

SciDAC Earth System Grid Center for Enabling Technologies

Semi-Annual Progress Report for the Period April 1, 2008 through September 30, 2008

Principal Investigators

Dean N. Williams³, Ian T. Foster¹, and Don E. Middleton⁵

The Earth System Grid Center for Enabling Technologies Team:

Rachana Ananthakrishnan¹, Frank Siebenlist¹

Arie Shoshani², Alex Sim²,

Robert Drach³

Phil Jones⁴

David Brown⁵, Julien Chastang⁵, Luca Cinquini⁵, Peter Fox⁵, Danielle Harper⁵, Nathan Hook⁵, Eric Nienhouse⁵, Gary Strand⁵, Patrick West⁵, Hannah Wilcox⁵, Nathaniel Wilhelmi⁵, Stephan Zednik⁵

Steve Hankin⁶, Roland Schweitzer⁶

David Bernholdt⁷, Meili Chen⁷, Jens Schwidder⁷

S. Bharathi⁸, Ann Chervenak⁸, Robert Schuler⁸, Mei-Hui Su⁸



**Climate simulation data are now securely accessed, monitored,
cataloged, transported, and distributed to the national and
international climate community**

¹ Argonne National Laboratory, Chicago

² Lawrence Berkeley National Laboratory

³ Lawrence Livermore National Laboratory

⁴ Los Alamos National Laboratory

⁵ National Center for Atmospheric Research

⁶ National Oceanic and Atmospheric Administration

⁷ Oak Ridge National Laboratory

⁸ University of Southern California, Information Sciences Institute

Table of Contents

<i>SciDAC Earth System Grid Center for Enabling Technologies</i>	<i>1</i>
1 Executive Summary	4
1.1 Overall goal for this reporting period	4
1.2 Highlights	5
1.2.1 LLNL ESG CMIP3 Portal and R&D Highlights	5
1.2.2 NCAR ESG Gateway Portal R&D Highlights	6
1.2.3 ORNL ESG Gateway Portal R&D Highlights	7
1.2.4 LANL ESG Node Highlights	7
1.2.5 LBNL Storage Resource Manager Highlights	7
1.2.6 PMEL Product Delivery Services Highlights	8
1.2.7 ANL Security, Data, and Services Highlights	8
1.2.8 ISI Monitoring, Data Catalogs, and Federation Highlights	9
2 Overall Progress	9
2.1 ESG-CET Gateway	9
2.2 ESG-CET Data Node Publisher	11
2.3 ESG-CET Cyber Security	13
2.4 Product Services	13
2.5 DataMover-Lite (DML)	13
2.6 BeStMan Java Client API and the Storage Resource Managers (SRM)	14
2.7 Data Catalogs	14
2.8 Monitoring	14
2.9 Packaging of ESG-CET	15
2.10 Data Transport Services	15
3 Architectural Design Diagrams, Requirement Documents and Use Cases	16
4 ESG-CET Group Meetings	16
4.1 ESG-CET Executive Meeting (April 1 – April 3)	17
4.2 ESG All-hand Meeting (April 29 – May 1)	17
4.3 Executive Meeting (July 13 – 16)	17
4.4 PI Meeting (GO-ESSP) Workshop (September 16 – 19)	17
5 Collaborations	17
5.1 University of Utah Visit to Discuss VisTrails, ESG, and CDAT Collaboration	17
5.2 Collaboration with the SciDAC VACET Visualization Team	18
5.3 Collaborations with the Southern California Earthquake Center (SCEC)	19

5.4	Scientific Data Management (SDM) Center for Enabling Technology (SciDAC CET)	19
5.5	Center for Enabling Distributed Petascale Science (CEDPS) (SciDAC CET)	20
5.6	SBIR CDAT-MODAVE – Collaboration with Tech-X	20
5.7	Global Organization for Earth System Science Portal (GO-ESSP) Workshop	21
5.8	Hybrid Coordinate Ocean Model (HyCOM) consortium (NOAA, Navy, et. al.)	21
5.9	NOAA Geophysical Fluid Dynamics Laboratory	21
5.10	NOAA Integrated Ocean Observing System (IOOS) and NOAA Office of Climate Observations (OCO)	21
5.11	NASA JPL visit to LLNL to discuss ESG-CET and Satellite Data	21
5.12	External Support of Grid-Enabled OPeNDAP	22
6	<i>Outreach, Papers, Presentations and Posters</i>	22
6.1	Outreaches Activities	22
6.1.1	World Meteorological Organization (WMO) Information System Intercommission Coordination Group (ICG-WIS)	22
6.1.2	DAARWG Workshop in Ashville, NC	22
6.1.3	Global Earth Observing System of Systems (GEOSS)	22
6.1.4	Steering Committee for HPDG08	22
6.1.5	EGU Session Convener	22
6.1.6	ESIn Editor	23
6.1.7	Ocean Observatory Initiative	23
6.2	Papers:	23
6.2.1	Paper and Presentation: SciDAC ‘08 Conference in Seattle, WA	23
6.2.2	Paper in the Bulletin of the American Meteorological Society (BAMS)	23
6.2.3	SciDAC Review Article	23
6.2.4	Earth Science Informatics Article	24
6.2.5	Federation	24
6.2.6	SciDAC Review Highlights	24
6.3	Talks:	24
6.3.1	Net@Edu Meeting	24
6.3.2	North American Regional Climate Change Assessment Program (NARCCAP) PI and Community Meeting	24
6.3.3	Virtual Climate Change Research Institute Workshop	24
6.3.4	World Meteorological Organization (WMO) ET-WISC Meeting	24
6.3.5	Washington trip to meet with Sponsors	25
6.3.6	ESnet Climate meeting at LBNL with Dan Hitchcock	25
6.3.7	BER Program Manager Wanda Farrell’s visit to ORNL	25
6.3.8	Presented ESG-CET to Christine Chalk	25
6.3.9	Presented ESG-CET to David Skinner and Horst Simon	25
6.3.10	BER Program Manager Anjuli Bamzai visit to ORNL	25
6.3.11	Global Organization for Earth System Science Portal (GO-ESSP) Workshop	25
6.3.12	Presentation at EGU 2008, IN09, The Grid for Geosciences	25
6.4	Posters:	25
6.4.1	SciDAC ‘08 PI Meeting in Denver, CO	25
6.4.2	European Geophysical Union Meeting	26

1 Executive Summary

This report, which summarizes work carried out by the ESG-CET during the period April 1, 2008 through September 30, 2008, includes discussion of highlights, overall progress, period goals, collaborations, papers and presentations. To learn more about our project, please visit the [Earth System Grid](#) website. In addition, this report will be forwarded to the DOE SciDAC project management, the Office of Biological and Environmental Research (OBER) project management, national and international stakeholders (e.g., the Community Climate System Model (CCSM), the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5), the Climate Science Computational End Station (CCES), the SciDAC II: A Scalable and Extensible Earth System Model for Climate Change Science, the North American Regional Climate Change Assessment Program (NARCCAP), and other wide-ranging climate model evaluation activities), and collaborators. This report and others like it can be found on the [ESG-CET](#) website.

The ESG-CET executive committee consists of David Bernholdt, ORNL; Ian Foster, ANL; Don Middleton, NCAR; and Dean Williams, LLNL. The ESG-CET team is a collective of researchers and scientists with diverse domain knowledge, whose home institutions include seven laboratories and one university: Argonne National Laboratory (ANL), Los Alamos National Laboratory (LANL), Lawrence Berkeley National Laboratory (LBNL), Lawrence Livermore National Laboratory (LLNL), National Center for Atmospheric Research (NCAR), Oak Ridge National Laboratory (ORNL), Pacific Marine Environmental Laboratory (PMEL), and University of Southern California, Information Sciences Institute (USC/ISI). All work is accomplished in close collaboration with the project's stakeholders and domain researchers and scientists.

1.1 Overall goal for this reporting period

During this semi-annual reporting period, the ESG-CET team continued its efforts to start newly needed software investigations and complete existing software prototypes and components, and launch software component interoperability— making diverse systems such as the “Gateway”, “Data Node”, “Product Service”, and “Security” work together for our distributed prototype testbed. The launch of the prototype testbed is scheduled for late October with the installation of ESG Gateways at NCAR and LLNL. Using the newly developed ESG Data Publisher, the ESG II CMIP3 (IPCC AR4) data holdings – approximately 35 TB – will be among the first datasets to be published into the testbed. The prototype testbed is scheduled to begin full testing with testbed partners (i.e., LLNL, NCAR, Geophysical Fluid Dynamics Laboratory (GFDL), ORNL, British Atmospheric Data Center (BADC), Model and Data / Max-Planck-Institute for Meteorology (M&D/MPIM)– the German Climate Computing Centre, and the University of Tokyo Center for Climate System Research) in January of 2009.

As we strove to complete software for our testbed, we continued to provide production-level services to the community. These services continued for CMIP3 (IPCC AR4), CCES, and CCSM, Parallel Climate Model (PCM), Parallel Ocean Program (POP), and Cloud Feedback Model Intercomparison Project (CFMIP), and NARCCAP data.

1.2 Highlights

1.2.1 LLNL ESG CMIP3 Portal and R&D Highlights

The [WCRP CMIP3 Multi-Model Database](#) portal continues to provide the world’s climate scientists with the most complete collection of simulation data. The portal serves currently over 35 TB of data to some 2,200 scientific groups worldwide with an average download rate of 600 GB/day. Thus far, more than [500 research papers](#) have been published based on the CMIP3 Multi-Model archive and over 500 TB of data have been downloaded.

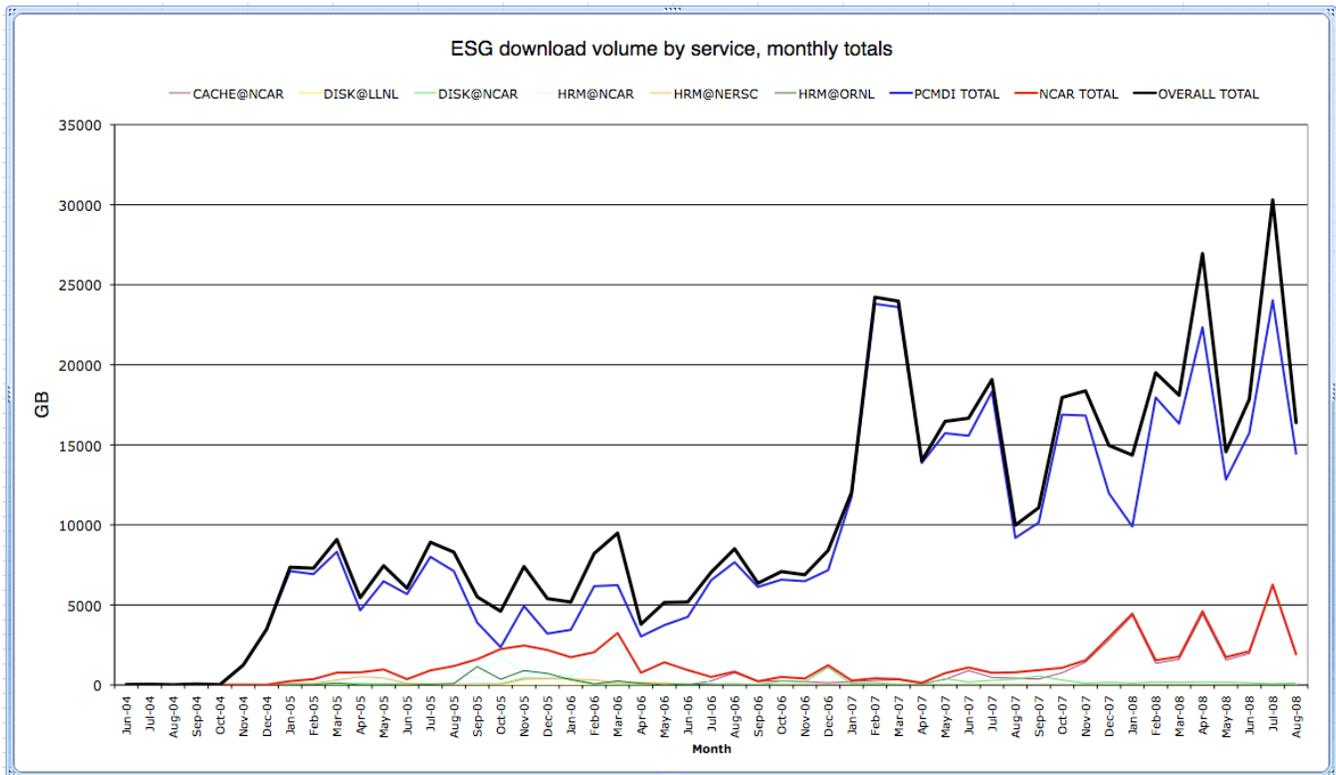


Figure 1: ESG Monthly LLNL CMIP3 and NCAR CCSM Gateway Downloads

As part of the Data Node development, LLNL is working to complete the “Data Publishing” component, which will allow data providers to publish data into the ESG-CET system. As a prototype, the “Publisher” can extract metadata from a list of directories (or a file) representing one or more datasets, into a database. Once extracted, the “Publisher” generates a THREDDS configuration catalog for each dataset, which in turn is used by the “Product Server” for re-initialization. The “Publisher” has the ability to use one of the two most popular open source relational database management system (RDBMS)—MySQL or Postgres—for adding, removing, accessing, and processing data into the database. The first stages of the “Data Node Publisher” were also implemented by initiating the interoperability of SSL client authentication between the “Publisher” and the “Gateway”. Work was completed to allow data providers to publish data via the command line. A graphical user interface is under development and scheduled to be released in October – in time for the launch of the prototype testbed.

For the CMIP5 (IPCC AR5) testbed, LLNL purchased and installed a Dell machine (hostname `pcmdi3.llnl.gov`) that has 8 CPUs running at 3 GHz each, 4 GB of physical memory, and 1 TB of disk

space on a RedHat Enterprise Linux 5.2 operating system. Ports 8080, 8443, and standard ftp ports were opened for other ESG-CET team members’ access to pcmdi3.llnl.gov for interoperability testing.

1.2.2 NCAR ESG Gateway Portal R&D Highlights

The NCAR ESG-CET Gateway portal continued to attract users and see continuous activity during this reporting period. We reached 9003 registered users and have over 200 TB of published data in the form of over one million files. Over the last several months, we have continued to publish new datasets, in particular some of the high-resolution IPCC runs. To date, the global user community has downloaded 66.5 TB of data. Registrations and data access have risen nicely over the last few months, as illustrated by the figure below.

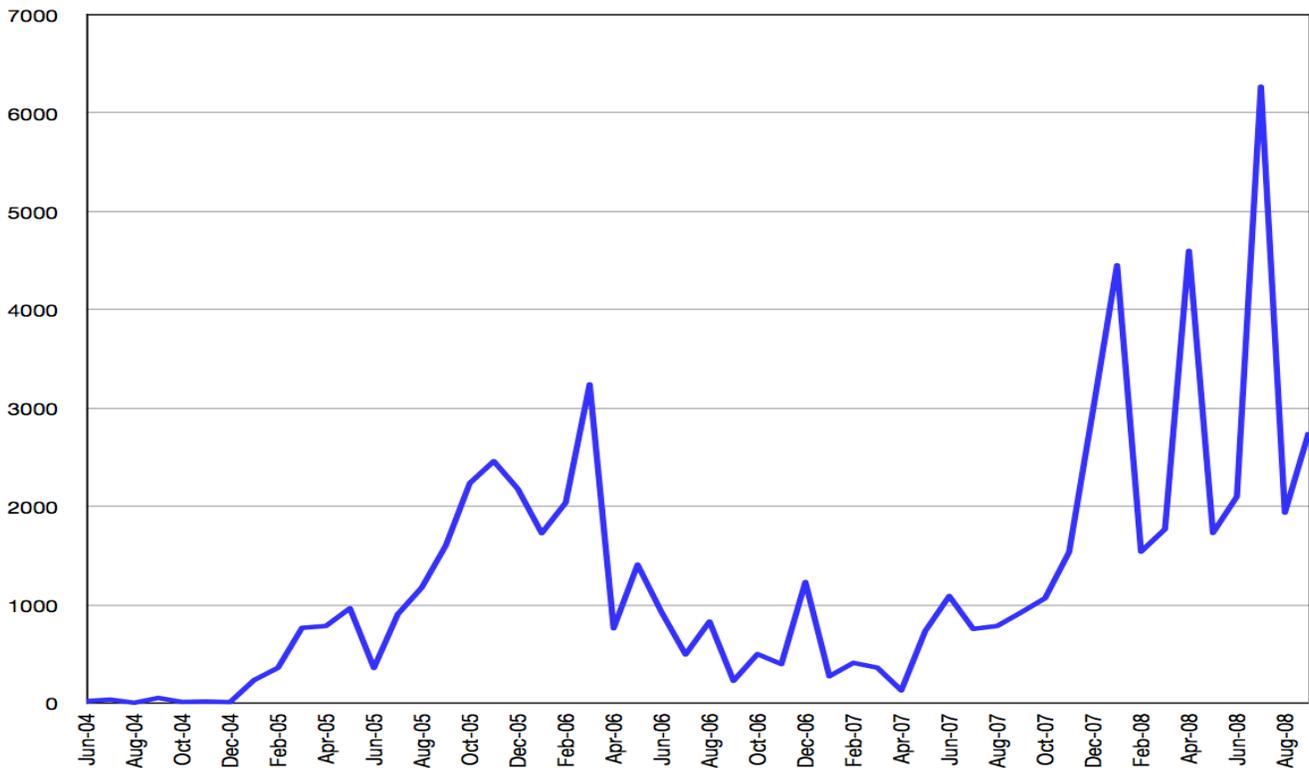


Figure 2: NCAR ESG Download Volume (GB) per Month, to 2008-09-24

The ESG-CET “Gateway” was instrumented to use the emerging OpenID technology for web-based Single Sign-On. Successful federated authentication, including exchange of user attributes, was demonstrated among ESG-CET “Gateways”, and with an external testbed data center – the BADC.

As mentioned above, a prototype data publishing functionality was developed and exposed as web services; thus the “Gateway” can be securely invoked by the “Data Node Publisher” application.

The ESG-CET “Gateway” software was also integrated with BeSTMan, the next generation of Storage Resource Management (SRM) software for deep storage data access. In addition, retrieval of data stored on the NCAR Mass Storage System (MSS) from the ESG-CET “Gateway” interface was successfully demonstrated.

For the “Gateway” web interface, new semantic-based capabilities were developed for faceted data search and models and simulations “track back” and intercomparison. This includes a new user interface that allows dynamic selection and updating of arbitrary search categories.

Finally, a prototype of the ESG-CET “Gateway” framework was successfully deployed in support of the 2008 NCAR Advanced Study Program Summer Colloquium on “Dynamical Cores of Atmospheric General Circulation Models”.

1.2.3 ORNL ESG Gateway Portal R&D Highlights

The C-LAMP team has made available the complete Experiment 1 data, which have been processed and published with help from LLNL. At present, 888 files – approximately 64 GB – of C-LAMP data are accessible to users.

After a testing period, our "ESGtg" data node was put into production and the old server "sleepy" was retired. “ESGtg” is located on the ORNL TeraGrid network enclave, with a 10 Gbps direct link to NCAR. It also has several TB of local disk space to serve as cache for the SRM. Recent logs indicate that “ESGtg” is sending an average of 120 data files per day to the NCAR ESG portal in response to user requests.

We are in the process of specifying and procuring a moderately sized server to support the prototype testbed deployment planned for the next period at ORNL. We also are beginning discussions with the ORNL computer centers (DOE-funded National Center for Computational Sciences (NCCS) and NSF-funded National Institute for Computational Sciences (NICS)) concerning the relationship that ESG will have with the central infrastructure of these centers over the longer term. Issues for discussion include whether it will be possible (given policy, security, and technical considerations) for ESG to utilize these central resources, or whether ESG needs to remain independent.

1.2.4 LANL ESG Node Highlights

LANL continued to serve data from its node – both through ESG and also through direct web access. With the help of NCAR, LANL worked through unresolved issues with ESG NCAR portal downloads.

Working with Jim Ahrens’ visualization team, LANL continued testing the ParaView parallel visualization package and evaluating it as a potential remote analysis and visualization tool to handle large data sets needed for ESG-CET.

Finally, work began on the re-design of the SCRIP interpolation package for eventual use in ESG-CET. New features include subroutine interfaces (rather than off-line computations), emerging grid specification standards (i.e., gridspec – a standard description of grids used in earth system models), and reformulated algorithms to improve robustness.

1.2.5 LBNL Storage Resource Manager Highlights

As a requirement for ESG-CET, the new SRM called BeStMan (Berkeley Storage Manager) was deployed at NCAR and LBNL/NERSC. BeStMan is based on the standard specification v2.2 (under Open Grid Forum – OGF). The NCAR deployment provides cache management for their gateway and MSS. The NERSC deployment provides access to their High Performance Storage System (HPSS) as well as cache management for file staging from HPSS. Extensive tests were conducted on both installations, and include support of multiple concurrent requests, each for thousands of files. It was necessary to customize the NCAR installation of their deep-storage-access MSS, since this requires passwords for reading and writing as well as project id. Work on the new ESG “Gateway” prototype

testbed is using these deployed BeStMan systems. Testing of the “Gateway” in its use of BeStMan is also in progress, and a plan for deploying BeStMan for ORNL’s HPSS is underway.

There is a need for BeStMan to work with Globus 4.2 (GT4.2), since BeStMan uses httpg (grid-enabled http protocol) and GT4.2 no longer supports the httpg protocol (since it is not an official standard). Discussions are underway with the Globus team at ANL to develop a short-term plan to support BeStMan in GT4.2. In the long run, BeStMan will use https.

1.2.6 *PMEL Product Delivery Services Highlights*

Collaborative work continues as we build the foundation to create the “Product Service” based on the Live Access Server (LAS). The “Product Service” implements the LAS service request protocol (XML) for delivering “information products” to end-users and to other tiers of the ESG-CET system – typically visualizations, tables, and file subsets. During the past year, the request protocol was formally extended to convey data transformations that are essential to ESG-CET – notably regridding, differencing, and averaging. The “Product Service” can link various “back end services” into useful work flows: relational database queries; netCDF file IO; the PMEL-developed Ferret application and the LLNL-developed Climate Data Analysis Tools (CDAT) application for graphics rendering services, and the OPeNDAP-DAPPER protocol for access to collections of time series and profile observations.

In recent work, new protocols have been added to LAS for the handling of in-situ observations (OPeNDAP-ERDAP and a prototype OGC/Sensor Observation Service (SOS) implementation). A new cache management thread also has been added to the LAS server, in order to better ensure that the available disk resources of the server are used effectively. In collaboration with other ESG partners, LAS also has been enhanced so as to auto-configure using the ESG “Data Node Publisher” strategy – essentially an ability to tell in-operation LAS servers of new datasets that are available for the generation of products at distributed sites.

1.2.7 *ANL Security, Data, and Services Highlights*

The ESG-CET “Security” team gathered requirements, evaluated options and provided a federated solution for single sign-on capabilities in the ESG-CET project. We have completed an evaluation of OpenID and Shibboleth for single sign-on solutions. Information on the evaluation is available at <http://www.ci.uchicago.edu/wiki/bin/view/FranksProjects/OpenIDEvalESG> and <http://www.ci.uchicago.edu/wiki/bin/view/FranksProjects/ShibbolethEvalESG>.

Based on the team’s consensus, OpenID was chosen as the protocol to adopt and use in the federation. We have defined an OpenID protocol profile for the ESG-CET federation to ensure that some optional security options of the protocol are always used for authentication.

We evaluated open source libraries that implement OpenID protocol, and chose OpenID4Java as the software to leverage in the ESG-CET infrastructure. The library provides basic implementation of the protocol for use by both identity providers and relying parties. The ANL team has designed and developed extensions to the library to support important features such as pluggable framework for identity provider validation and a framework to push attributes as part of the authentication protocol. These changes have been successfully integrated with the ESG-CET “Gateway”, which now supports OpenID protocol. We are currently working with the OpenID4Java development community to contribute the changes to the code base.

For supporting single sign-on with non-browser clients, we investigated the use of MyProxy Online Certificate Authority (CA) to issue user credentials and assign client machines. We set up a MyProxy

server to demonstrate these features and worked with the OpenDAP team to leverage these credentials for accessing their server. We are currently working on utilizing the MyProxy feature to embed user attributes in the issued credentials for authorization. This solution will be integrated into the ESG “Gateway” software in time for the launch of the October prototype testbed.

The “Security” team is also exploring solutions for server-side processing of data and is investigating the use of tools such as LAS and Dapper. In collaboration with scientists from ANL's Mathematics and Computer Science Division, we have set up remote access to NCEP-2 data using LAS with a backend Hyrax server, and are currently gathering user requirements for backend processing of the data. We are also working with scientists from ANL's Environment Science Division to process station/point data and provide access to ARM data using LAS. Current infrastructure allows remote access to ARM data and derived aggregation data set via Dapper.

1.2.8 *ISI Monitoring, Data Catalogs, and Federation Highlights*

Highlights of the reporting period include the development and prototyping of a new version of the ESG-CET “Monitoring” system based on the most recent Globus Toolkit release, and also providing new features that will benefit ESG and allow improved monitoring of its infrastructure. These new features include an improved Trigger Service that detects failures and takes corrective actions when necessary. Other new features include a script to generate the ESG status page automatically, and an improved configuration of the monitoring infrastructure to improve robustness. This new “Monitoring” infrastructure has been tested on NCAR nodes, and ORNL is now in the process of testing the infrastructure.

The ISI team has led a series of discussions on the requirements and architecture of data catalogs and federation for the new ESG infrastructure. The goal of these discussions has been to explore architecture alternatives, understand tradeoffs, and clarify requirements.

2 Overall Progress

During this reporting period, progress was made in key areas that are necessary to meet ESG-CET objectives, goals and milestones. This section provides greater technical depth and presentation of the components needed for the prototype testbed.

2.1 ESG-CET Gateway

In the past six months, the “Gateway” team has made considerable progress towards the development of the next generation ESG-CET “Gateway”, both as a standalone data portal component and in interaction with a network of associated “Data Nodes”. Additionally, the ESG-CET collaboration, together with external partners like BADC, is converging on a final architecture for cross-Gateway federation, and for interoperability with other data centers. The goal is to first establish a prototype testbed in the autumn of 2008 that will allow concrete experimentation with available data services, scalability, and performance testing. Further development will be toward finalizing the delivery of CMIP5 (IPCC AR5) data by late 2009.

The following specific milestones were achieved since the last progress report:

- Metadata: The metadata object models, and the associated relational model for data persistence, have been considerably expanded to include both detailed information about models and simulations (in collaboration with the Earth System Curator project), and extensive descriptive metadata for observational systems. The database schema of the ESG-CET “Gateway” is now

composed of more than 100 tables, and its relational constraints, trigger functions, and table additions are still under active development.

- Data Browsing: The user interface and underlying services for browsing data collections have been considerably upgraded, including the capability to browse data along multiple axes (by project, by experiment, by topic etc.), to display additional informational metadata, to execute administrator-level data operations, and to view associated metrics.
- Data Publishing: Preliminary infrastructure was developed for ingesting data into the system by parsing metadata information stored in THREDDS XML catalogs. Additionally, this functionality was exposed to remote web service clients, and successfully invoked by the “Data Node Publisher” application. “Security” was provided by certificate-based mutual authentication between the remote client and the “Gateway”. Current development work is focusing on improving the robustness and performance of the publishing operations, as well as implementing data un-publishing functionality.
- Deep Storage Data Access: In collaboration with the LBNL team, the “Gateway” software was integrated with the next generation of SRM software for deep storage data access (i.e., BeStMan). This work included developing services for submitting data requests to BeStMan via the new web-service based interface, as well as the user interface for generating, monitoring and displaying user requests. Some work still remains to be done in terms of integrating the “Gateway” and BeStMan software packages with the latest version of the Globus libraries.
- Authentication: Following the project decision to rely on the emerging OpenID protocol and technology for web-based Single Sign-On, the ESG-CET “Gateway” was instrumented as both an OpenID Consumer and an OpenID Provider (i.e., it was given the capability of acting as either party in the authentication process). Single Sign-On for web browser based access was successfully demonstrated for a system of two ESG-CET “Gateways”, and for an ESG-CET “Gateway” together with a prototype data portal set up by BADC. Additionally, in collaboration with the ANL team, the OpenID authentication process was augmented by the transmission of site-specific user attributes using the OpenID protocol for attribute exchange, and with the capability of using a white list of trusted identity providers. Currently the OpenID Provider functionality is being factored out as a standalone component, so that the ESG-CET “Gateway” could be deployed in conjunction with an external, site-specific Identity Provider.
- Authorization: Prototype services and web-based user interfaces (UIs) have been developed for assigning authorities to users and authorizing users to access data resources in a globally federated environment. These components will have to be re-factored somewhat as the Virtual Organization (VO)-level security architecture is being modified and finalized to allow interoperability with external partners.
- Semantic Search: The “Gateway” team actively collaborated with the Earth System Curator project to significantly expand the semantic-based capabilities of the ESG-CET “Gateway” framework for data discovery, model track back, and model intercomparison. In particular, a new user interface for faceted search was developed (modeled after the *mpace* prototype), that allows the user to select any number of interesting search categories, in any order, and to dynamically adjust the options available in each category based on the previous selections (see figure 2). The capabilities of the ESG-CET “Gateway” semantic framework were successfully utilized to set up a prototype portal in support of the 2008 NCAR Advanced Study Program Summer Colloquium on “Dynamical Cores of Atmospheric General Circulation Models” for comparing output from a large number of models in a set of predetermined experiments (with prescribed forcings, boundary conditions, transports, etc.).

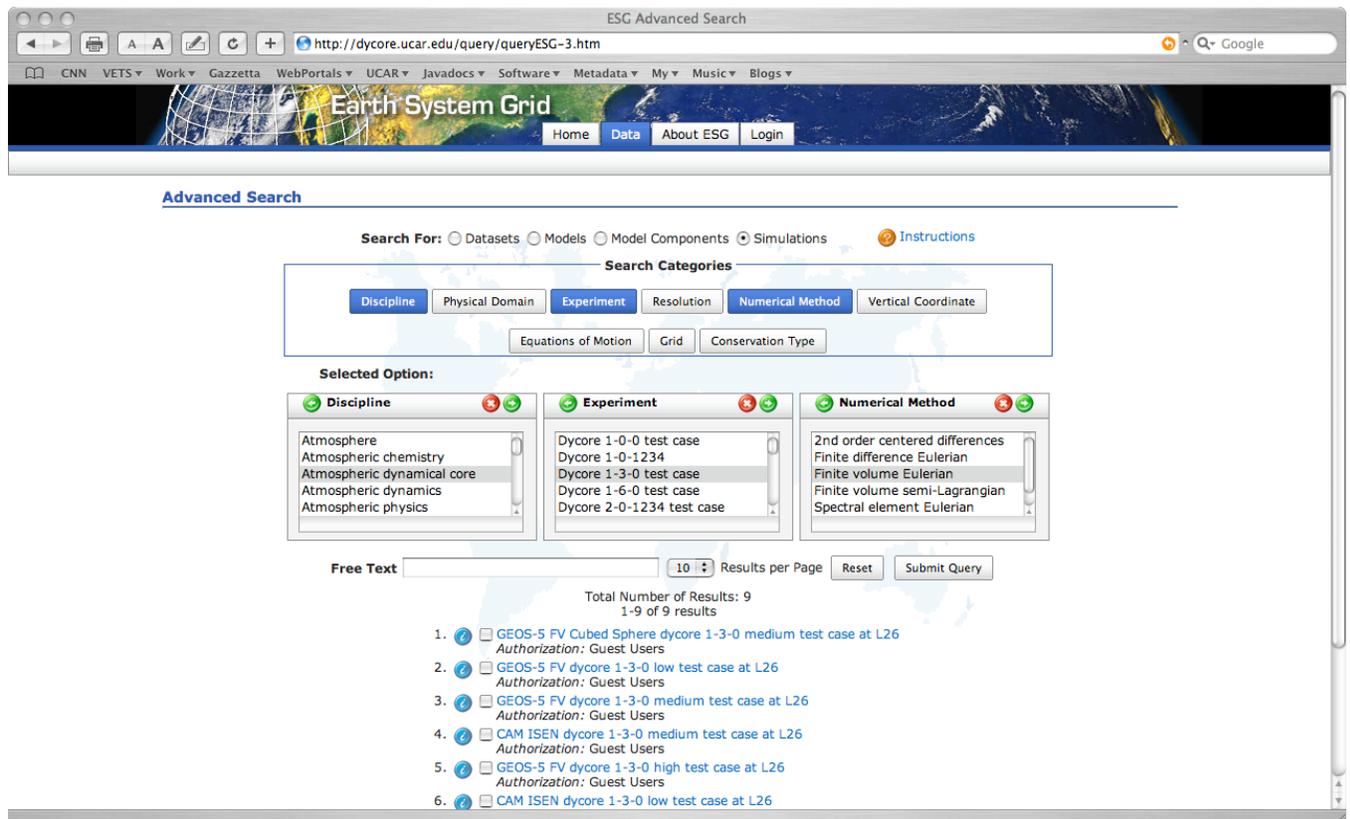


Figure 3: ESG-CET Webpage Showing Advanced Faceted Search Capability

- **Metrics:** A first version of application-programming interfaces (APIs) and services was developed to store and report “Gateway” metrics, including user logins, metadata searches and data downloads. Additionally, full click-streaming functionality has been integrated to track the user actions throughout the web portal session.

2.2 ESG-CET Data Node Publisher

The “Data Publisher” is the component of the ESG “Data Node” software that ingests data and metadata into the ESG catalogs. This publication process involves several steps, including: scanning data files and ingesting the information into the node database, gathering additional metadata not included in the data files, performing basic quality control, summarizing the metadata in one or more XML catalogs, notifying other node components that new catalogs have been created, and notifying the “Gateway” that new datasets have been added. A prototype publisher client has been developed and tested successfully on the ESG “Gateway” prototype testbed. The testing revealed a need to manage publication of multiple datasets in a single operation, as well as to query the node database to supply summaries of datasets already published.

Additionally, the critical interaction between the “Publisher” and “Gateway” publication web services has been designed and prototyped. For “Security” the “Publisher” supports SSL client authentication using short-lived X.509 certificates. The MyProxy credential management service will be used to generate and manage ESG certificates.

Interaction between the “Publisher” and other node servers has also been tested. While configuration of the node architecture is somewhat flexible, each node will likely include at least two components: a THREDDS Data Server (TDS) to serve individual files and aggregations of files, and a Live Access Server (LAS) to generate visualization and data products. When a new dataset is published, the servers are notified that a change has been made. The “Publisher” does this by invoking re-initialization services; a process has been demonstrated for the TDS.

Integrated into the backend of LAS “Product Services” is the Climate Data Analysis Tools (CDAT) that perform comparison, analysis, and visualization of large model output datasets, and that deliver results to climate scientists as customized end products accessed through a web browser. With CDAT analysis capabilities in LAS, climate scientists on the ESG grid can transparently access large datasets from disparate data centers and perform complex, configurable, ESG “Data Node” operations in a lightweight user interface environment.

Work also continued on the front-end graphical user interface (GUI). In the previous reporting period, we converged on an interface for publication of a single dataset. Since then, we have converged on an interface that will allow the publication of multiple datasets and the ability to operate on single or multiple datasets. At this point in time, the GUI form of the publisher has the ability to generate THREDDS catalogs and is friendly to use. Testers were able to understand the interface and run it in a few short minutes with very little instructions. For the prototype testbed, a query interface and a data provider authentication under construction.

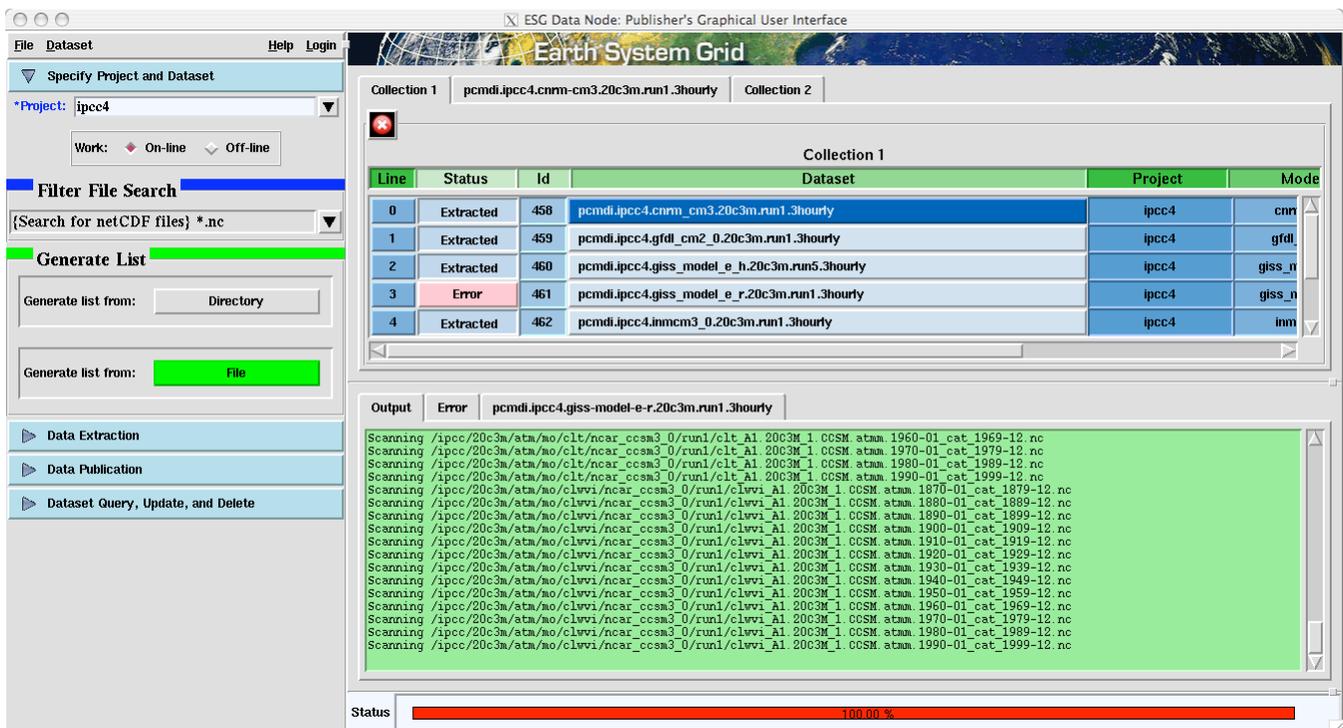


Figure 4: ESG-CET Data Node Publisher GUI Front-end Showing a Collection of Datasets and Scanned Extraction Results

2.3 ESG-CET Cyber Security

Secure access to data and resources plays a crucial role in the ESG. The “Security” model must safeguard data, resources, and the credentials of both users and services—but without imposing an undue burden on users. Finding the right balance between the required “Security” level of the overall system and its practical utility is a challenge.

In addition, the scope of ESG continues to enlarge with the requirement to federate additional national and foreign sites (e.g., GFDL, BADC, and the University of Tokyo Center for Climate System Research). The use cases associated with this federation translate into a requirement for a Single Sign-On solution for the browser clients as well as the web service and GridFTP clients.

The overall ESG “Security” architecture must be flexible enough to accommodate site-specific needs of individual groups, as well as the general infrastructure needs. After working on the Single Sign-On solutions, we are currently focusing on the distributed authorization infrastructure, where the authentication services, the attribute services and the policy enforcement points may reside in different administrative domains. The use cases that drive the associated requirements are developed in lockstep with our BADC collaborators to ensure future integration.

2.4 Product Services

The ESG-CET is intended to serve customers of a broad spectrum of sophistication. These users span numerical modelers who want access to “raw” model output files and verbatim subsets of model output, to climate impacts investigators who want rapid access to these data without the complexities of model-specific coordinate systems, to those users who only want to quickly visualize the overall behaviors of models. The petascale nature of the ESG data holdings requires that significant levels of data reduction take place at the server in order to satisfy these customers – both through straightforward subsetting and decimation, as well as through specific analysis operations such as the computing of spatio-temporal averages. In the ESG architecture, we refer to the steps that convert raw data into analysis results and visualizations as “Product Services”.

As described in section 1.2.6, the Live Access Server (LAS) has been extended into a generalized workflow engine for the creation and delivery of ESG products. A service-oriented approach in which “Back-End Services” are accessed via Simple Object Access Protocol (SOAP) has been employed in order to make the architecture adaptable to the range of products that it must provide. New output product capabilities were added including dynamic (adaptive, zoomable) Google Earth[®] visualizations, on-the-fly custom animations, and a “slide sorter” tool that allows scientists to inspect and difference gridded fields.

Recent work has been devoted to automate configuration procedures that allow ESG “Data Nodes” to publish newly available datasets to their LAS servers. For LAS to communicate with its AJAX-based user interface components, detailed metadata of datasets are required – variable names, units, space/time domains, and grid-relationships. All of this must be passed through the publication process to LAS servers without interrupting the services they are providing.

2.5 DataMover-Lite (DML)

Currently DML communicates to the “Gateway” only in order to download files using Lightweight Authorized HTTP File Server (LAHFS) based http. For future version, it is desirable that DML communicates directly with the SRM at the “Gateway” or directly with SRM or GridFTP servers on “Data nodes”.

Work has started on DML communication with SRMs for direct data downloading. DML will use the SRM standard interface to communicate with SRM servers based on the user request after the ESG “Gateway” authenticates and authorizes the user for data access. Files downloads will use GridFTP protocol in addition to LAHFS based http protocol.

2.6 BeStMan Java Client API and the Storage Resource Managers (SRM)

Integrating BeStMan into the new ESG “Gateway” for communicating with deep storage repositories was accomplished by developing of a Java client API for the ESG “Gateway” based on the standard SRM specification. Some dependent Java client libraries had to be updated to work with the Spring Framework of the ESG “Gateway”. The ESG “Gateway” team now uses this Java client API to communicate with BeStMan servers.

There recently were many users inquiries about why they had to wait for a long time until their requests were filled, and support-emails sent, regarding unresponsive requests. In analyzing the SRM logs, we identified the problem due to many files that were still in the disk cache even though they could be removed. This problem occurred because of increased usage of ESG to download files, an indication of growing interest in ESG. To overcome this problem in the short-term, the ESG portal disk cache was updated from 300GB to 2.5TB for SRM to manage and serve ESG users. This problem will be further mitigated in the next version of ESG by taking advantage of the automatic release of files by “DataMover-Lite” DML, or explicitly by the user as soon as files are transferred to the client.

Because of wrong information about the file sizes, in the metadata catalog, users also reported data downloading errors. Using the log information of the SRMs, we were able to identify this problem. Republishing the data into the metadata catalog to the SRM using the command ‘srm-ls’ provided the correct sizes of the files, thereby fixing the problem.

2.7 Data Catalogs

Support for the Replica Location Service (RLS) catalog servers that are deployed in the current ESG II infrastructure is ongoing. During the last six months, this support included extensive work to improve the reliability of RLS servers that operate at large scale. A test environment for a 64-bit platform, which should benefit ESG sites like LANL that run RLS on a 64-bit architecture, has been deployed.

2.8 Monitoring

The “Monitor” team continues to provide service infrastructure that allows ESG to detect and repair component failures. These monitoring services are essential for the reliable operation of the ESG portals and services. When failure messages occur, the “Monitoring” team helps to identify the causes of failures in the ESG infrastructure.

In the last six months, ESG “Monitoring” support included several efforts:

- Working to understand monitoring requirements in the new ESG architecture and, to this end, lead discussions of future ESG “Monitoring” needs at ESG weekly AccessGrid (AG) Sessions.
- Developing and deploying a new prototype version of the ESG “Monitoring” system that uses the latest release of the Globus Toolkit (GT 4.2). This release includes a new “Trigger Service” that is easier to maintain. The “Trigger Service” checks for specific error conditions and performs actions when those conditions are detected.
- Developing a script to automatically generate an ESG status page.

- Site-specific file format configurations for both aggregators and triggers were added for additional robustness in their deployment. This configuration file format is defined with the greater flexibility that will be needed within a “Gateway” site.
- A new ESG II monitoring infrastructure was developed and deployed on the NCAR CCSM portal and nodes in order to receive better feedback information.
- Plans are underway to deploy the new monitoring infrastructure on ORNL’s ESG II node. The “Monitor” team is currently investigating the new Globus stack, and is identifying a machine on which to run the new monitoring infrastructure.

2.9 Packaging of ESG-CET

One of ESG-CET’s priorities has been on software packaging and management in anticipation of widespread deployment of the ESG “Gateway” and “Data Node” software stacks as the new software goes into production. We are considering the Pacman package management system used by the Open Science Grid (OSG) for their Virtual Data Toolkit (VDT) grid software distribution, and have had significant interactions with OSG throughout this work. One obstacle is that VDT is based on Globus 4.0, whereas a number of developments in the ESG software stack are dependent on Globus 4.2. The VDT team plans to update their distribution, but without yet having a specific timeline. We have not yet brought our findings to the full ESG team for discussion, but our assessment is that there are significant benefits to be gained from adopting Pacman, using the VDT distribution as our infrastructure wherever possible, and using the OSG’s Build & Test Facility to evaluate the ESG software stack on a broad spectrum of platforms.

2.10 Data Transport Services

The High Altitude Observatory (HAO) effort continues along the theme of high performance data access and transport using the OPeNDAP Hyrax software infrastructure with ESG-specific customizations supporting the ESG “Gateway” and “Data Node” components of the overall architecture. This support includes responding to other parts of the architecture such as the “Product Services” and the adopted “Security” implementation. HAO participates in all-hands meetings and regular AccessGrid sessions as well as smaller group tasks.

HAO has made the following progress on the following components since the last reporting period:

Since ESG uses the Back-End Server (BES) portion of Hyrax, we have created the Point-to-Point Transmission protocol C Application Programming Interface (PPTCAPI) specifically for use in the Data System Interface (DSI) module. PPTCAPI is in the community OPeNDAP BES software repository and will be added to the standard BES releases. While working on PPTCAPI we also improved the PPTClient class for memory efficiency, which is required for the ESG data volumes. The BES was also extended to support multi-file access, thus re-enabling the ESG aggregation capability. This work completes the first major task for the HAO effort, which was to re-implement the OPeNDAPg (from ESG II) capability as a combination of community released OPeNDAP Hyrax software with a few extensions that ESG requires (e.g., aggregation, security). The final documentation and packaging for this software will be completed early in the next reporting period, and the release will be made available to ESG team members as well as selected community collaborators. Specific features include:

- High Performance Architecture for Data Operations: Remote netCDF Invocation (RNI). To satisfy both use cases for ESG and performance and scalability non-functional requirements, HAO has developed a flexible architecture for aggregating, operating upon, and transporting

very large data objects. Since ESG uses the netCDF Markup Language (NcML) for data and metadata, RNI allows any client to operate on these virtual datasets using a familiar netCDF API but the location of the operations can be on local files, netCDF files, or OPeNDAP objects which also can reside on local, or many (and distributed) computers. Further the operations can be close to the data or close to the application in a user-transparent manner.

- GridFTP Protocol Support in the OPeNDAP Hyrax and Client Software: HAO has re-implemented the GridFTP support as a transport layer for the OPeNDAP Hyrax software – two external groups are testing (see the Collaborations Section for more details). The intention is to make this a capability for ESG internal data transport, an implementation that fully supports the available Grid security infrastructures (i.e., certificates).
- “Product Services” Support: HAO is working with the ESG “Product Services” group to design an implementation of a BES command set and data handler for the I/O Service Provider (IOSP) (Unidata's THREDDS, Java-based I/O API) and general Java plug-in application support for the OPeNDAP BES. This capability will enable various product servers, such as the Ferret Data Server (FDS) to be plugged into multiple instances of the BES and to operate in parallel. HAO has tested Java-Virtual-Machine (JVM) techniques for this purpose, and will continue to work on this in the next reporting period.
- “Security” - Authentication and the Middle Tier of the OPeNDAP Server Architecture: HAO worked with the “Security” team to implement the ESG Test MyProxy installation. HAO created a DSI that authenticates connections against the Argonne MyProxy and uses PPTCAPI to talk to the RNI module running on a BES. HAO added a new connector class to the RNI client library that uses the gridftp_client API to talk to the DSI module. This software will be released for testing in the next reporting period.
- BES Authorization: HAO examined the NCAR-developed API for authorizing requests, and completed a prototype authorization query using the Axis/C APIs for the RNI BES module. This work will be completed early in the next reporting period, and an end-to-end verification of the new capability will be available for ESG use.
- Security of the Code Base: In conjunction with the OPeNDAP Hyrax core development team and with input from the NOAA Computer Incident Response Team (CIRT), HAO has also performed security audits and solutions to the code base being developed for the ESG specific implementations. The developed code will be security certified for government web application installations (since NOAA uses the National Institute of Standards and Technology (NIST) government type certification procedures).

3 Architectural Design Diagrams, Requirement Documents and Use Cases

For those with “access” to the ESG-CET internal website, all architectural design and requirement and use-case documents can be viewed at <http://esg-pcmdi.llnl.gov/internal>.

4 ESG-CET Group Meetings

The ESG-CET executive committee holds weekly conference calls on Tuesday at 10:00 a.m. PDT. At these meetings priorities and issues are discussed that make up the agenda for the weekly project meetings held on Thursdays at 12:00 p.m. PDT via the AccessGrid (AG). At the AG meetings, the entire team discusses project goals, design and development issues, technologies, timelines, and milestones. Given the need for more in-depth conversation and examination of work requirements, the following face-to-face meetings were held during this next reporting period:

4.1 ESG-CET Executive Meeting (April 1 – April 3)

The ESG-CET executive committee convened at the Office of Advanced Scientific Computing Research (OASCR) Computer Science Principal Investigator Meeting held in Denver, CO to discuss the upcoming ESG-CET All-hands meeting. Discussions also included project management, technical direction, collaborations, and overall project direction.

4.2 ESG All-hand Meeting (April 29 – May 1)

In late-April and early May, the ESG-CET team convened over a three-day period for the ESG Prototyping All-hands Meeting held at the National Center for Climate Research (NCAR) in Boulder, CO. We discussed the future ESG-CET interactions with science stockholders and subproject collaborators, and planned a path forward to meet the end-of-year ESG-CET prototype testbed deadline. The specific meeting topics included gap analysis, virtual organization authorization, federation, publishing, product services, security, metrics, and metadata replication.

4.3 Executive Meeting (July 13 – 16)

The ESG-CET executive committee convened at the SciDAC 2008 Conference, held in Seattle, WA to discuss the conference proceedings and conference paper. In addition to meeting with SciDAC and BER Program Managers, the overall state of ESG-CET was discussed.

4.4 PI Meeting (GO-ESSP) Workshop (September 16 – 19)

At the Global Organization for Earth System Science Portal (GO-ESSP) Workshop, ESG-CET PIs Don Middleton and Dean Williams, discussed the state of ESG-CET and planned for future direction of the project (i.e., upcoming project review, prototype testbed, current and future collaborations, papers and presentations, etc.).

5 Collaborations

To effectively build an infrastructure capable of dealing with petascale data management and analysis, we established connections with other funded DOE Office of Science SciDAC projects and programs at various meetings and workshops, such as the SciDAC 2008 Conference held in Seattle, WA. In particular, collaborations were established with the following groups:

5.1 University of Utah Visit to Discuss VisTrails, ESG, and CDAT Collaboration

As part of the SciDAC VACET collaboration effort, Dean Williams visited the Scientific Computing and Imaging (SCI) Institute at the University of Utah to discuss the possibilities of VisTrails in the ESG framework. (VisTrails is an emerging paradigm for capturing complex analysis processes at various levels of details and provides provenance information necessary for reproducible results in a shared environment.) The LLNL group was able to install VisTrails and implemented a small subset of CDAT commands callable from within the VisTrail visual environment. Although CDAT has been developed with climate applications in mind, most of the concepts of the system can be used to develop a more general data analysis tool by merging the functionalities of CDAT and VisTrails.

VisTrails Spreadsheet

CDAT Standard Window

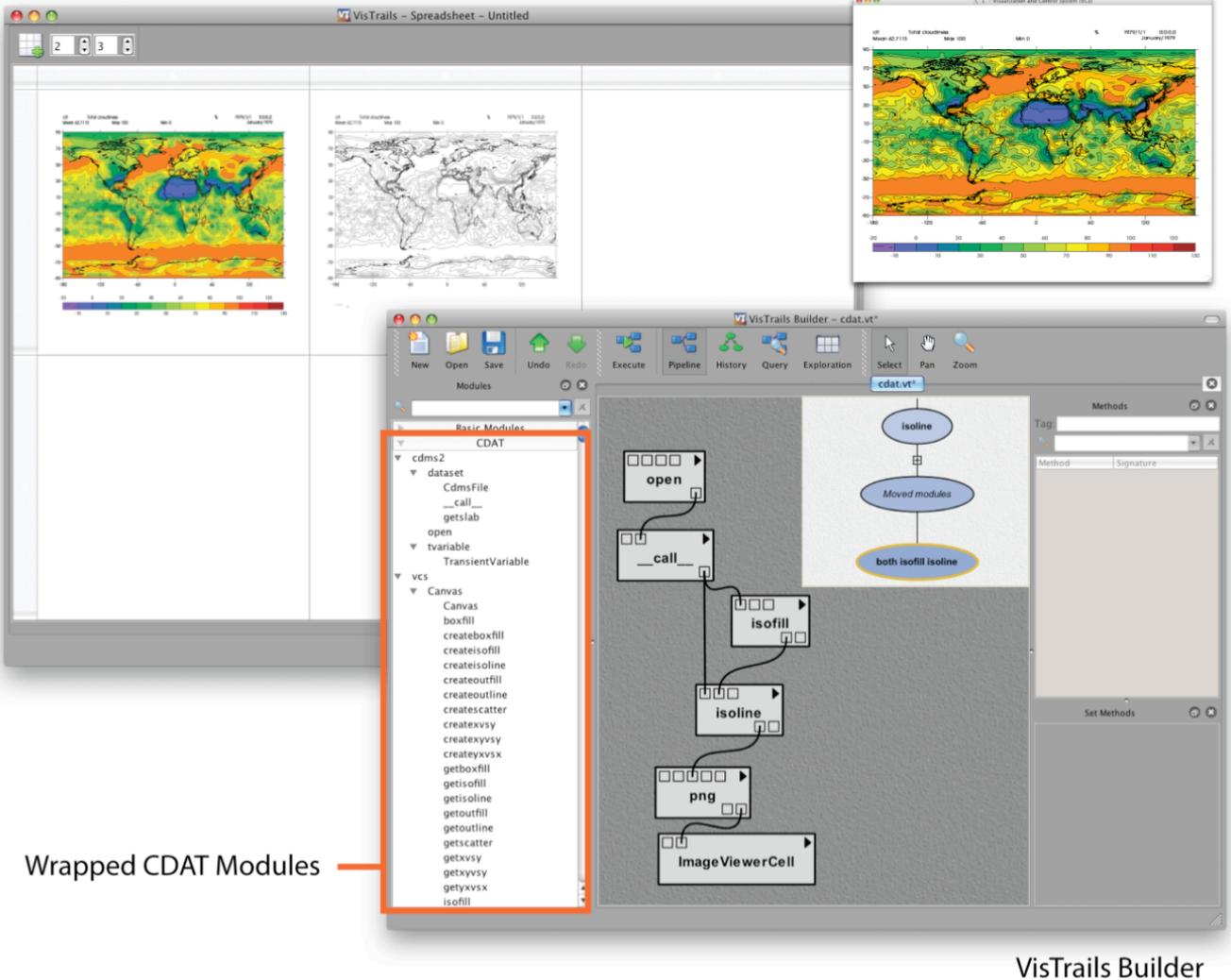


Figure 5: Example of CDAT Workflow Built Inside VisTrails

5.2 Collaboration with the SciDAC VACET Visualization Team

During the 2008 SciDAC conference the VACET team demonstrated the new 3D visualization system add-on to the 5.0 beta release of the CDAT package to John Drake and Phil Jones. Both were enthusiastic about the 3D capabilities, their tight integration into the existing pipeline, and especially the flexible annotations offered by the new system. John and Phil re-iterated their concerns about the high spatial resolutions of upcoming simulations and the inability of existing tools to adequately handle these large data sets. All parties agreed that the ViSUS package and its ability to process tera-scale datasets in a streaming and/or out-of-core fashion is an important and welcome addition to the CDAT package.

Going forward, John and Phil suggested two climate-specific extensions for the current tool chain: Mapped z-elevations and mapped logical grids. Mapped z-elevations are important for climate visualization since most common climate simulation codes operate on slices of constant air pressure rather than constant elevation. While both measurements are related, they are not identical and their

mapping must be taken into account for an accurate visualization. The current version of the ViSUS system already has the capability to display height fields, as mapped 2D planes, and this capability will be extended to support general elevation mappings. While older simulation codes used polar grids with discontinuities at both geographic poles, newer packages use warped polar grids that move the singularities to less significant locations (e.g., into the middle of Canada), or more generally mapped grids.

While these structures logically remain regular grids geometrically they can be mapped almost arbitrarily onto the earth depending on the specific needs of a given simulation. Multiple options to support mapped grids in the ViSUS system were discussed with higher dimensional coordinates being the clear favorite. This concept entails storing the geometric locations of samples along with the standard physical variables (pressure, temperature, etc.). Since the grid logically remains regular, the "enhanced" dataset can easily be handle by the existing streaming infrastructure. The mapping can now be seen as a special case of multi-modal visualization where, for example, temperature is one field and x,y,z, positions are a second field. Custom rendering code will then combine both fields to create a mapped temperature field.

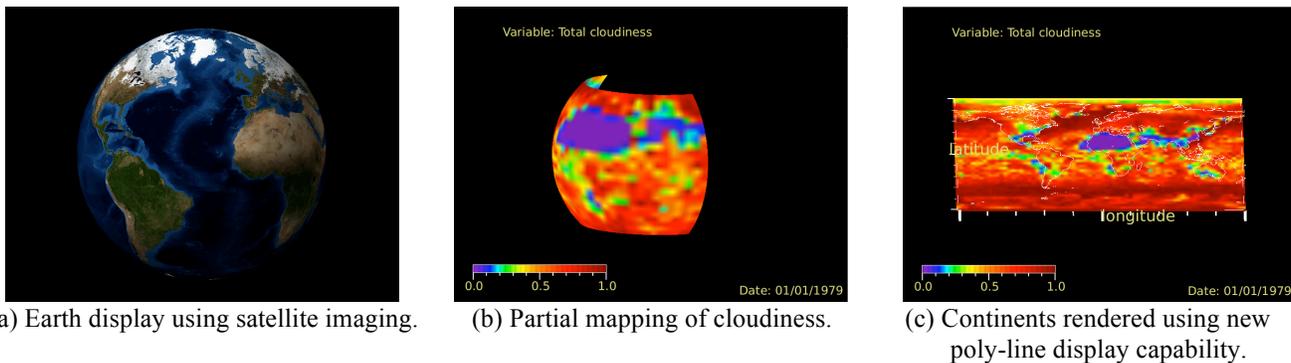


Figure 6: Example 3D Capabilities Provided by the VACET Team

One of the first new features that the VACET team will introduce is the potential management of multiple meshes with different domains and resolutions that may be incompatible. In particular, the VACET team will work on the visualization of atmospheric and ocean data that are generated on independent domains.

5.3 Collaborations with the Southern California Earthquake Center (SCEC)

Ann Chervenak initiated discussions on possible collaborations regarding metadata schemas and catalogs with the ESG-CET and the Southern California Earthquake Center (SCEC) project. She also led phone meetings between the two projects. Participants from SCEC included Thomas Jordan (PI), Phil Maechling, David Okawa and Tran Huynh – all from the University of Southern California’s Information Sciences Institute (USC/ISI).

5.4 Scientific Data Management (SDM) Center for Enabling Technology (SciDAC CET)

Based on their experience with DataMover-Lite, Arie Shoshani and his team developed a new client version of an SRM (called “SRM-Lite”) in order to move files to and from sites that have highly secure systems by using one-time-passwords. This tool is similar in design to DML, but runs as a client version of SRM to pull/push files from behind a firewall. Unlike DML, SRM-lite cannot use a GUI remotely, and therefore has only a command-line interface. For example, Chi-Fan Shih of NCAR’s Data Support

Section used this tool to move ESG data from a disk system inside the firewall to one outside the firewall. SRM-Lite, based on high-performance scp (hpn-scp) for higher throughput, is being used successfully to move the large volumes of data.

5.5 Center for Enabling Distributed Petascale Science (CEDPS) (SciDAC CET)

Collecting statistics on data movement and storage usage is very important to ESG. In order to collect such data from BeStMan, Arie Shoshani’s team collaborated with members of the Center for Enabling Distributed Petascale Science (CEDPS) project to extract such data automatically from the SRM logs. They have worked with Dan Gunter and Keith Beattie at LBNL for CEDPS troubleshooting to collect logging information into DB, display with NetLogger and web-frontend. They also plan to use their statistics collection tools to automatically generate summary statistics and help with troubleshooting.

5.6 SBIR CDAT-MODAVE – Collaboration with Tech-X

Dean Williams is working with Alex Pletzer (from Tech-X) and V. Balaji (from the Geophysical Fluid Dynamics Laboratory, GFDL) on the Mosaic Data Analysis and Visualization Extension (MoDAVE). An SBIR-funded project, MoDAVE started in July 2008 with the aim of extending CDAT to handle the large multi-block datasets that are increasingly becoming the norm, as atmospheric models move away from longitude/latitude based grids. Since its inception, MoDAVE has focused on developing a cubed-sphere netCDF reader using CDAT's Climate Data Management System 2 (cdms2) object to infer the connectivity between mosaic tiles. Knowledge of the tile connectivity is necessary in order to accurately visualize cell-centered data on the cubed-sphere (finite-volume-based discretization leads to cell-centered scalar fields). A general API (mvViz3d) was defined to allow multiple 3D visualization engines to be plugged into CDAT. At present mvViz3d supports two implementations: one based on the Visualization Toolkit (VTK) and the other on VisIt (www.llnl.gov/visit/); both can be used to render 3D iso-surfaces on the sphere. In addition, we will continue to explore ways to parallelize data analysis and visualization so as to be able to handle high-resolution (< 10km) datasets. CDAT is one of several standard analysis tools that are needed to operate on ESG’s “Product Services” back-end.

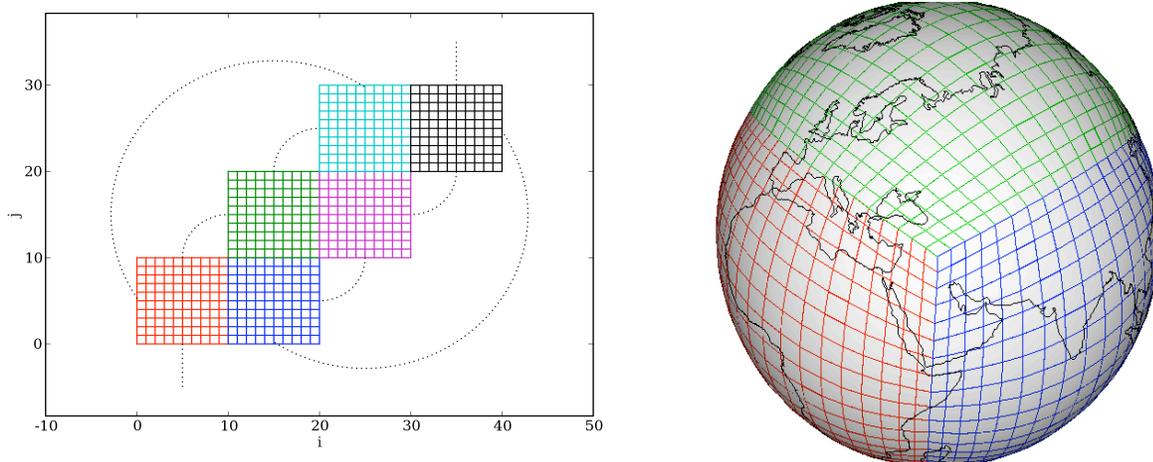


Figure 7: Example of a Mosaic Grid: Cubed Sphere Connectivity

5.7 Global Organization for Earth System Science Portal (GO-ESSP) Workshop

Steve Hankin (this year's workshop host), Don Middleton, and Dean N. Williams are three of seven GO-ESSP steering committee members who coordinated the seventh annual GO-ESSP workshop held September 16 – 19 at the [Seattle Washington Public Library](#). In addition, Steve, Don, and Dean chaired six of the seven workshop sessions. The GO-ESSP workshop focuses on facilitating the organization and implementation of an infrastructure for full data sharing among a consortium spanning continents, countries, and intergovernmental agencies. All ESG-CET testbed partners (i.e., LLNL, NCAR, GFDL, BADC, DKRZ, and the University of Tokyo) were present. The workshop, in part, covered testbed security concerns and addressed issues of collaboration. By 2011, this organization envisions allowing users open access to petabytes of multi-model generated data, as well as in-situ, satellite, biogeochemistry, and ecosystems data.

5.8 Hybrid Coordinate Ocean Model (HyCOM) consortium (NOAA, Navy, et. al.)

NOAA/PMEL (Steve Hankin) is a partner in the Hybrid Coordinate Ocean Model (HyCOM) consortium [<http://hycom.rsmas.miami.edu/>]. The HyCOM Consortium, which has developed a high-resolution (1/12 degree) operational, global ocean modeling capability under cooperative US Navy and NOAA funding. The HyCOM model presents unique technical challenges, such as the complicated coordinate system that it employs and its large data volumes, but the needs of HYCOM overlap in many respects with the ocean components of the climate models to be utilized in CMIP5 (IPCC AR5). A significant and productive two-way technology transfer has been developed in support of ESG-CET, and of HyCOM in particular.

5.9 NOAA Geophysical Fluid Dynamics Laboratory

The NOAA GFDL Fluid Dynamics Laboratory is an active contributor to CMIP5 and an active participant in the ESG-CET. V. Balaji [Head, GFDL Modeling Systems Group] is a frequent participant and active contributor in ESG-CET telcons and meetings leading to a vigorous bi-directional exchange of ideas and technology. NOAA/PMEL (Steve Hankin) shares a Memorandum of Understanding (MOU) with GFDL for the development of the Laboratory's data portal, also leading to an active two-way technology transfer between NOAA and ESG-CET.

5.10 NOAA Integrated Ocean Observing System (IOOS) and NOAA Office of Climate Observations (OCO)

NOAA has been designated as the lead agency in the development of the U.S Integrated Ocean Observing System (IOOS). PMEL is a member of the IOOS Integrated Products Team (IPT) and is also the developer of the ocean Observing System Monitoring Center (OSMC) on behalf of NOAA/OCO. Through the PMEL membership in the ESG-CET a number of useful collaborations benefits are being explored and are likely to be realized in time for the CMIP5 (IPCC AR5) work. IOOS and OSMC are both sources of integrated ocean observations that are potentially useful to IPCC scientists in the evaluation of climate model outputs. PMEL will be helping to bring these collections of observations into the ESG-CET framework for the benefit of IPCC scientists and others.

5.11 NASA JPL visit to LLNL to discuss ESG-CET and Satellite Data

Dan Crichton and Amy Braverman from NASA's Jet Propulsion Laboratory (JPL) are actively pursuing initiatives in climate modeling, and so are interested in establishing an active relationship with ESG. Their recent visit (along with other JPL scientists) to LLNL explored how to enable the climate

community to gain access to JPL satellite data, and allow modelers at JPL to gain access ESG data. JPL is now engaged in a pilot activity that demonstrates access between JPL and LLNL and, if successful, will encourage deeper collaborations between NASA and ESG-CET.

5.12 External Support of Grid-Enabled OPeNDAP

Karen Schichardt and Jeff Daily of PNNL, who are working with Dr. David Randall (CSU) on an NSF STC for extreme weather modeling, also are collaborating with the Data and Transport ESG-CET team. OPeNDAP efforts at NCAR, security infrastructure, and GridFTP protocol developments now are helping this group integrate their efforts into the ESG-CET environment.

6 Outreach, Papers, Presentations and Posters

Outreach activities, papers, talks, and posters presented during this time period:

6.1 Outreaches Activities

6.1.1 *World Meteorological Organization (WMO) Information System Intercommission Coordination Group (ICG-WIS)*

Don Middleton serves as an expert advisor to this high-level coordination group for the development of next generation globally federated data systems for climate, weather, hydrology, oceanography, and other U.N./WMO program elements. Middleton attended a meeting of this group in August in Brazilia, Brazil and verbally briefed the council on ESG-CET activities and implications related to the next IPCC process and related federated data systems.

6.1.2 *DAARWG Workshop in Ashville, NC*

Dean Williams accepted the invitation to join the NOAA Advisory's Board Data Archive and Access Requirements Working Group (DAARWG), an important working group of NOAA's Science Advisory Board (SAB). DAARWG evaluates data archiving and access requirements from all of NOAA's observing systems and computational models, as well as from relevant non-NOAA sources. Its charter is to provide scientific advice and broad direction regarding the wide range of data, information, and products that NOAA should archive, and ways in which this agency can best provide access to them.

6.1.3 *Global Earth Observing System of Systems (GEOSS)*

GEOSS is a primary thrust of the Group for Earth Observations (GEO), which is a United Nations program coordinated under the auspices of the World Meteorological Organization (WMO). Don Middleton attended several days of GEOSS meetings in Boulder, CO. in September 2008. Middleton also gave an invited presentation entitled "Global Federated Climate and Weather Data Systems" and focused on ESG-CET in the area of climate and the THORPEX Interactive Grand Global Ensemble (TIGGE) in the area of weather.

6.1.4 *Steering Committee for HPDG08*

Peter Fox is a member of the Steering and Programme Committee for HPDG08 (High Performance Data Grid 2008) to be held in Dunedin, New Zealand Dec 3-4, 2008.

6.1.5 *EGU Session Convener*

Peter Fox convened a European Geophysical Union meeting session on "The Grid for Geosciences and Geoscience Applications for the Grid".

6.1.6 *ESIn Editor*

Peter Fox is an *Earth Science Informatics* (ESIn) editor for a special issue on “Grid computing for Geosciences, Earth Science Informatics”.

6.1.7 *Ocean Observatory Initiative*

Ian Foster accepted an invitation to join the Advisory Committee for the NSF Ocean Observatory Initiative. He attended his first meeting in Dallas Texas and was able to discuss opportunities for joint work with ESG-CET.

6.2 Papers:

6.2.1 *Paper and Presentation: SciDAC '08 Conference in Seattle, WA*

Dean Williams presented a PowerPoint presentation on ESG-CET, titled, “Data Management and Analysis for the Earth System Grid”. The presentation covered the scientific challenges facing the climate community in terms of data management and analysis of hundreds of petabytes simulated data.

In addition, the ESG-CET team presented a paper to the SciDAC 2008 conference proceedings, with the citation being: D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, S. Bharathi, D. Brown, M. Chen, A. L. Chervenak, L. Cinquini, R. Drach, I. T. Foster, P. Fox, S. Hankin, V. E. Henson, P. Jones, D. E. Middleton, J. Schwidder, R. Schweitzer, R. Schuler, A. Shoshani, F. Siebenlist, A. Sim, W. G. Strand, N. Wilhelmi, M. Su, 2008: Data management and analysis for the Earth System Grid in the *Journal of Physics* Conference Series, SciDAC 2008 conference proceedings.

6.2.2 *Paper in the Bulletin of the American Meteorological Society (BAMS)*

An article by the ESG-CET team follows on the Meehl et al. 2007 *BAMS* article on *the CMIP3 multi-model datasets*:

D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, S. Bharathi, D. Brown, M. Chen, A. L. Chervenak, L. Cinquini, R. Drach, I. T. Foster, P. Fox, D. Fraser, J. Garcia, S. Hankin, P. Jones, C. Kesselman, D. E. Middleton, J. Schwidder, R. Schweitzer, R. Schuler, A. Shoshani, F. Siebenlist, A. Sim, W. G. Strand, and N. Wilhelmi, M. Su, 2008: The Earth System Grid: Enabling access to multi-model climate simulation data. *Bulletin of the American Meteorological Society*. (Scheduled for publication in January 2009.)

6.2.3 *SciDAC Review Article*

The ESG-CET team completed the SciDAC Review Article entitled, “The Planet at Their Fingertips: Climate Modeling Data Heats Up”. The article talks about the increasing importance of climate modeling and the tremendous need for the Earth System Grid to allow fast and accurate access to hundreds of petabytes.

The authors of the article include the entire ESG-CET team:

D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, S. Bharathi, D. Brown, M. Chen, A. L. Chervenak, L. Cinquini, R. Drach, I. T. Foster, P. Fox, S. Hankin, V. E. Henson, P. Jones, D. E. Middleton, J. Schwidder, R. Schweitzer, R. Schuler, A. Shoshani, F. Siebenlist, A. Sim, W. G. Strand, N. Wilhelmi, M. Su. “The Planet at Their Fingertips: Climate Modeling Data Heats Up”, SciDAC Review Journal. (Scheduled to appear in Winter Issue #10 of the SciDAC Review.)

6.2.4 *Earth Science Informatics Article*

Developing Service-Oriented Applications in a Grid Environment: Experiences Using the OPeNDAP Back-End-Server, 2008, J. Garcia, P. Fox, P. West. S. Zednik.

Submitted to Earth Science Informatics as part of the special issue on “Grid Computing for Geosciences”.

6.2.5 *Federation*

Robert Schuler and Ann Chervenak, wrote a federation document for the ESG-CET team entitled, “Search Metadata: Storage and Sharing Considerations”. In addition, they gave a talk on “Federated Metadata” at the ESG All-Hands Meeting in April 2008.

6.2.6 *SciDAC Review Highlights*

The ESG-CET team prepared an article for the SciDAC Review highlighting ESG-CET. Several of the SciDAC projects that reported at the SciDAC 2008 conference were asked to present a one-page summary of their project, together with an image and a caption.

6.3 Talks:

6.3.1 *Net@Edu Meeting*

Don Middleton provided the keynote address at the annual Net@Edu meeting in Tempe, Arizona in February 2008 (this was not reported in the previous ESG-CET report for that time period). Entitled “*Cyberinfrastructure and Emerging Scientific Data and Knowledge Systems*”, Middleton’s presentation highlighted ESG-CET as a primary focal point of the talk.

6.3.2 *North American Regional Climate Change Assessment Program (NARCCAP) PI and Community Meeting*

NARCCAP is a multi-agency (NSF, DOE, NOAA, EPA) regional climate modeling program that is closely affiliated with ESG-CET, and providing data services to the community using ESG cyberinfrastructure. Don Middleton contributed to the organization of this meeting and presented an opening talk entitled “*NARCCAP Data: The Earth System Grid (ESG), Data & Knowledge Systems, and a Few Useful Tools*”. We also held a hand-on training session for community members on using the ESG-based NARCCAP data system, and it was well received and very effective.

6.3.3 *Virtual Climate Change Research Institute Workshop*

Ann Chervenak gave a presentation on “The Earth System Grid: Turning Climate Datasets into Community Resources” at the Virtual Climate Change Research Institute Workshop held at USC on June 25, 2008. A group of paleoclimatologists is proposing a new NSF Science and Technology Center and hopes to collaborate with ESG-CET.

6.3.4 *World Meteorological Organization (WMO) ET-WISC Meeting*

In June 2008, Don Middleton attended a meeting in Darmstadt, Germany of the Expert Team on the World Meteorological Organization (WMO, a program of the United Nations) Information System (WIS) Global Federated Systems (ET-WISC). Middleton provided a presentation on federated scientific data systems included a briefing on ESG-CET for the WMO expert team.

6.3.5 *Washington trip to meet with Sponsors*

On June 27th, Dean Williams spent a day in Washington, D.C. with sponsors from the DOE Office of Science SciDAC and from the Office of Biological and Environmental Research (OBER). Dean presented information on ESG-CET and discussed the progress of this project. The meeting also provided an opportunity to introduce Dean and ESG-CET to Susan Turnbull, the new ESG-CET SciDAC Program Manager.

6.3.6 *ESnet Climate meeting at LBNL with Dan Hitchcock*

David Bernholdt contributed to a presentation by John Drake (lead PI for the SciDAC CCSM effort) discussing data volumes and networking needs anticipated for CMIP5 (IPCC AR5), presented at the BER-ESnet workshop on 4 August. Dean Williams was also present at this meeting and presented information on ESG-CET.

6.3.7 *BER Program Manager Wanda Farrell's visit to ORNL*

On August 6th, BER Program Manager Wanda Farrell visited ORNL. David Bernholdt contributed to a presentation by John Drake on modeling-related activities at ORNL, including ESG-CET.

6.3.8 *Presented ESG-CET to Christine Chalk*

On August 6th, BER Program Manger Christine Chalk visited LLNL. Dean Williams presented ESG-CET and discussions took place covering the use of ESG-CET for SciDAC's Biology & Environment scientific area.

6.3.9 *Presented ESG-CET to David Skinner and Horst Simon*

On August 13th, Dean Williams visited NERSC to discuss the ESG-CET and the possible need for more storage space on NERSC's long-term storage facility. In addition, future conversations were planned on NERSC as an ESG "Data Node" and as a compute server.

6.3.10 *BER Program Manager Anjuli Bamzai visit to ORNL*

On September 11th, BER Program Manager Anjuli Bamzai visited ORNL, where David Bernholdt gave a presentation on ESG.

6.3.11 *Global Organization for Earth System Science Portal (GO-ESSP) Workshop*

At the GO-ESSP workshop, the following presentations were presented by ESG-CET team members: Luca Cinquini - "The ESG-CET Gateway: A Distributed and Federated Architecture for Data Search, Access and Analysis", Nate Wilhelmi - "ESG-CET Demonstration", Eric Nienhouse - "The Challenges of Building Rich Web Clients Geo-Spatial Applications – Lessons Learned", and Roland Schweitzer - "Server-side OPeNDAP Analysis – Concrete steps toward a generalized framework via a reference implementation using F-TDS".

6.3.12 *Presentation at EGU 2008, IN09, The Grid for Geosciences*

Grid-enabled OPeNDAP Hyrax servers: parallel/multiple back-end services and remote netCDF invocation (RNI), by J. Garcia, S. Zednik, P. West and P. Fox.

6.4 Posters:

6.4.1 *SciDAC '08 PI Meeting in Denver, CO*

David Bernholdt, Ian Foster, Don Middleton, and Dean Williams (the ESG-CET executive team)

attended the SciDAC PI meeting held in Denver, CO, where Don and Dean presented a poster on ESG-CET. Besides providing a great opportunity for the executive to intermingle with other Advanced Scientific Computing Research (ASCR) computer science efforts, this meeting provided an opportunity to discuss topics in areas that may assist future ESG development, including performance tools, application development frameworks, data management/analytics, and visualization.

6.4.2 *European Geophysical Union Meeting*

In April of 2008, Don Middleton attended the annual European Geophysical Union (EGU) meeting in Vienna, Austria. Middleton presented a poster on ESG-CET, which was developed in concert with the ESG-CET Executive team.