LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-CONF-450053

Earth System Grid Center for Enabling Technologies: Building a Global Infrastructure for Climate Change Research

D. N. Williams, J. Ahrens, R. Ananthakrishnan, G. Bell, S. Bharathi, D. Brown, M. Chen, A. L. Chervenak, L. Cinquini, R. Drach, I. T. Foster, P. Fox, S. Hankin, D. Harper, N. Hook, P. Jones, D. E. Middleton, R. Miller, E. Nienhouse, R. Schweitzer, R. Schuler, G. Shipman, A. Shoshani, F. Siebenlist, A. Sim, W. G. Strand, F. Wang, H. Wilcox, N. Wilhelmi

August 17, 2010

SciDAC Conference
Chattanooga, TN, United States
July 12, 2010 through July 15, 2010

Disclaimer

Earth System Grid Center for Enabling Technologies:
Building a Global Infrastructure for Climate Change Research

The Earth System Grid Center for Enabling Technologies Team:

D. N. Williams[3,11], J. Ahrens[4], R. Ananthakrishnan[1], G. Bell[3], S. Bharathi[9], D. Brown[6], M. Chen[8], A. L. Chervenak[9], L. Cinquini[5], R. Drach[3], I. T. Foster[1,11], P. Fox[10], S. Hankin[7], D. Harper[6], N. Hook[6], P. Jones[4], D. E. Middleton[6,11], R. Miller[8], E. Nienhouse[6], R. Schweitzer[7], R. Schuler[9], G. Shipman[8], A. Shoshani[2], F. Siebenlist[1], A. Sim[2], W. G. Strand[6], F. Wang[8], H. Wilcox[6], N. Wilhelmi[6]

[1] Argonne National Laboratory, Chicago, IL, USA.
[2] Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
[3] Lawrence Livermore National Laboratory, Livermore, CA, USA.
[4] Los Alamos National Laboratory, Los Alamos, NM, USA.
[5] National Aeronautics and Space Administration, Pasadena, CA, USA.
[6] National Center for Atmospheric Research, Boulder, CO, USA.
[7] National Oceanic and Atmospheric Administration (PMEL), Seattle, WA, USA.
[8] Oak Ridge National Laboratory, Oak Ridge, TN, USA.
[9] University of Southern California, Information Sciences Institute, Marina del Ray, CA, USA.
[10] Rensselaer Polytechnic Institute, Troy, NY, USA.

[11] E-mail: williams13@llnl.gov, don@ucar.edu, itf@mcs.anl.gov

Abstract. Established within DOE's Scientific Discovery through Advanced Computing (SciDAC-) 2 program, with support from ASCR and BER, the Earth System Grid Center for Enabling Technologies (ESG-CET) is a consortium of seven laboratories (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], and Pacific Marine Environmental Laboratory [PMEL]), and two institutes (Rensselaer Polytechnic Institute [RPI] and the University of Southern California, Information Sciences Institute [USC/ISI]). The consortium's mission is to provide climate researchers worldwide with a science gateway to access data, information, models, analysis tools, and computational capabilities required to evaluate extreme-scale data sets. Its stated goals are to (1) make data more useful to climate researchers by developing collaborative technology that enhances data usability; (2) meet the specific needs that national and international climate projects have for distributed databases, data access, and data movement; (3) provide a universal and secure web-based data access portal for broad-based multi-model data collections; and (4) provide a wide range of climate data-analysis tools and diagnostic methods to international climate centers and U.S. government agencies. To this end, the ESG-CET is working to integrate all highly publicized climate data sets—from climate simulations to observations—using distributed storage management, remote high-performance units, high-bandwidth wide-area networks, and user desktop platforms in a collaborative problem-solving environment.

1. The Climate Data Challenge and the Earth System Grid

Climate change research requires a wide variety of data-management, -analysis and -storage resources. These data sources reside on distributed national and international data holdings, and large-scale processing is required to compare observational data, which is recorded at a very high sampling frequency, with simulation data. As data has become increasingly abundant and complex, it has also become increasingly difficult to transport and effectively process. For many scientific disciplines, the amount of data being generated, inspected, and studied ranges up to tens of petabytes, with estimates for the near future breaking the exascale barrier (ASCR Science Network Requirements, 2009). At these data sizes, transferring data is prohibitively expensive; instead, the simpler approach is to move the computation to where the data sets are stored. Based on this approach, the Earth System Grid (ESG) (Williams, D. N., 2009) has developed a system of Gateways and Data Nodes where stored data can be accessed and analyzed in place. Gateways allow climate researchers to gain entrance (e.g., via a Web portal) to data, information, models, analysis tools, and computational resources, while Data Nodes are located where the data resides and are responsible for providing services for access to data.

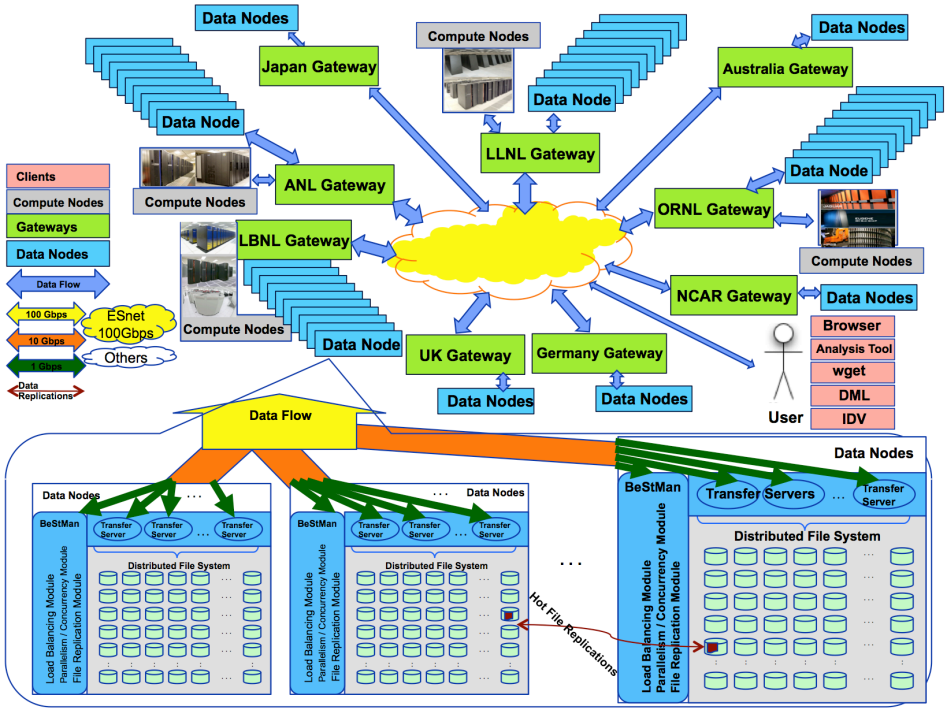

Figure 1: *The topology of the ESG enterprise system with network connections and computing resources, providing a network of geographically distributed gateways, data nodes, and computing resources in a globally federated, built-to-share scientific discovery infrastructure.*

Figure 1 shows ESG ongoing partnerships and established relationships with a variety of data, research, and technology efforts. To meet the needs of climate community, these collaborations will be augmented with additional observational and model-generated data sets, including: the Intergovernmental Panel on Climate Change (IPCC) Coupled Model Intercomparison Project, phase 5 (CMIP5), the Community Climate System Model version 4 (CCSM4), the Community Earth System Model version 1 (CESM1), the Carbon-Land Model Intercomparison Project (C-LAMP), the Parallel Ocean Program (POP), the North American Regional Climate Change Assessment Program (NARCCAP), the National Aeronautics and Space Administration (NASA) satellite observational data (CloudSat, and others), the National Oceanic and Atmospheric Administration (NOAA) observational data, and many other data sets critical to understanding climate change. In addition to data, ESG is

unifying computational and analytical climate resources under one unique powerful knowledge-discovery enterprise system and continues to evolve to maintain its reputation as a reliable resource for serving the climate-science community as it seeks to derive fundamental and practical understanding of climate and climate change.

2. Data Download Highlights

ESG has become a very important project for a number of climate initiatives (e.g., CMIP5, CCSM, NARCCAP, NASA observations – more listed above) representing important data, so in order to access and download data, ESG users are required to have an ESG account at one of the four ESG sites. Many types of data (e.g., CMIP5, CCSM, POP, NARCCAP, C-LAMP, AIRS, etc.) are available for free download at one of the four ESG Gateways located at: the Program for Climate Model Diagnosis and Intercomparison (PCMDI), http://pcmdi3.llnl.gov/esgcet/home.htm; NCAR, http://www.earthsystemgrid.org/; ORNL, http://esg2-gw.ccs.ornl.gov/esgcet/home.htm; and NASA, http://esg-gateway.jpl.nasa.gov/home.htm. The graphs below illustrate that use of the ESG continues to grow at a rapid pace as indicated over the last five years. Portals are recording data download volumes in the tens of terabytes (TB, where 1 TB is 1×10^{12} bytes) per month, and the cumulative total since January 2005 has now reached over one petabyte (PB, where 1 PB is 1×10^{15} bytes) of data downloaded. This certainly is a huge milestone!

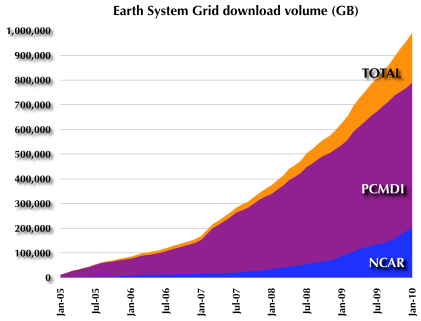

Figure 2: *In February, ESG has reached the 1 PB download milestone!*

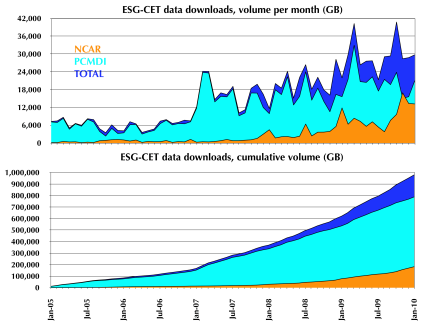

Figure 3: *The cumulative user-ship of the ESG-CET portals has continued to show growth since 2004. The total number of users by the end of 2010 is very likely to be 20,000. In a given month, between 500 and 800 users are active—up to 30 or more per day.*

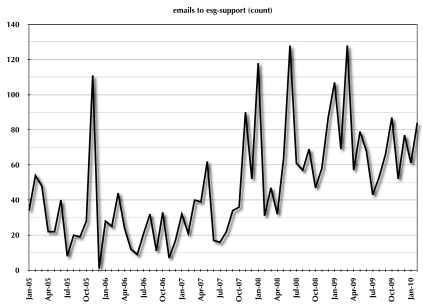

Figure 4: *ESG-CET provides support via email (esg-support@earthsystemgrid.org), and tracking the number of e-mails per month shows that the user community is well engaged with ESG-CET. An average day sees between 3 and 4 e-mails, with spikes correlated to releases of additional data sets and codes to the ESG-CET community. Many of these e-mails are related to questions regarding the served data, as well as standard support questions involving resetting usernames and passwords.*

3. ESG-CET is Positioning for CMIP5 (IPCC AR5) Data

Just as ESG was critical for the success of CMIP3 (Williams et al., 2009), it has been given the responsibility of meeting the data management needs of CMIP5, which will provide results informing the IPCC's Fifth Assessment Report (AR5) (IPCC 2007). Under the leadership of PCMDI at LLNL and with the help of NCAR, ORNL and others in the national and international community—including centers in the U.K., Germany, Japan, and Australia—an internationally federated, distributed data archival and retrieval system is being established. The CMIP5 data archive will dwarf that of its predecessor (CMIP3). Instead of the 35 terabytes collected for CMIP3, modeling groups will generate tens of petabytes, and the ~2 petabyte subset of this data expected to be of highest interest to researchers will be replicated at several data centers around the world. PCMDI will supervise the distributed CMIP5 data archive and oversee the effort to provide access to this valuable collection of output from model simulations and observations.

CMIP5 model output will mainly be made available through ESG data nodes located at the modeling centers, but PCMDI will serve as a hub for gaining access to this distributed data archive. In addition, as noted above, PCMDI will replicate in its own archive the data expected to be in particularly high demand. To further enhance accessibility to this subset of the data, it will subsequently be replicated at additional global sites (in the U.K., Germany, Japan, and Australia), where it will be served through Gateways coordinated throughout the ESG network (see **Figure 1**). In responding to this challenging undertaking, PCMDI has organized partnerships with global data centers funded to assist with CMIP5 data retrieval and dissemination and to create an internationally distributed data archival and retrieval federation, known as the Earth System Grid Federation (ESGF). The federation utilizes software primarily developed by the ESG-CET.

Ongoing partnerships and established relationships to a wide variety of data, research, and technology efforts position ESG to continue to speed progress in climate science. Although in many cases ESG tools and technologies primarily benefit the climate community, it sometimes has proved to be general enough to serve other science communities.

4. Gateway Data Discovery and Access

The Gateway software stack has been guided by two main goals: the imminent transition of the multiple ESG operational portals at PCMDI, NCAR, ORNL, and NASA to the new infrastructure, and support for the upcoming CMIP5 model output streams. The Gateway functionality has been augmented and revised in many respects, including the following major areas of development:

- **Faceted search.** A completely new user interface for the ESG-faceted search has been developed, and featured prominently on the home pages of the PCMDI and NCAR Gateways. The new interface follows paradigms that are becoming common across many e-business web sites, thus facilitating the user experience as well as increasing the scalability and maintainability of the application.
- **Data Reference Syntax (DRS).** The Gateway application has been enabled with the capability to ingest and expose metadata conforming to the DRS specification that will be used to organize and index the CMIP5 data archive. This includes parsing of the DRS information from the Thematic Realtime Environmental Distributed Data Services (THREDDS) catalogs, persistence in the relational database, and configuration of DRS facets in the metadata query interface.
- **Data versioning.** The Gateway underlying domain model has been heavily refactored to support data versioning (i.e., the need to publish, update and retract new versions of the data in response to quality control results). Versioning support had implications at all levels of the Gateway software: the relational object model, the publishing services, the metadata query services, and the user interface.
- **Data migration.** A procedure for migrating the relational database content when upgrading the underlying schema has been setup based on the open source Liquibase project. This

procedure addresses a critical operational need since it will allow Gateway administrators to upgrade to a new major software release without having to republish all the data or ask users to register again.

- **Authorization service.** The Gateway application has been augmented with an embedded Authorization Service, which can deliver Security Assertion Markup Language (SAML) signed assertion in response to authorization queries by remote clients (in particular, the Data Node plug-ins for the GridFTP and THREDDS Data Server (TDS) servers).
- **Metadata exchange.** The Gateway infrastructure for metadata exchange, based on the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) protocol, has been updated to support versioning and replication, as well as the capability to execute selected harvesting by project so that a Gateway needs only import those records that are relevant to its user community, and not others.
- **Data download.** The user workflows for selecting and downloading the data have been thoroughly revised to make them more friendly, efficient, and to support a variety of options including selection across data collections and gateways, retrieval of files from deep storage via BeStMan, generation of wget scripts and integration with the Data Mover-Lite (DML) desktop client.
- **Model metadata.** The Earth System Curator project has continued to work together with ESG to expand the Gateway functionality for ingesting and servicing model metadata, including the capability to connect data sets to the models and simulations that produced them, full handling of the CMIP5 conformance properties, ingestion of Common Information Model (CIM) metadata from the Common Metadata for Climate Modelling Digital Repositories (METAFOR) Questionnaire application, and several improvements to the user interface (the model "trackback" pages). Progress in this area was guided by a series of demonstrations that the Curator project organized to solicit continuous feedback from national and international partners, such as the Global Organization for Earth System Science Portals (GO-ESSP) and METAFOR communities.
- **Gateway user interfaces.** The Gateway User Interface has been revised and improved throughout the site, including major changes in the pages for metadata search, data download, model trackback, user and group administration.

The Gateway software has been deployed, tested, and officially release to the community according to a carefully planned schedule aimed at progressively enhancing the functionality of the application, while at the same time soliciting feedback from a progressively larger user base.

5. Data Node Server-Side Analysis

In the ESG Data Node software stack (see **Figure 5**), we have begun the development of a coordinating entity called the Data Node Manager. The architecture of this component is unique from the other components because it addresses cross cutting concerns that span all the constituent elements of the stack. The Data Node Manager logs, monitors, and collects metrics over the entire software stack. It has been designed from conception to be highly fault tolerant and perform under load with graceful degradation properties.

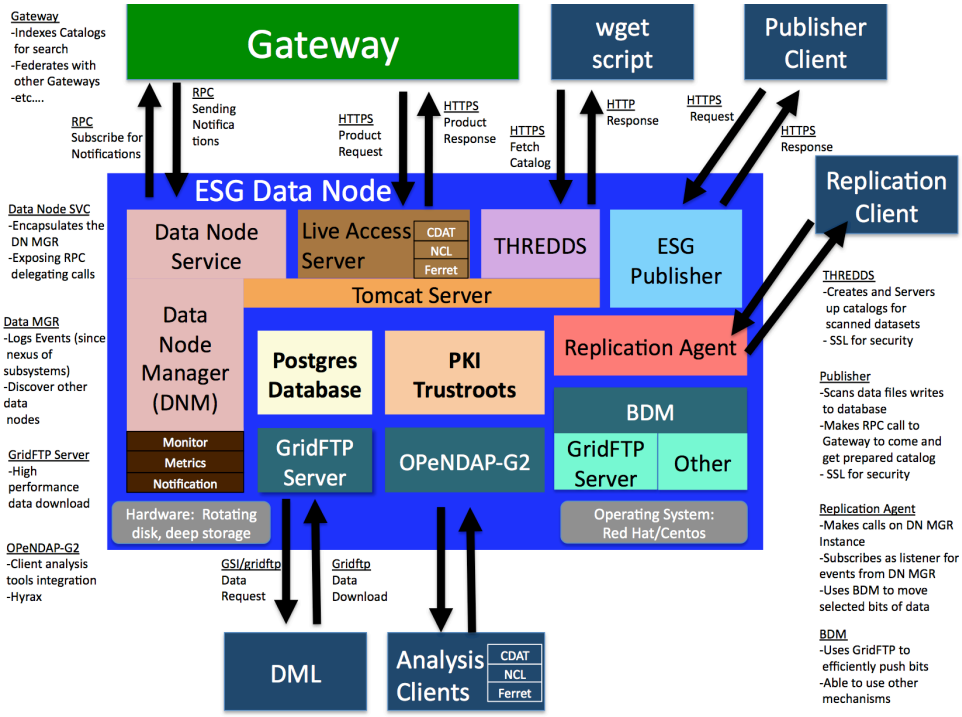

Figure 5: *This diagram shows the operations performed by the Data Node within the ESG infrastructure.*

In the ESG infrastructure, the actual data holdings reside on a potentially large number of federated nodes collectively referred to as the ESG Data Node. In addition to hosting those data, the Data Node includes the metadata services needed to publish data to portals and execute data-product requests through these portals. Personnel can set up nodes at local institutions, and the single ESG Gateway serves data requests to many associated nodes. For example, more than 30 institutions are expected to operate Data Nodes as part of the IPCC Coupled Model Intercomparison Project, phase 5 (CMIP5).

The ESG infrastructure includes the following components, which are also shown in **Figure 5**:

- **Support for high-performance data download client.** The Gateway application is augmented with the DML client, which supports high-performance, multiple-file download. Users can download and configure DML directly through the Gateway and then start it via a file list in XML format. With this system, users can request access to files stored on a local rotating disk or those in deep storage archive, which are then transferred to the Gateway cache.
- **Data replication.** Data replication is an important facet of international collaborations such as ESG. No one facility can host the accumulated holdings for the data-intensive climate research community; therefore, archives at one institution are replicated at a collaborator's site via software links. ESG has a robust data replication client, developed and improved by analyzing replication use cases and requirements to ensure that the IPCC archive stored at LLNL is available to all of the participating the European, Asian, and Australian data centers.
- **Modeling metadata.** The ESG collaboration is supporting the Earth System Curator (ESC) project by developing the full infrastructure so that it accurately captures and displays model metadata within the Gateway Web application.
- **Federated authentication.** ESG is designed as a federated system allowing user access via the ESG Gateway and supporting interoperability with other non-ESG partner data centers. The ESG infrastructure includes the OpenID protocol, which supports cross-site authentication between the many Gateways as well as with European Web portals. OpenID is

a single-sign-on technology that allows users to register at only one site, where their credentials are stored, and then carry their authenticated identity as they navigate and access data throughout the ESG federation.

- **User attribute services.** ESG, in collaboration with its European partners, chose the SAML as the enabling technology to exchange user attributes and access control information among sites. Each ESG gateway or partner data center will deploy a SAML-based attribute service, which other gateways can securely query to retrieve attribute information about a specific user. This information is required to authorize users both to access specific data sets controlled by a group at another site and to store complete data access metrics.
- **Data publishing operations.** After the Data Node software stack is installed, users can run software scripts required to publish all current ESG data holdings (i.e., climate model simulations and observations) to the Gateway. This support for full publication ensures that users have access via the Data Node software stack application to all holdings, including nonstandard model runs, multiple deep-storage archives, and multiple data access services.

The ESG Data Node software stack is distributed nationally and internationally via a script or a virtual machine (VM), a software platform that allows a complete and sovereign operating system (a guest OS) to run as an application inside another operating system (the host OS). The guest OS executes software applications identically to a physical machine. The VMs contain the fully installed CentOS (with requisite libraries and other functionality installed) and the Data Node software stack. (This setup eliminates the need for the system to check all software prerequisites). Hardware virtualization via VM improves security, insulates the hardware from attack and user error, offers code portability and ease of backup, and protects the system against potential software conflicts.

6. Data Node Security Infrastructure

ANL, NCAR, and the British Atmospheric Data Centre (BADC) have closely collaborated to define and implement a new ESG security architecture that would support secure data access by browsers and rich analysis and visualization clients throughout the federation. The new architecture is based on technologies such as OpenID, SAML, and X509 certificates and is scheduled to replace the current ESG token-based access. As part of this collaboration-wide effort, ESG has delivered the following components:

- A Java-based OpenID Relying Party (ORP), which is a web application (running within a Tomcat servlet container) that is responsible for either validating a user certificate or redirecting a user browser to an OpenID Identity Provider, with the net effect of establishing a secure authentication cookie that can be used by other Data Node software components as a proof of the user's identity.
- A servlet filters infrastructure that can be used in conjunction with the Open Runtime Platform (ORP) application to secure a generic Java-based data server, such as the THREDDS Data Server that will be part of the standard ESG Data Node application stack. Existing implemented components include an authentication filter to consume the cookie set by the ORP and an authorization filter that contacts the Gateway Authorization Service.

7. Bulk Data Mover (BDM)

The BDM is responsible for the successful replication of large data sets. Climate data sets are characterized by large numbers of small files. To handle this issue the ESG uses the BDM software as a higher-level data transfer management component to manage the file transfers with optimized transfer queue and concurrency management algorithms.

The BDM can accept a request composed of multiple files or an entire directory. The files or directory are described as Universal Resource Locators (URLs) that indicate the source sites that contain the

files. The request also contains the target site and directory where the replicated files will reside after successful transfers. If a directory is provided at the source, then the BDM will replicate the structure of the source directory at the target site. The BDM is capable of transferring multiple files concurrently as well as using parallel transmission control protocol (TCP) streams. The optimal level of concurrency or parallel streams is dependent on the bandwidth capacity of the storage systems at both ends of the transfer as well as achievable bandwidth on the wide-area-network (WAN). Setting up the level of concurrency correctly is an important issue, especially in climate data sets, because of the small files. Concurrency that is too high becomes ineffective (high overheads and increased congestion), and concurrency that is too low will not take advantage of available bandwidth. A similar phenomenon was observed when setting up the level of parallel streams in a single file transfer, such as GridFTP.

The BDM is designed to work in a "pull mode", where the BDM runs as a client at the target site. This choice is made because of practical security aspects: site managers usually prefer to be in charge of pulling data rather than having data pushed at them. However, the BDM could also be designed to operate in a "push mode" or as an independent third-party service. Because a large-scale data replication can take a long time (from many minutes to hours and even days) the BDM must be an asynchronous service. That means that when a replication request is launched, a "request token" is returned to the client. The client should be able to use that request token to check the status of the request execution at any time. Another obvious implication to the long lasting nature of large-scale replication is the need for automatic monitoring and recovery from any transient failures, which is an important part of the BDM's design.

8. Multi-phase Transfer Request Management

The tasks that the BDM performs to accomplish a successful replication are organized into three phases, as shown in ***Figure 6***. The initialization phase plans and prepares file replications from the data source to the local target storage. It includes the following tasks: 1) storage allocation verification at the target site, this requires collecting the total data size from the source site; 2) generating a request plan, the plan includes the initial level of concurrency, number of parallel streams, and buffer size for the request; 3) returning initial request estimation to the client; and 4) mirroring the directory structure of the source at the target site. It then generates an execution plan that includes pair-wise source-to-target URLs for all the files to be replicated. This is used by the execution phase.

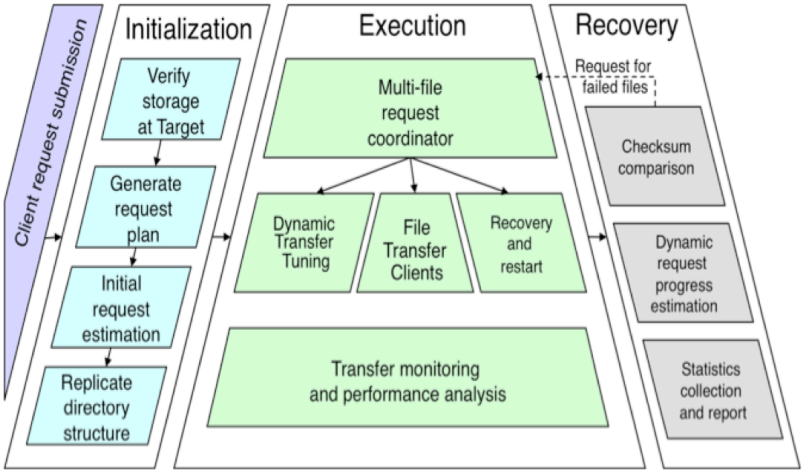

Figure 6: The design of the BDM

The Execution phase transfers the requested files, while monitoring and analyzing transfer performance for dynamic adjustment on the transfer properties. It consists of four modules: 1) the Multi-File Request Coordinator uses the information from the "execution plan" and transfer properties including the concurrency level, and accordingly instantiates the file transfer client; 2) the File Transfer Client can support any transfer protocols or services preferred by the virtual organization, supported transfer protocol could be GridFTP, HTTPS, secure copy (SCP), secure FTP (SFTP), etc.; 3) the Recovery and Restart module continuously monitors the health of the system and the files being transferred; and 4) the module responsible for monitoring and adjusting concurrency collects dynamic transfer performance, and if significant discrepancies from the estimated performance are noticed, it adjusts the number of concurrency and parallel streams.

The Recovery phase interacts dynamically with the components of the execution phase to validate the completed request by collecting statistics, generating dynamic progress estimation on-demand, and validating transferred files at the end of the request. It has three functions: 1) it collects statistics from the execution of the replication request; 2) it generates dynamic progress estimation on-demand when a client asks for request progress status, this module needs the information on file transfers that completed, are in-progress, or are pending, as well as bandwidth usage statistics and estimation; and 3) the file validation module can be running as soon as files are transferred, or at the end of the request, depending on the site preference. The reason for preferring file validation by checksum comparison after all transfers complete is that calculating checksums is computationally intensive and may disturb the running transfers. This module is also responsible for re-submitting files whose checksums indicated data corruption.

9. Product Services

The ESG is intended to serve information products to users representing a broad spectrum of sophistication—from 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 and to users who only want to quickly visualize the overall behaviors of models. Highlights of the enhancements to LAS made through ESG support are:

- **LAS integration.** Integration of LAS into the authorization and access control framework used by ESG-CET (on-going development).
- **Data intercomparison capabilities through the vizGal interface**. This interface is under active development to enhance or add capabilities for intercomparison along slices in any dimension, including comparison of time-series, Hoffmuller plots, and vertical section plots (see **Figure 7**).
- **THREDDS catalog "cleaning tools".** An important part of the ESG-CET science mission is to make meaningful intercomparisons between disparate data collections. Unfortunately, many interesting data sources to which models can be compared (i.e., satellite data collections, operational model outputs, and gridded oberservational assimilations) are not kept in well-organized archives with fully developed file-level metadata. To help organize such data collections, we have built THREDDS catalog scanning tools that can identify data sets with excellent metadata from large collections, automatically create aggregations from collections of individual time-series files and create catalogs that reflect only the best of large often jumbled data repositories. These clean catalogs can let scientists from ESG-CET and elsewhere quickly identify data sets, which can be immediately used to further their scientific goals.

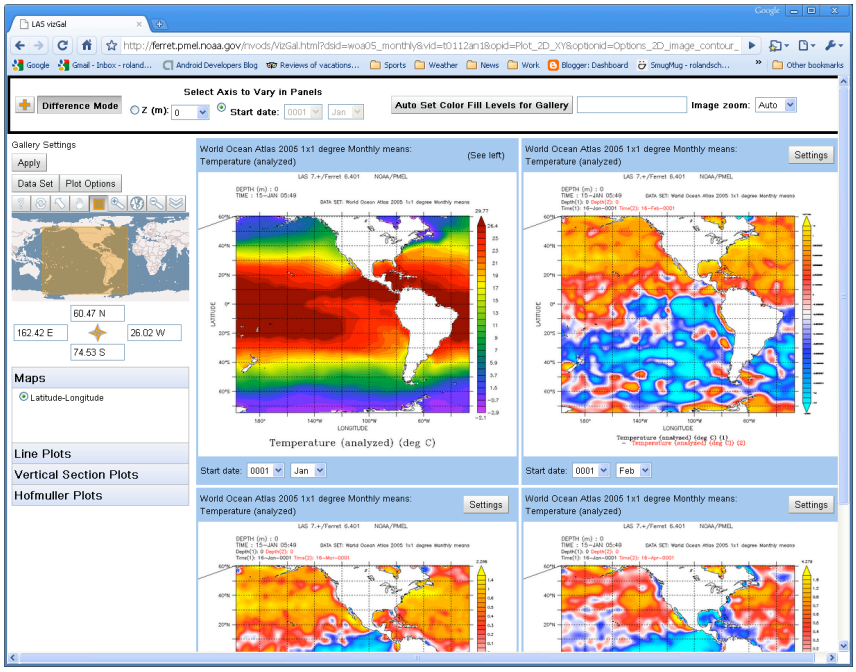

Figure 7: *Bowser displaying the new vizGal interface with multiple views showing model intercomparison.*

10. Conclusions

In summary, ESG has come to be relied upon by the climate science community to provide fundamental support for its research. ESG has been entrusted to unify computational and analytical climate resources under one unique and powerful knowledge discovery enterprise system. It continues to evolve to maintain its reputation as a reliable resource for serving the climate science community as it seeks to derive fundamental and practical understanding of climate and climate change.

11. Acknowledgements

This work is supported through the U.S. Department of Energy Office of Science, Offices of Advanced Scientific Computing Research and Biological and Environmental Research, through the SciDAC program. Argonne National Laboratory is managed by Argonne Chicago LLC under Contract DE-AC02-06CH11357. Information Sciences Institute is a research institute of the Viterbi School of Engineering at the University of Southern California. Lawrence Berkeley National Laboratory is managed by the University of California for the U.S. Department of Energy under contract No. DE-AC02-05CH11231. Lawrence Livermore National Laboratory is managed by the Lawrence Livermore National Security, LLC for the U.S. Department of Energy under contract No. DE-AC52-07NA27344. Los Alamos National Security is managed by LLC (LANS) for the U.S. Department of Energy under the contract No. DE-AC52-06NA25396. National Center for Atmospheric Research is managed by the University Corporation for Atmospheric Research under the sponsorship of the National Science Foundation. Oak Ridge National Laboratory is managed by UT-Battelle, LLC for the U.S. Dept. of Energy under contract DE-AC-05-00OR22725. Pacific Marine Environmental Laboratory is under the National Oceanic and Atmospheric Administration's line office of Ocean and Atmosphere Research, lies within the U.S. Department of Commerce.
The ESG-CET executive committee consists of Dean N. Williams, LLNL; Ian Foster, ANL; and Don Middleton, NCAR.

Lawrence Livermore National Laboratory, Livermore, CA 94550, LLNL-CONF-404832.

12. References

- ASCR Science Network Requirements, 2009: Office of Advanced Scientific Computing Research Network Requirements Workshop 2009. http://www.er.doe.gov/ascr/ProgramDocuments/Docs/ASCR-Net-Req-Workshop-2009-Final-Report.pdf.
- IPCC (Intergovernmental Panel on Climate Change), 2007: *Climate Change 2007*. www.ipcc.ch.
- Williams, D. N., 2009: The planet at their fingertips: climate modeling data heats up. SciDAC Review, 9. http://www.scidacreview.org/0902/html/esg.html.
- 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, D E Middleton, J Schwidder, R Schweitzer, R Schuler, A Shoshani, F Siebenlist, A Sim, W G Strand, M Su, N. Wilhelmi, "The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data", in the Bulletin of the American Meteorological Society, January 2009.