

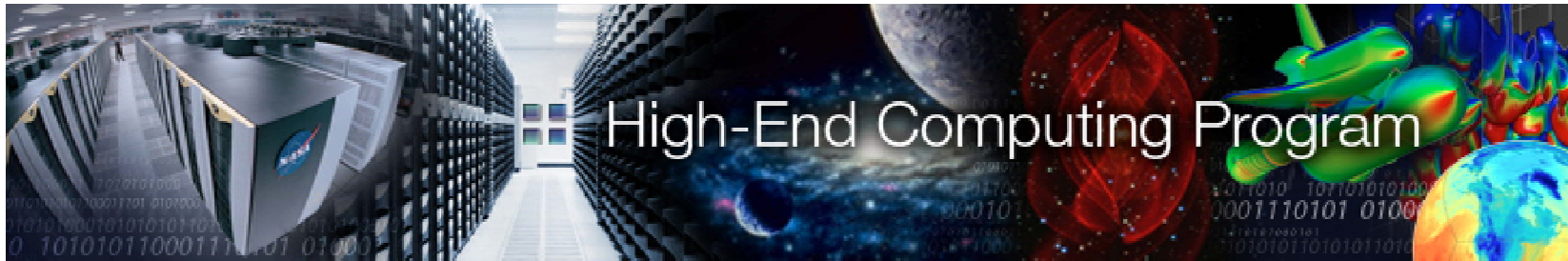
Data Management System: Toward an iRODS-Based Approach to Scientific Data Services

John Schnase, William P. Webster, Lynn A. Parnell
and Daniel Duffy
NASA Center for Climate Simulation
Goddard Space Flight Center

2011 IEEE Mass Storage Systems & Technologies
Research Track/Short Papers
May, 2011

<http://nccs.nasa.gov>
<http://cisto.gsfc.nasa.gov>
<http://www.hec.nasa.gov>

NASA High-End Computing Program



HEC Program Office
NASA Headquarters
Dr. Mike Little
Scientific Computing Portfolio Manager



**High-End Computing
Capability (HECC) Project**
**NASA Advanced
Supercomputing (NAS)**
NASA Ames
Dr. Rupak Biswas

**NASA Center for Climate
Simulation**
Goddard Space Flight Center
Dr. Phil Webster



NCCS Mission

- *Traditional*
 - Enable scientists to increase their understanding of the Earth and the universe by providing state-of-the-art high performance computing, storage, network, and application solutions
 - Provide large-scale compute engines, analytics, data sharing, and high-end computing services
- *Future*
 - Develop a data services capability to better support the climate research communities and prepare the way for technology advances



New Science Goals

- *Intergovernmental Panel on Climate Change (IPCC) Assessment Report 5 (AR5)*
 - Provide the data management services and analytical tools necessary to support the publication requirements of the IPCC
- *Observation/Simulation Data Integration*
 - Bring the climate modeling and observational communities together to work toward the goal of integrating model outputs and observational data
- *Next Generation HEC Requirements for Modeling and Assimilation*
 - Contribute emerging technologies to address computing the ever increasing requirements for Earth system modeling
 - Continuing to push the resolution of global models to the highest possible resolutions



Technology - integrated Rule-Oriented Data System (iRODS)

- Open source data grid software developed by the Data Intensive Cyber Environments (DICE) group, University of North Carolina
- Targets large repositories and digital preservation
- Supports the federation of independent, distributed collections
- Supports server-side workflows that are implemented by chaining execution rules together based on data policies
- Includes features such as domain-specific validation, automatic replication, and digital signature/checksum computation
- Validates assertions about data such as integrity and authenticity



iRODS Data Grids

- **Observation Data**

- Moderate Resolution Imaging Spectro-radiometer (MODIS) observational data
 - 54 million registered files, 630 TB of data, and over 300 million defined metadata values
- Small-scale, multi-product, application-specific data service
 - The Invasive Species Data Service (ISDS) manages a collection of MODIS data products for ecological forecasting applications

- **Simulation Data**

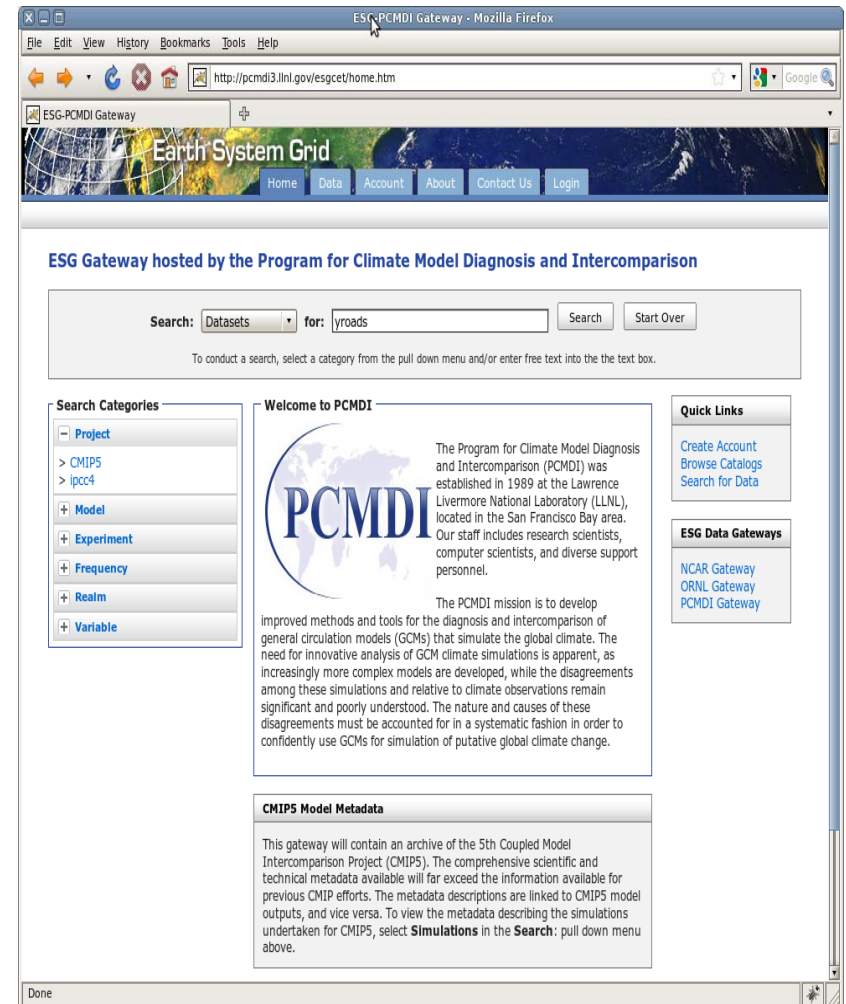
- Modern Era Retrospective-Analysis for Research and Applications (MERRA) simulation data
 - 360 files, 47 GB of data, and 4000 metadata values
- Year of Tropical Convection (YOTC) data sets
 - 134,000 files, 12 TB of data, and 400,000 metadata values

The image shows four overlapping screenshots of NASA data collection pages. Each page displays metadata for a specific data collection, including details like Collection, Data, Type, Format, Customers, Distinction, Interfaces, and Status. The collections shown are modis_Zone (MODIS Atmosphere), isds_Zone (Invasive Species Data Service), merra_Zone (Modern Era Retrospective-Analysis for Research and Applications), and yotc_Zone (Year of Tropical Convection).



Next Steps – Data Services

- *Earth System Grid*
 - Use iRODS to address the publication requirements of IPCC/AR5, which requires integration of ESG
- *Data Integration*
 - Create the framework for integrating observational and simulation data
- *Cloud Provisioning*
 - Incorporate cloud computing to better serve the extended research community



The screenshot shows a web browser window titled "ESG-PCMDI Gateway - Mozilla Firefox". The address bar displays "http://pcmdi3.llnl.gov/esgce/home.htm". The website header features the "Earth System Grid" logo and navigation links for Home, Data, Account, About, Contact Us, and Login. Below the header, a search bar is visible with "Datasets" selected in the pull-down menu and "yroads" entered in the text box. The main content area includes a "Welcome to PCMDI" section with a world map logo and text describing the Program for Climate Model Diagnosis and Intercomparison (PCMDI). A "Search Categories" sidebar on the left lists Project, Model, Experiment, Frequency, Realm, and Variable. A "Quick Links" sidebar on the right contains links for "Create Account", "Browse Catalogs", and "Search for Data". A "CMIPS Model Metadata" section is also present at the bottom of the main content area.

