

CCSM Portal/ESG/ESGC Integration (a PY5 GIG project)

Lan Zhao, Carol X. Song
Rosen Center for Advanced Computing
Purdue University

With contributions by: Kathy Saint/SGI,
Cecelia DeLuca/NOAA CIRES, Don Middleton/NCAR



TeraGrid[®]

CCSM/ESG/ESGC Collaboration (PY5)

• Goal

– Semantically enabled environment that includes modeling, simulated and observed data, visualization and analysis

- building upon ESG, ESG-Curator (ESGC) and Purdue CCSM portal
- include one or more gateways

• Initial phase (PY5):

– Establish the initial phase of a prototype environmental science gateway on the TeraGrid

– Enable users to launch climate model runs on the TeraGrid and publish metadata and data in the data curation system.

– Integrate the Purdue CCSM portal with ESG and ESGC

- CCSM portal: enable CCSM runs on the TeraGrid
- ESG: enable management, discovery, access and analysis of climate data in a distributed environment
- ESGC: support end-to-end modeling in the Earth Sciences – linking models and data with metadata

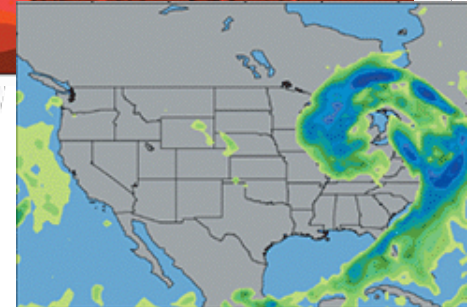
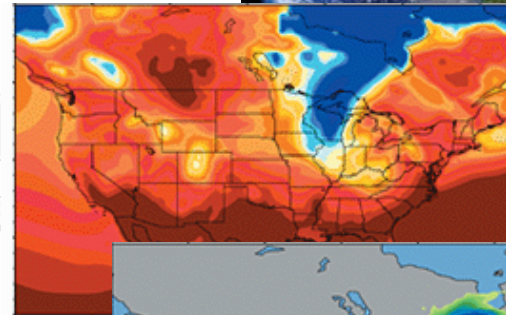
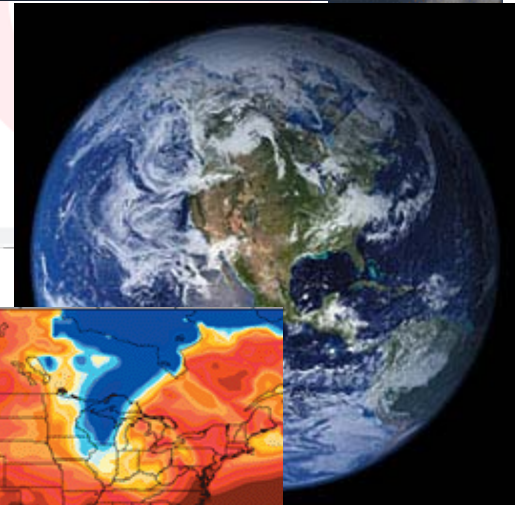


CCSM Overview

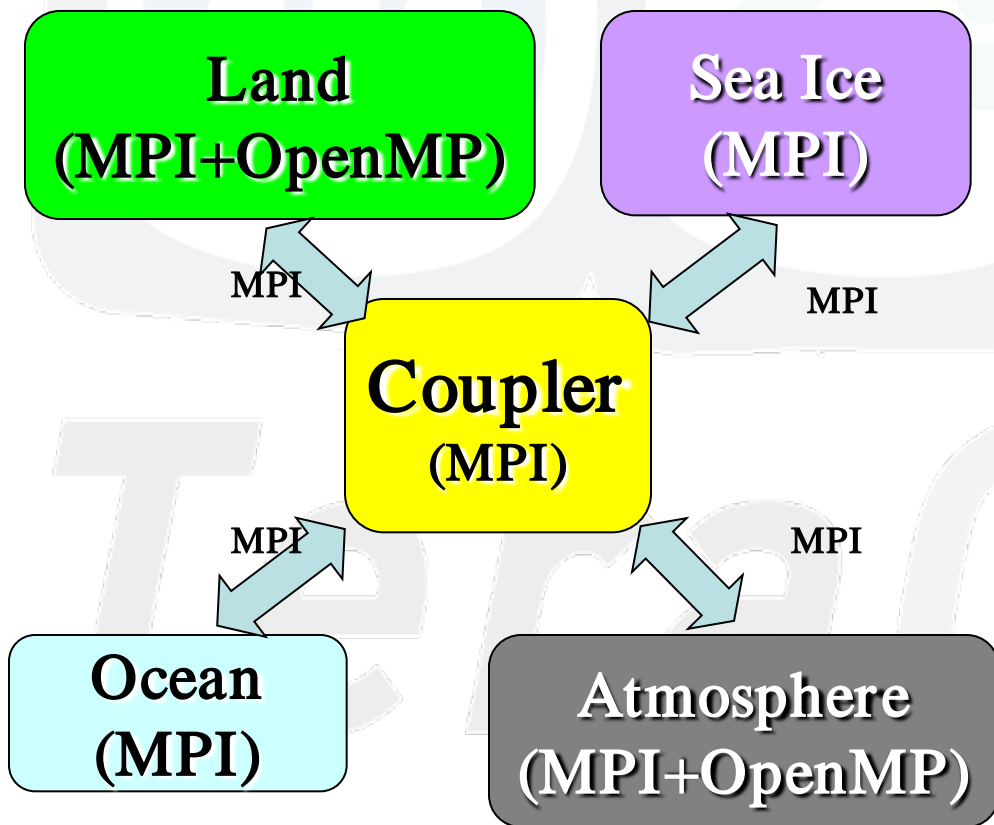
- *The Community Climate System Model (CCSM)*

<http://www.cesm.ucar.edu> is a coupled climate model for simulating the earth's climate system.

- Initially developed at the National Center for Atmospheric Research (NCAR) at Boulder, Colorado.
- Provides the modeling framework for confronting scientific questions about the Earth's past, present and future climate states



CCSM Overview



- Four models (components)
- Each model has *Active, Data* and *Dead* versions
- Models communicate with a *Coupler* component every time step



CCSM on the TeraGrid

- CCSM has high computational and storage needs.
- A typical model run on an IBM “bluesky” system, at a dataset resolution of T42_gx1v3, has the following requirements :
 - History-File Volume: 6.5 Gbytes/model year
 - Restart-File Volume: 0.9 Gbytes/model year
 - Simulation Years/Day: 7.5 on 104 CPUs
- TeraGrid provides suitable computational, storage and networking resources to run CCSM.

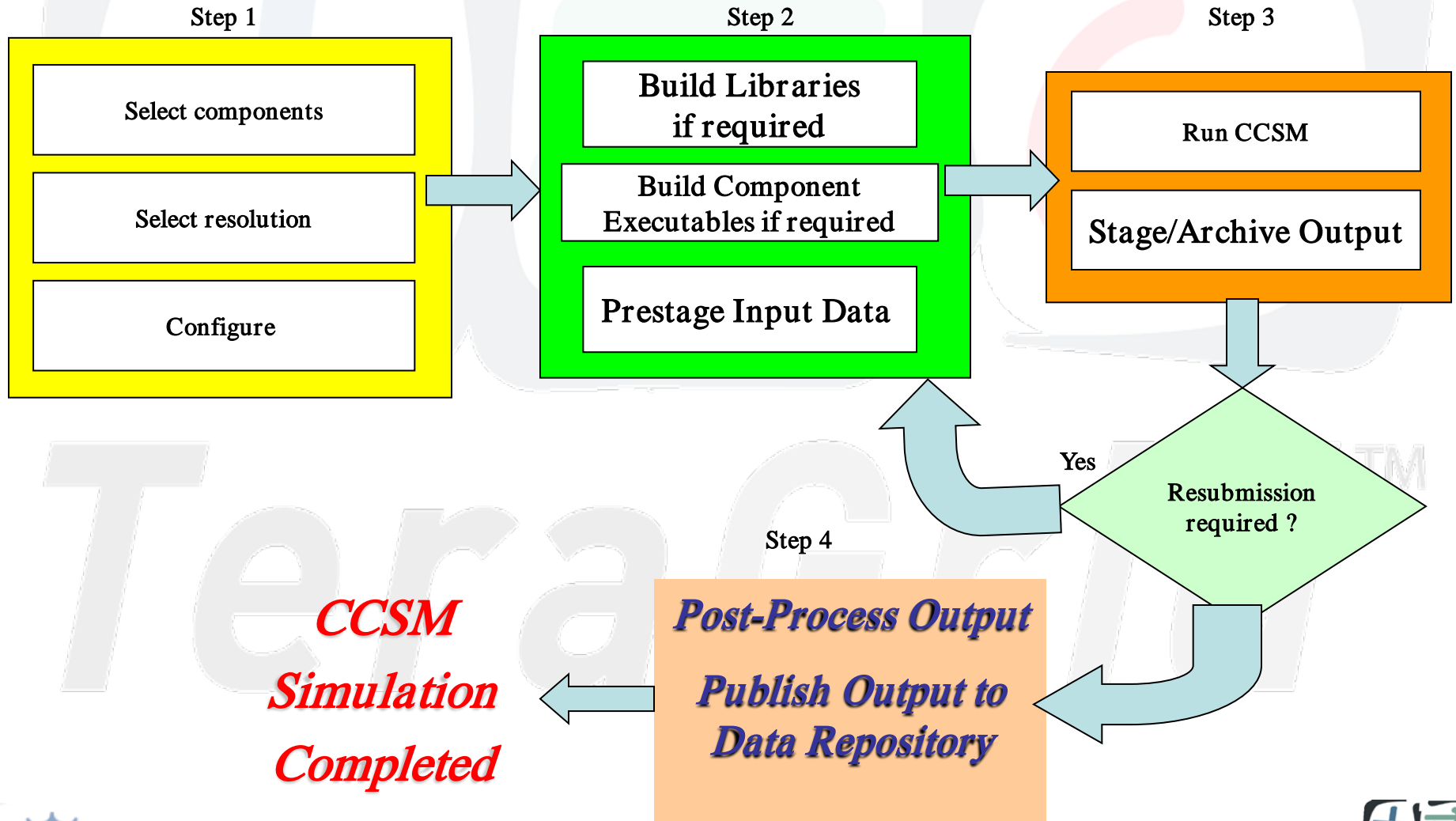


Challenges

- Even for experienced CCSM Users, the following challenges exist in using CCSM on the TeraGrid
 - Porting and Validation on a new platform
 - Performance Tuning
 - Learning curve for TeraGrid tools, protocols, specifics of batch queuing systems
 - Making changes to CCSM software stack to accommodate specifics of TG software
 - Large data (need storage and data management tools)
 - Collaboration and Dissemination of Results



Steps in a typical CCSM Simulation

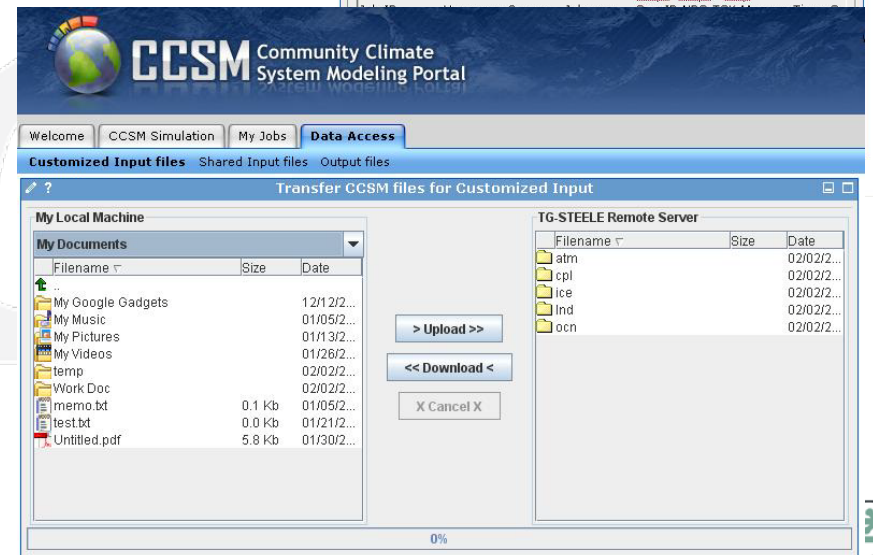
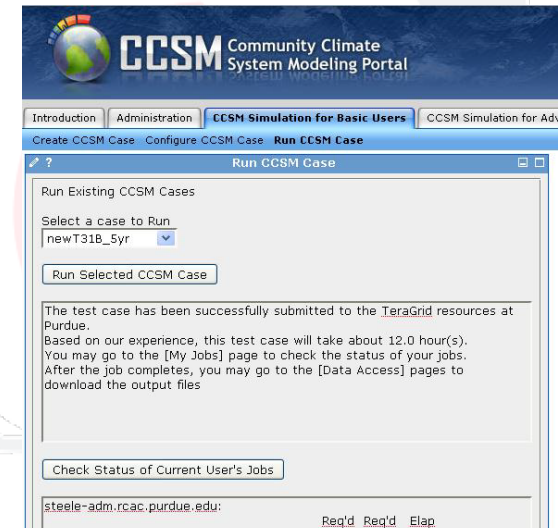


Climate Modeling Portal

Community Climate System Model (CCSM) to simulate climate change on Earth

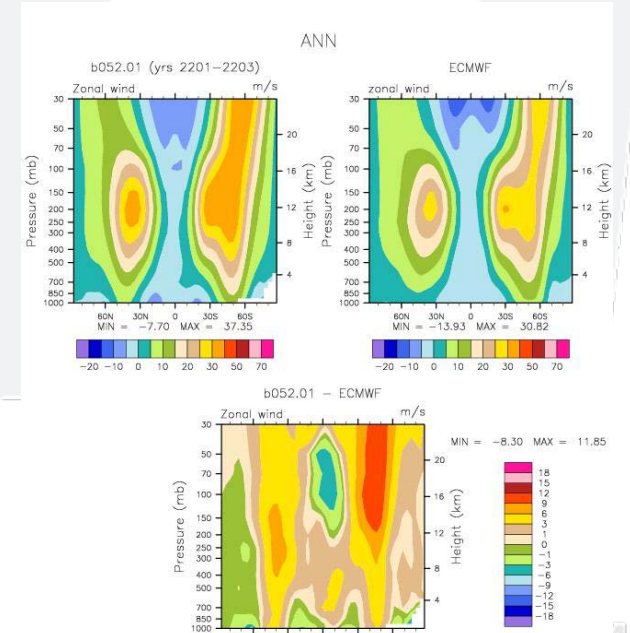
- Easy to use interface to compose, run, and monitor CCSM jobs using TeraGrid resources

- Basic user interface and advanced user interface
- Still allow control in editing simulation configuration
- Open to both Purdue and non-Purdue users
- Link to Purdue LDAP authentication
- Data upload/download
- Use Purdue Steele Linux cluster
- Use TG community account
- Data post processing and visualization
- Job management and status tracking



Climate Modeling Portal

- TRAC allocation on Steele, Queenbee and Ranger
- Being used in 2009 Fall class
 - POL 520/EAS 591: Models in Climate Change Science and Policy
 - Semester-long projects, generate policy recommendations based on scientific, economic, and political models of climate change impacts



Tera

CCSM
Community Climate System Modeling Portal
Welcome, Lan Zhao

Introduction | CCSM Simulation for Basic Users | CCSM Simulation for Advanced Users | My Jobs | Data Access

Manage My Jobs | Monitor My Jobs

My Jobs

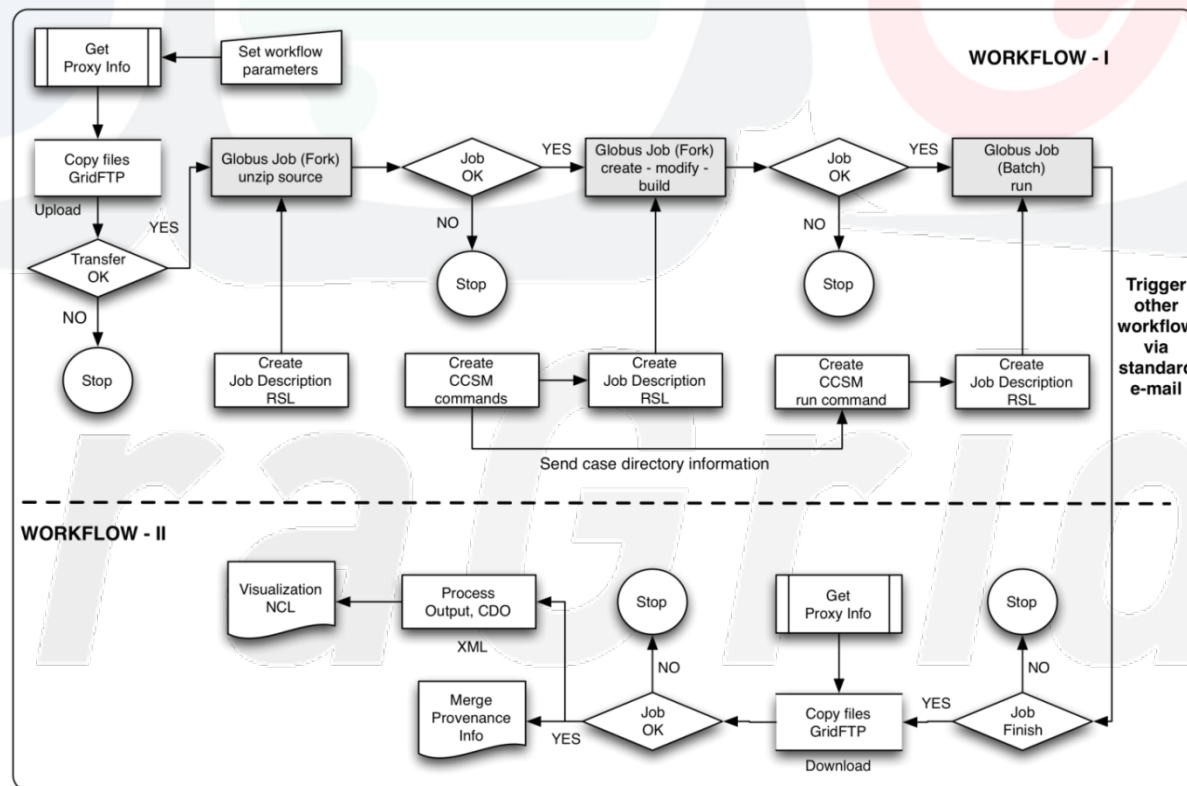
Update Status
Delete Selected Cases

Case Name	Job ID	Resolution	Component	Teragrid site	Creation Time	Submission Time	Queue	Month	Job Status	Status Details	Delete Case/Files
t31b2	1710911	T31_gx3v5	B	Purdue	2009-07-13 14:04:40.0	2009-07-10 22:06:40	tg_workq	24	COMPLETED	Details	<input type="checkbox"/>
test2	1613134	T31_gx3v5	B	Purdue	2009-06-30 15:58:03.0	2009-06-30 13:46:17	standby-8	1	COMPLETED	Details	<input type="checkbox"/>
t31b3	no pbs job id	T31_gx3v5	B	Purdue	2009-04-10 16:30:19.0	0000-00-00-NA	NOT_AVAILABLE	24	CASE CREATED	N/A	<input type="checkbox"/>



CCSM Self-Describing Workflows

- Turuncoglu (2009) used Kepler to implement a CCSM4 (development version of CCSM) workflow as part of the Curator project
- Kathy Saint has been updating and simplifying the workflow using web services

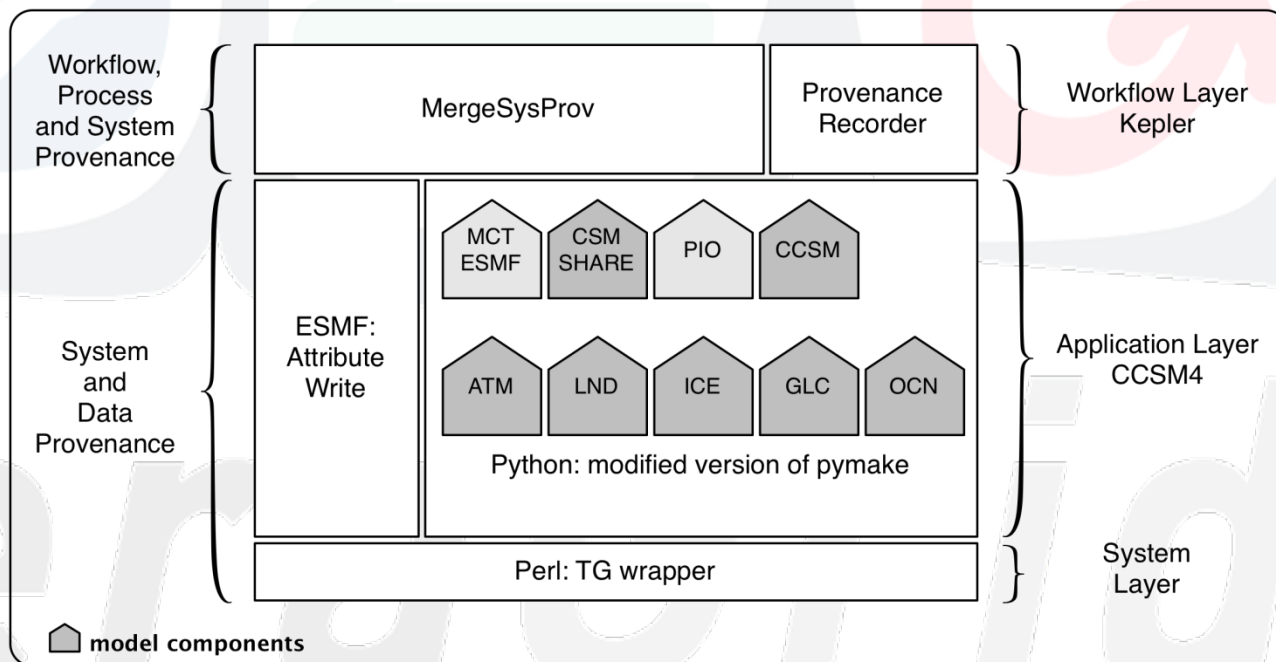


Workflow includes uploading source code; creating, building and running case; and collecting provenance data. Workflows connected via email message containing job description XML file or standard workflow definition file.



Provenance collection

- Multiple levels of metadata (system, data, process, workflow) describing the CCSM4 runs were collected automatically using a variety of tools

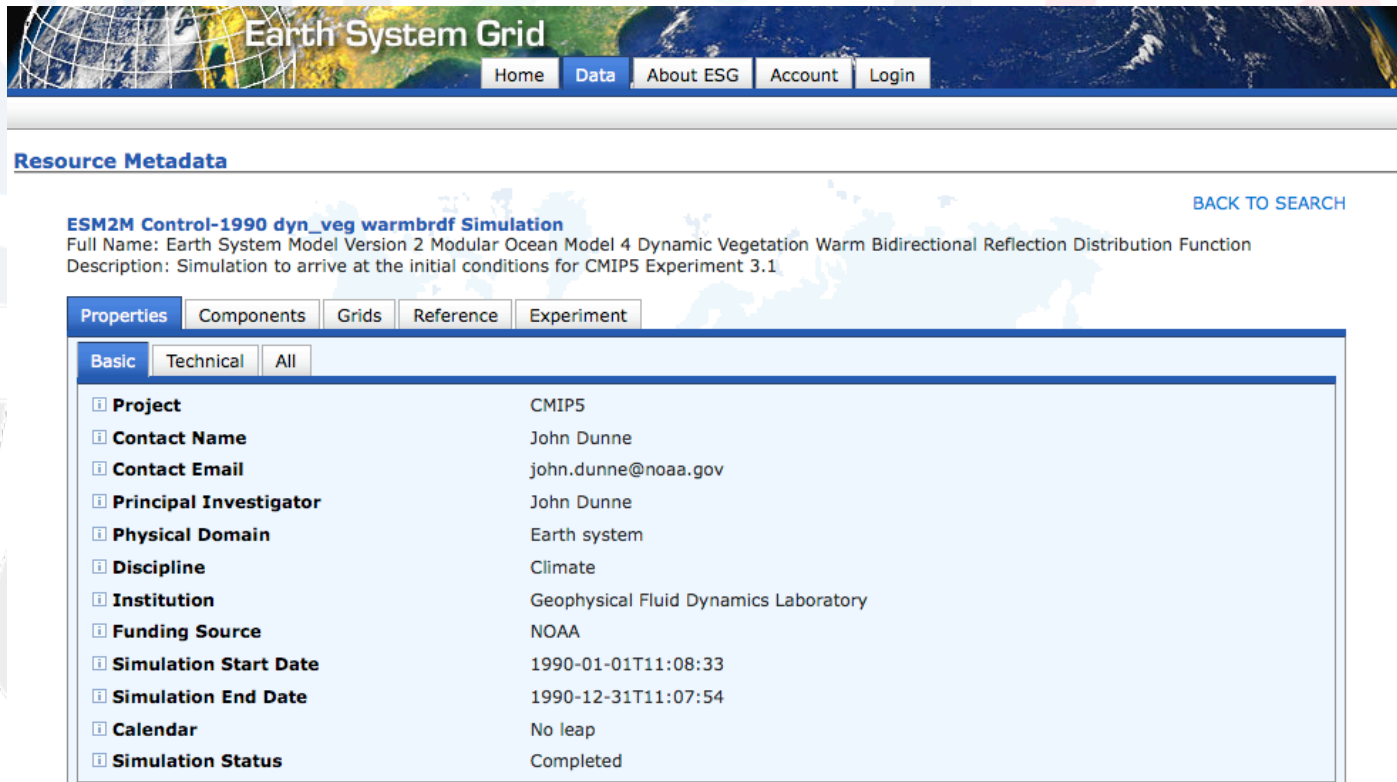


pymake – provided by ORNL and NCSU
tgwrapper.pl – uses SoftEnv and Modules applications



Metadata display

- The collected metadata can be ingested into the ESGC portal, where it can be searched, browsed, and compared



The screenshot shows the Earth System Grid (ESGC) portal interface. At the top, there is a navigation bar with links for Home, Data, About ESG, Account, and Login. Below this is a section titled "Resource Metadata" with a "BACK TO SEARCH" link. The main content area displays the title "ESM2M Control-1990 dyn_veg warmbrdf Simulation" and its full name and description. A tabbed interface allows users to view Properties, Components, Grids, Reference, and Experiment. The "Properties" tab is active, showing a list of metadata fields and their values.

Properties	
Project	CMIP5
Contact Name	John Dunne
Contact Email	john.dunne@noaa.gov
Principal Investigator	John Dunne
Physical Domain	Earth system
Discipline	Climate
Institution	Geophysical Fluid Dynamics Laboratory
Funding Source	NOAA
Simulation Start Date	1990-01-01T11:08:33
Simulation End Date	1990-12-31T11:07:54
Calendar	No leap
Simulation Status	Completed

A version of this user interface will be used to display metadata from CCSM and other models for experiments conducted for the 5th IPCC Assessment Report



CCSM/ESG/ESGC Collaboration (PY5)

- **Task 1: Publish Purdue CCSM archive data to ESG**
 - Data federation: access Purdue's climate data archive through ESG
 - Integrate ESG publishing interfaces on Purdue resources
 - Downloaded the scripts for ESG data node installation
 - Need sudo or root access to run the script, security requirements
 - Setting up a test server at Purdue
 - Software stack: TDS, PostgreSQL, Python/CDMS, ESGCET, Tomcat, Globus toolkit (GridFTP server + MyProxy client)
 - How to access the data archive from SRB, OpenDAP, and ESG?



CCSM-NCAR Collaboration (PY5)

- **Task 2: Enable ESGC to run CCSM simulations on TG**
 - Provide web service interfaces for remote model run invocation
 - CreateCase, ConfigureCase, RunCase, TrackStatus, GetResult, ListCases
 - Issue: CCSM v4 (unreleased, used by ESGC) vs. CCSM v3 (used by CCSM portal)?
 - Developed a prototype Java client that invokes simple web services interfaces to run a T31_gx3v5/B simulation using CCSM v3.
 - Working with Kathy Saint to learn more about CCSM v4, define use case, workflow steps/interfaces, and security model.
- **Task 3: Publish model run datasets and metadata back to ESG from both ESGC and Purdue CCSM**
 - Availability of TeraGrid-produced climate model datasets in ESG archives
 - Data publishing and wide area transport
 - the scripts ESGC uses to collect metadata are integrated with CCSM v4



CCSM-NCAR Collaboration (PY5)

- **Design and discussion**

- Conference call biweekly
- Exchanged documentation
- Learning about ESG/ESGC systems, investigating on how to integrate and potential problems
- Discussion with NCAR about design details

- **Plan for next quarter:**

- Task 1: set up a test server. Install a ESG data node. Publish example data to ESG.
- Task 2: provide an example java client that invokes a set of web services to run CCSM v3 at Purdue TG. Collect interface and security requirements. Install CCSM v4 and learn how it works in comparison with v3. Define service interfaces.
- Task 3: learn more once the CCSM v4 testbed is set up.



Thanks

- ***CCSM Portal***

<http://www.purdue.teragrid.org/ccsmportal>

- **For more details, contact**

- Lan Zhao lanzha@purdue.edu
- Carol X. Song cxsong@purdue.edu

For Kepler workflows and ESGC:

- Kathy Saint ksaint@sgi.com
- Cecelia DeLuca cecilia.deluca@noaa.gov
- Don Middleton don@ucar.edu
- Sylvia Murphy sylvia.murphy@noaa.gov



TeraGrid™