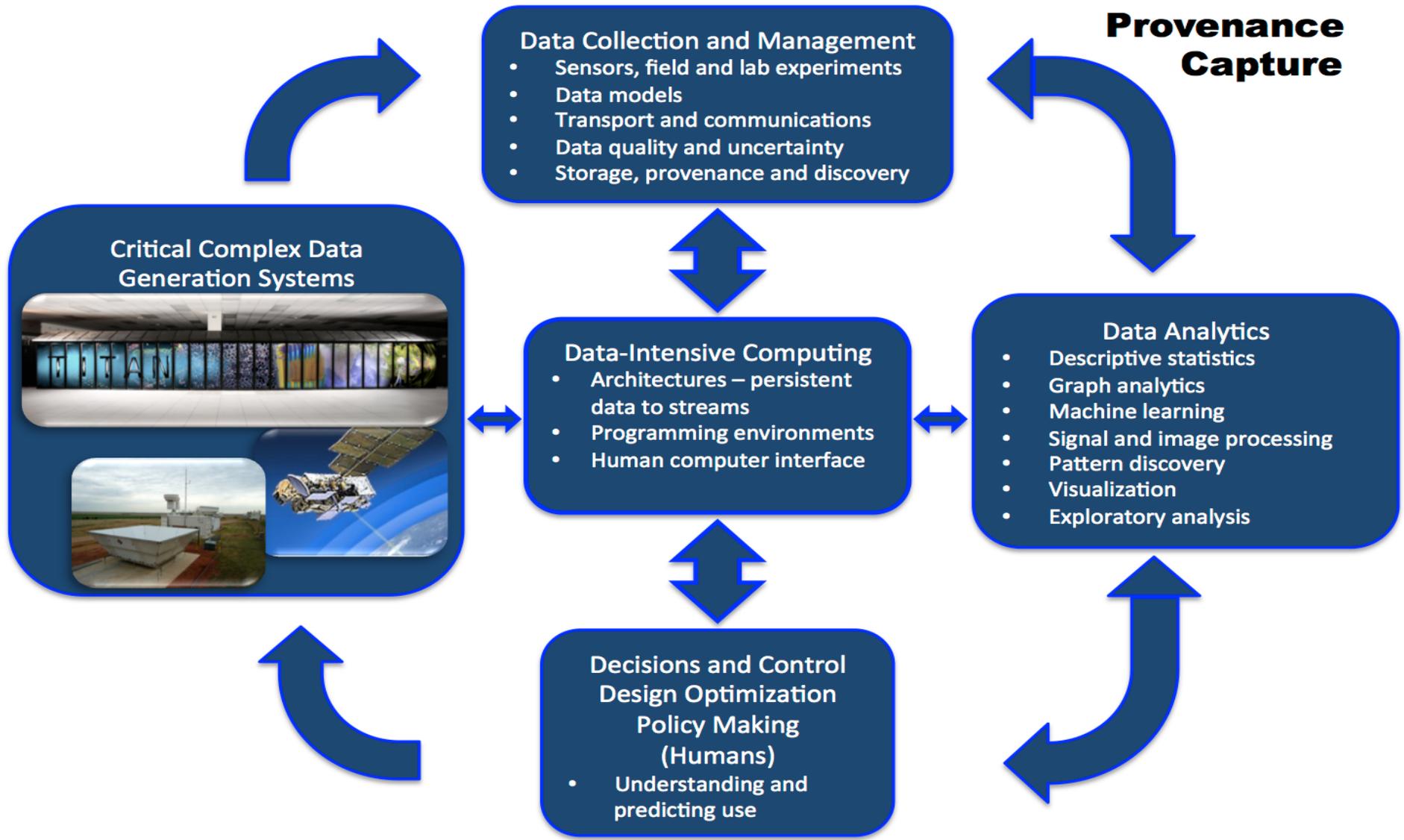
Earth System Grid Federation: Data Discovery and Workflows

Dean N. Williams

On behalf of the Climate Community

**Intelligent Optical Network Infrastructure Workshop
August 05, 2014**

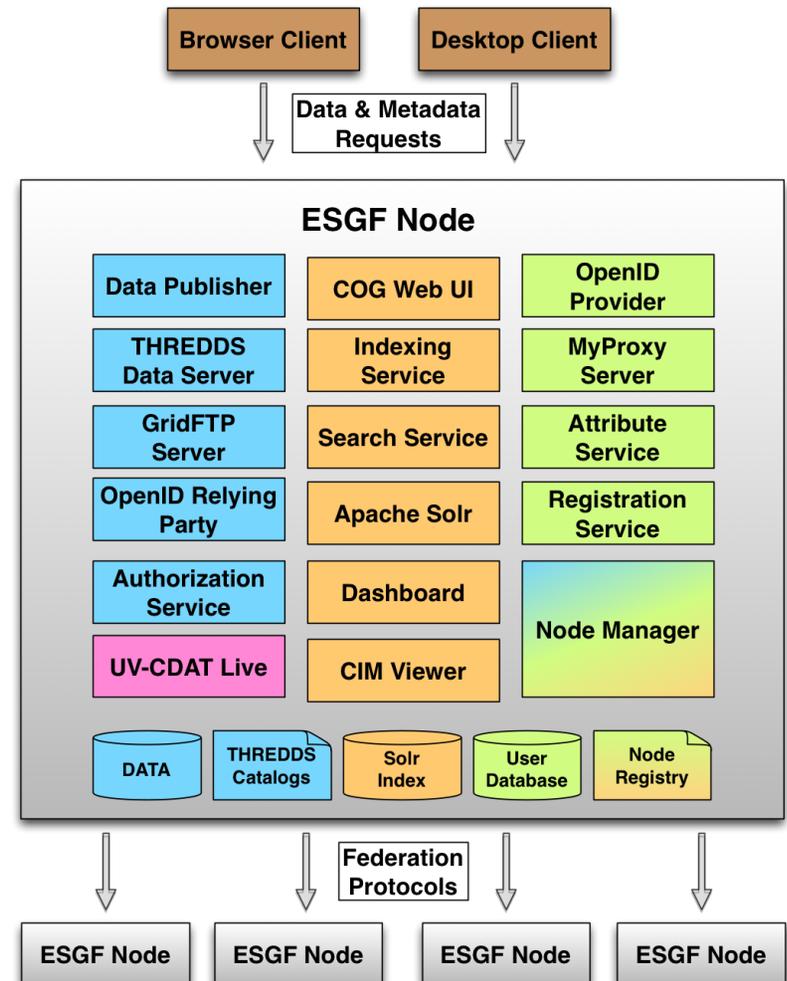
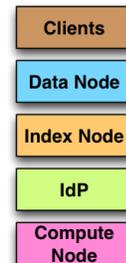
Integrated Data Ecosystem



Earth System Grid Federation

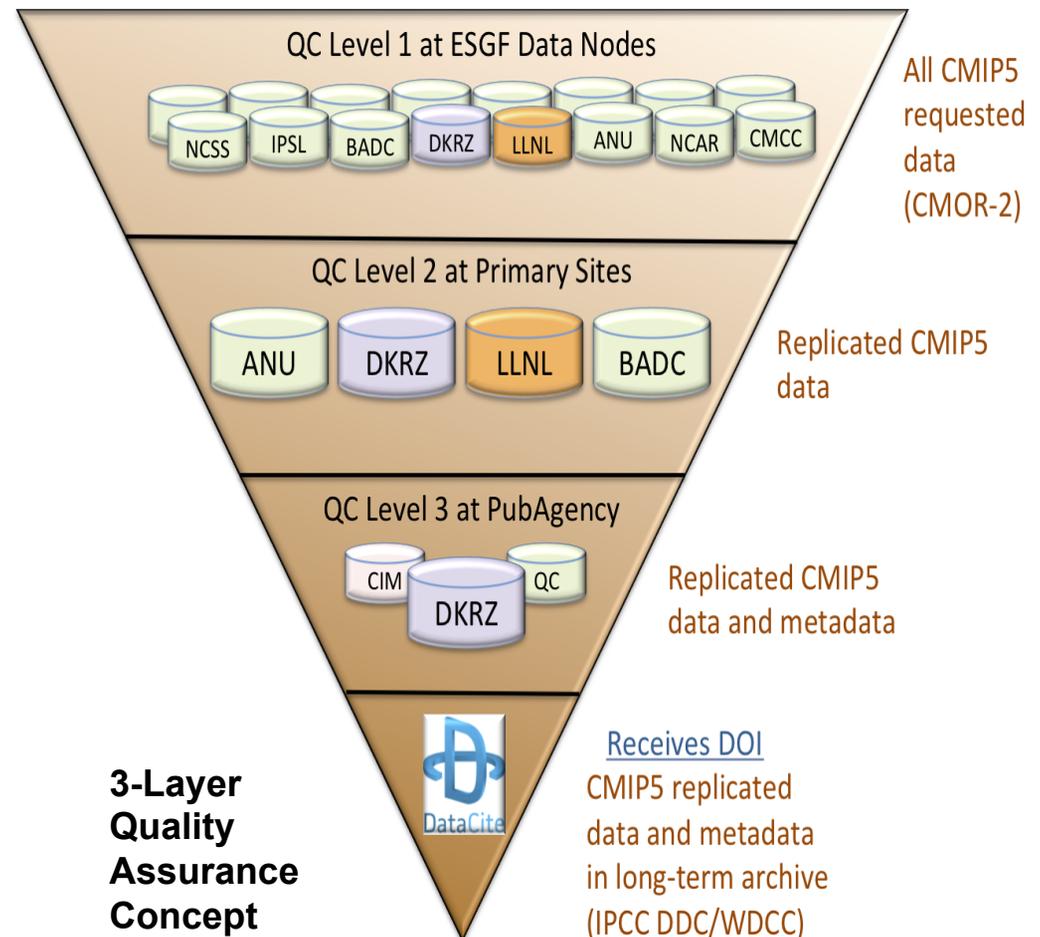
- ESGF is a coordinated multiagency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate
- ESGF climate data holdings and growing (40+)
- 120 core ESGF developers and more than 70 ESGF nodes in 30+ countries
- ESGF distributed data archival and retrieval system
- ESGF provides data quality control processing
- New ESGF technologies developed and deployed by LLNL and our partners

- Automated publication
- Security
- Versioning and replication
- Automated quality control (QC)
- Tracking and feedback notification
- User interface development
- Data transfer (WGET, BDM, GridFTP, Globus, BeStMan)
- DAP services (THREDDS Data Servers [TDS], OPeNDAP)
- Installation
- Dashboard (system monitor service)
- Improve use ability
- Computing hardware
- Local and remote analysis (over 40 core developers)
- Diagnostics and metrics and exploratory and analysis visualization
- Workflow and provenance capture
- International Climate Network Working Group (ICNWG)



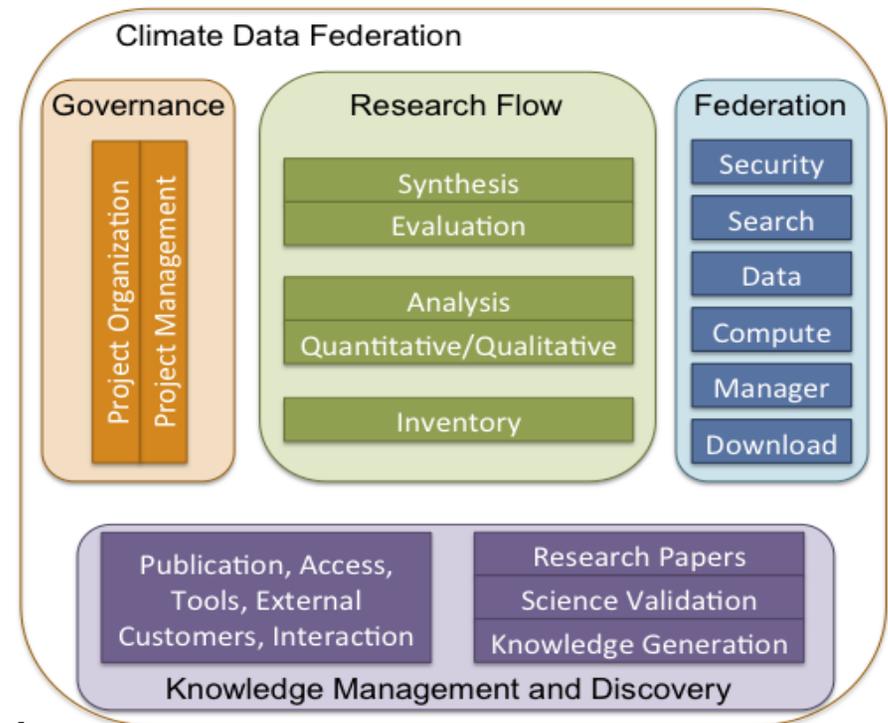
ESGF provides data quality control processing

- **Publishing data** to an ESGF portal performs QC Level 1 (QCL1) check
 - ✓ QCL1 data are visible to users and are identified as QCL1 on the UI
- **DKRZ (MPI) quality control code** is run on data to perform QC Level 2 (QCL2) check
 - ✓ QCL2 data are visible to users and are identified as QCL2 on the UI
- **Visual inspections** are performed for inconsistencies and metadata correctness at QC Level 3 (QCL3) check
 - ✓ QCL3 data are visible to users and are identified as QCL3 on the UI
 - ✓ **Digital Object Identifiers (DOIs)** are given to data sets that pass the QCL3 check

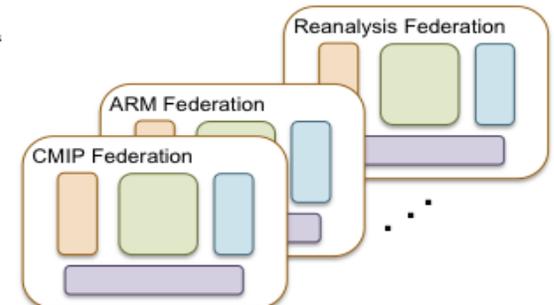


ESGF software system integration data federation services (i.e., data services)

- **NetCDF Climate and Forecast (CF) Metadata Convention**
 - ✓ LibCF, Mosaic
- **Climate Model Output Rewriter 2 (CMOR-2)**
- **Regridders: GRIDSPEC, SCRIP, and ESMF Publishing**
- **Search and Discovery**
- **Replication and Transport**
 - ✓ GridFTP, OPeNDAP, DML, Globus, BeSTMan (HPSS)
 - ✓ Networks
- **Data Reference Syntax (DRS)**
- **Earth System Documentation (ES-DOC)**
- **Quality Control**
 - ✓ QC Level 1, QC Level 2, QC Level 3, Digital Object Identifiers (DOIs)
- **Websites and Web Portal Development**
 - ✓ Data, Metadata, Journal Publication Application
- **Notifications, Monitoring, Metrics**
- **Security**
- **Product Services**



Peer Node Science Discovery & Information Exchange



New ESGF browser and desktop clients

The screenshot shows the ESGF Portal homepage. At the top, there are logos for ESGF (Earth System Grid Federation) and PCMDI. Below the logos is a navigation bar with links for Home, Search, Tools, Login, and Help. A large world map is featured in the center. Below the map, there is a section for 'Peer Nodes' listing various research institutions like ANL, BADC, BNU, CMCC, DKRZ, NOAA-ESRL, NOAA-GFDL, IPSL, NASA-GSFC, NASA-JPL, NCI, NERSC, ORNL, PCMDI, and SMHI-LIU-NSC. To the right of the peer nodes is a section for 'About esgf-pcmdi-9' and 'Resources'. At the bottom, there is a 'Quick Links' section and a footer with contact information and legal notices.

The screenshot shows the WCRP CMIP6 website. At the top, there are logos for ESGF-CoG Node at PCMDI and WCRP CMIP6 (World Climate Research Programme). Below the logos is a navigation bar with links for Home, About Us, Developers, and Governance. The main content area is titled '6th Coupled Model Intercomparison Project (CMIP6)' and contains text about the project's goals and challenges. To the right of the main text is a sidebar with sections for 'Visitors', 'Metrics', and 'Administrators'. Below the main text is a circular diagram illustrating the CMIP6 experimental design, with segments for 'Systematic Biases', 'Response to Forcings', 'Variability, predictability, future scenarios', and 'Short-term hindcasts'. At the bottom of the page, there is a footer with contact information and legal notices.

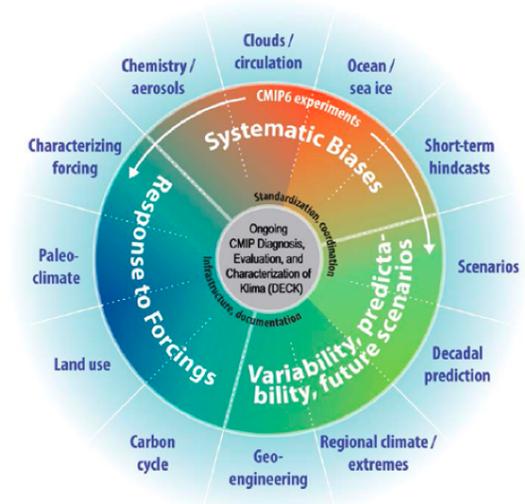


Figure and text from Meehl et al, 2014: Climate Model Intercomparisons: Preparing for the Next Phase, EOS Transactions, 95.4, 77-78.

UV-CDAT big data access and displays

- Client analysis tool access
- Thick and smart client analysis tool
- Seamlessly integrate of new components and packages into the framework
- Quickly explore massive amounts of data in unique ways to form new hypotheses and verify simulation data against observational data

UV-CDAT Application (thick-client)

UV-CDAT Live web browser (smart-client)

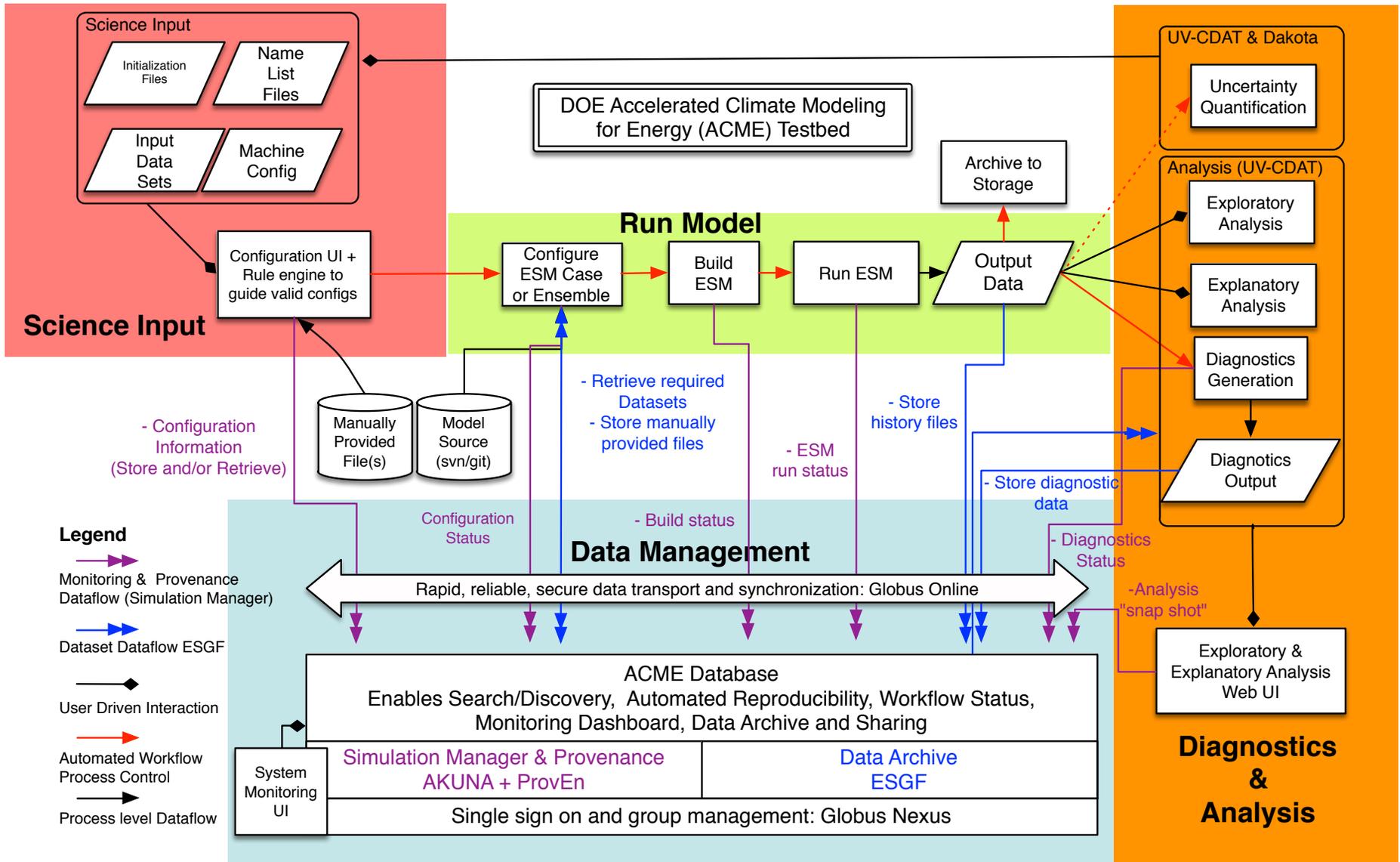
International Climate Network Working Group (ICNWG)



Participating network collaborators

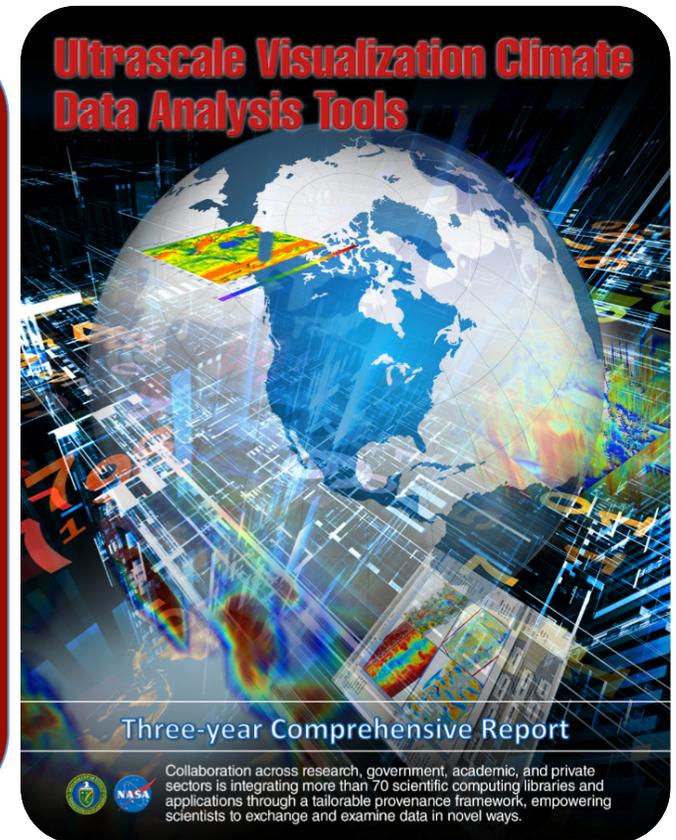
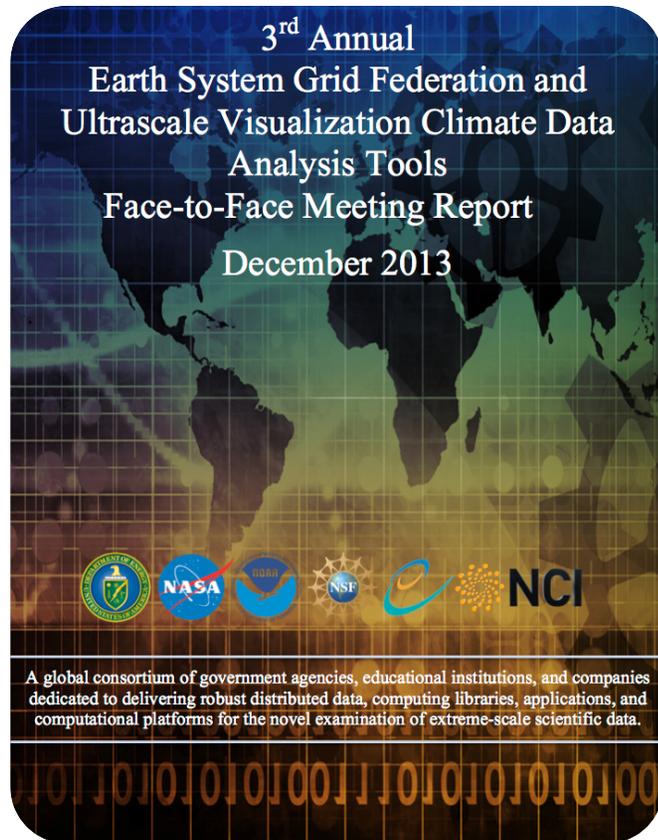


Accelerated Climate Modeling for Energy (ACME) End-to-end workflow and data infrastructure architecture



Data discovery and workflows

- <http://esgf.org>
- <http://uv-cdat.org>
- <http://pcmdi9.llnl.gov>
- <http://icnwg.llnl.gov>
- <http://aims.llnl.gov>
- <http://climate.llnl.gov>



4th Annual Earth System Grid Federation and Ultrascale Visualization Climate Data Analysis Tools Face-to-Face Conference will be December 9-11 in Livermore, California



**Lawrence Livermore
National Laboratory**