

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. Ahrens⁴, R. Ananthkrishnan¹, M. Balman², G. Bell³, S. Bharathi⁹, D. Brown⁶, M. Chen⁸, A. L. Chervenak⁹, L. Cinquini⁵, R. Drach³, I. T. Foster^{1,11}, P. Fox¹⁰, S. Hankin⁷, D. Harper⁶, N. Hook⁶, P. Jones⁴, D. E. Middleton^{6,11}, N. Miller¹, E. Nienhouse⁶, R. Schweitzer⁷, G. Shipman⁸, A. Shoshani², F. Siebenlist¹, A. Sim², W. G. Strand⁶, F. Wang⁸, C. Ward⁹, P. West¹⁰, H. Wilcox⁶, N. Wilhelmi⁶, S. Zednik¹⁰

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 Jet Propulsion Laboratory, 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, foster@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 system of federated science gateways to access data, information, models, analysis tools, and computational capabilities required to evaluate peta-scale to exa-scale data sets. Its goals are to (1) make data more useful to climate researchers by developing collaborative technology that enhances data usability; (2) meet the specific needs of national and international climate projects 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, ESG-CET is working to integrate in a collaborative problem-solving environment 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.

1. The Climate Data Challenge and the Earth System Grid

Climate change research places high demands on data-management, -analysis and -storage resources. These sources reside on distributed national and international data holdings. Large-scale processing is required to compare observations, which vary greatly in their temporal and spatial sampling, with simulated outputs from models. As data has become increasingly abundant and complex, transporting and effectively processing it has also become increasingly difficult. 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 provide climate researchers a doorway (e.g., via a Web portal) to data, information, models, analysis tools, and computational resources, typically providing an integrated view of data in many data nodes; data nodes are located where the data resides and are responsible for providing data access services.

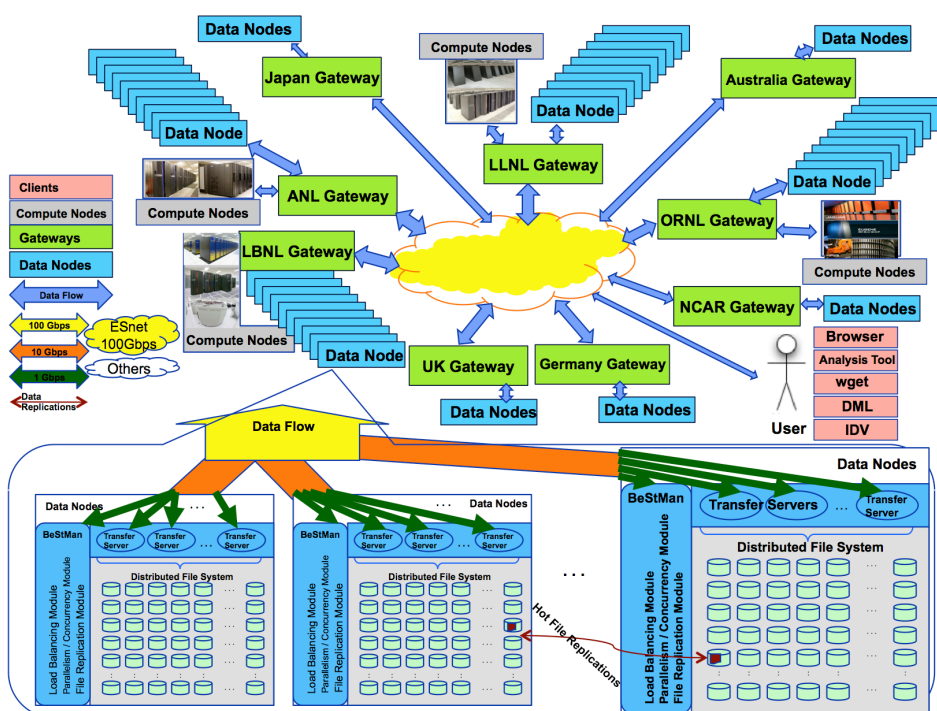


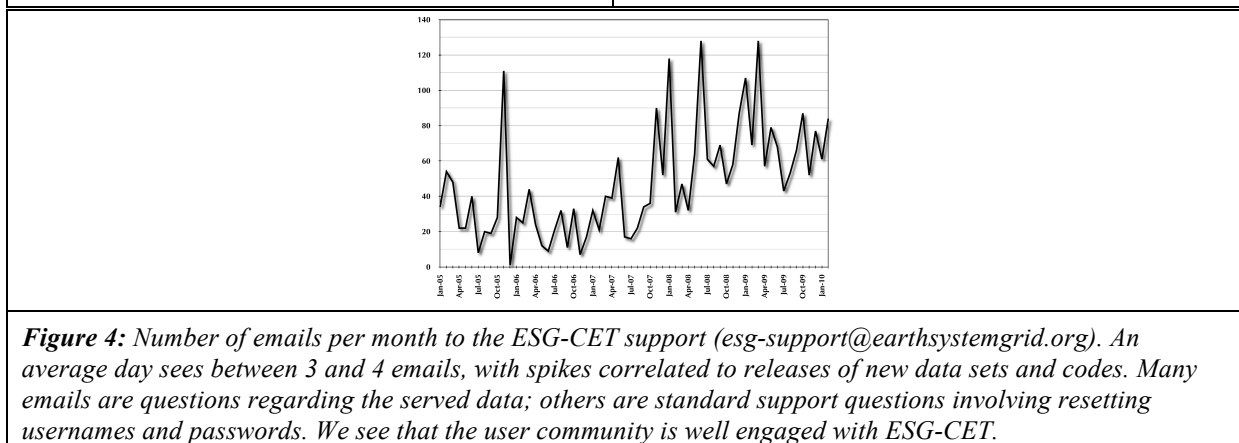
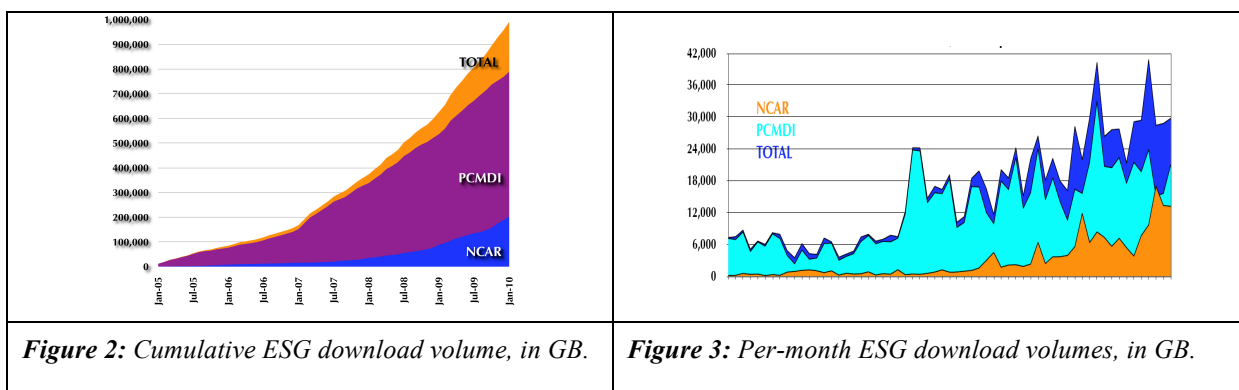
Figure 1: The topology of the ESG enterprise system with network connections and computing resources, which provides a network of geographically distributed gateways, data nodes, and computing resources in a globally federated, built-to-share scientific discovery infrastructure.

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 those from 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 the state of the climate and changes to it. In addition to these data, ESG is unifying computational and analytical climate resources under one unique powerful knowledge-discovery enterprise system. ESG 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. Volume of Data Downloaded

ESG has become an important resource for a number of climate initiatives (e.g., CMIP5, CCSM, NARCCAP, NASA observations, and more as listed above) that are central to climate investigations. In order to access and download data, ESG users must create an account at an ESG gateway. Many data (e.g., CMIP5, CCSM, POP, NARCCAP, C-LAMP, AIRS, etc.) are available for free download at one of the four ESG current 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://esg.ccs.ornl.gov/esgcet/home.htm>; and NASA/JPL, <http://esg-gateway.jpl.nasa.gov/home.htm>. Figures 2, 3, and 4 illustrate the growth in ESG usage over the last five years. Data download volumes routinely total in the tens of terabytes (TB, or 1×10^{12} bytes) per month. The cumulative total since January 2005 reached **one petabyte** (PB, 1×10^{15} bytes) in February 2010. The total number of registered users by the end of 2010 is expected to exceed 20,000. In a given month, between 500 and 800 users are active—up to 30 or more per day.



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

Just as ESG was critical to the success of CMIP3 (Williams et al., 2009), ESG has also been given the responsibility to meet 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 United Kingdom (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—"4" was skipped in CMIP numbering in order to align with AR numbering). Instead of the 35 TB collected for CMIP3, modeling groups will generate tens of PB. A subset of results (~2 PB) that is expected to be of highest interest to researchers will be replicated at PCMDI and several other data centers. 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 modeling centers, but PCMDI will serve as a hub for access to this distributed data archive. 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, ESG tools have sometimes proved to be general enough to serve other science communities.

4. Gateway Data Discovery and Access

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

- **Faceted search.** The ESG-faceted search module is a search interface that allows easy and rapid browsing of data sets in a way that never allows the user to feel lost in the data. 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 DRS is a common naming system to be used in files, directories, metadata, and URLs to identify data sets wherever they might be located within the distributed CMIP5 ESGF archive. The gateway code can now 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 underlying distributed data model has been heavily re-factored to support the tracking of multiple versions of data. Thus, data publishers can publish, update, and retract new versions 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 set up 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.
- **Attribute and Authorization service.** The gateway application has been augmented with attribute and authorization services, which can deliver Security Assertion Markup Language (SAML)-signed assertions in response to attribute and authorization queries by remote clients respectively, facilitating a fine-grained attribute based authorization policy enforcement. The data node plug-ins for the GridFTP (Allcock et al., 2005) and THREDDS Data Server (TDS) servers leverage these services to determine access to requested data.

ESG, in collaboration with its European partners, chose 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.

- **Metadata exchange.** The gateway infrastructure for metadata exchange, based on the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH), 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 supportive of 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 DataMover-Lite (DML) desktop client.
- **Model metadata.** The Earth System Curator project has continued to work 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.
- **Federated authentication.** ESG is designed as a federated system allowing user access via the ESG gateways and supporting interoperability with other non-ESG partner data centers (Siebenlist et al., 2009). The ESG infrastructure leverages 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 and create an identity 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. All web-based access to data node services leverage this authentication protocol for end user authentication. Some data node services require the use of Public Key Infrastructure (PKI). ESG gateways operate Credential Translation Services that allow a user to use a single OpenID login to obtain PKI credentials. This approach results in a seamless access experience for the end user.

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

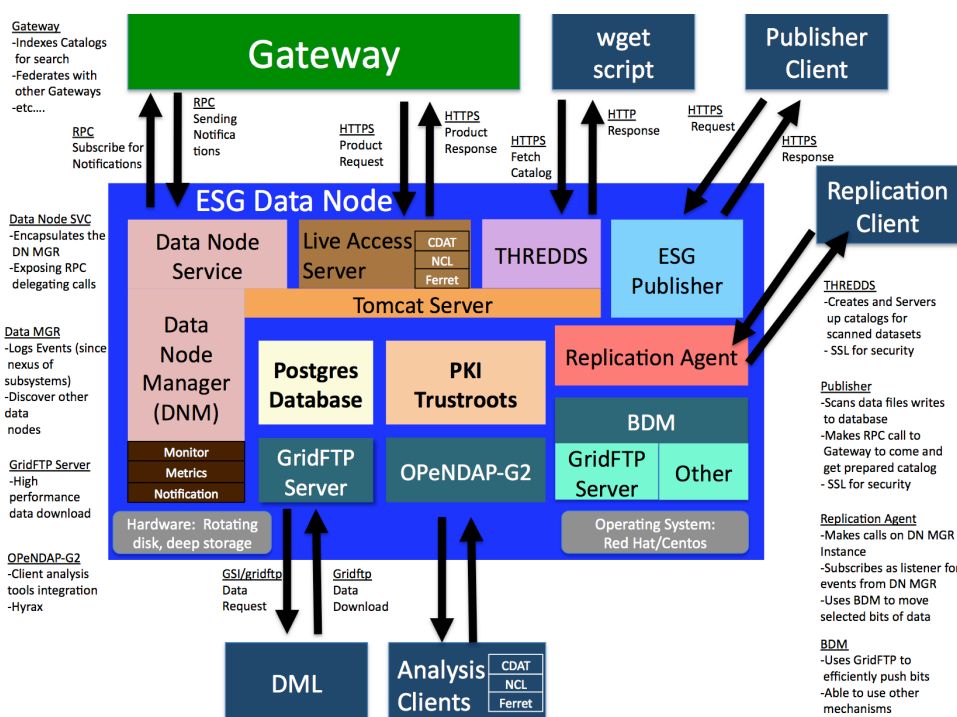


Figure 5: The operations performed by the data node within the ESG infrastructure.

In the ESG infrastructure, the actual data holdings reside on a multiple federated data nodes. In addition to hosting those data, a data node includes the metadata services needed to publish data to gateways and execute data-product requests through these portals. Personnel can set up nodes at local institutions, and a single ESG gateway can serve data requests to many associated nodes. For example, more than 30 institutions are expected to operate data nodes as part of the IPCC 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 provides 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 participating European, Asian, and Australian data centers.
- **Modeling metadata.** The ESG collaboration is working with the Earth System Curator (ESC) project at developing the full infrastructure so that it accurately captures and displays model metadata within the gateway Web application.
- **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 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 collaborated closely to define and implement a new ESG security architecture to 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 will 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 presented via SSL authentication 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 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.
- Data download clients that can be used with the new security infrastructure, including wget scripts that use client certificates to securely download the requested file by using SSL for authentication. A Java Webstart application client for the Credential Translation Service, which is a MyProxy Online CA, will also be integrated to allow users to download credentials to their local machine, to use with DML clients.

7. Bulk Data Mover

The Bulk Data Mover (BDM) is responsible for the successful transport of large data sets. Climate data sets are characterized by large volume of files with extreme variance in file sizes. The ESG uses BDM as a higher-level data transfer management component to transfer files in data sets into ESG data nodes. BDM achieves high performance using a variety of techniques, including advance transfer queue management, multi-threaded concurrent transfer connections, data channel caching, load balancing over multiple transfer servers, adaptive transfer adjustment, and dynamic parameter tuning.

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 BDM will copy the structure of the source directory at the target site. 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.

BDM is designed to work in a “pull mode”, where it 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, BDM could also be designed to operate in a “push mode” or as an independent third-party service. Because large-scale data replication can take a long time (from many minutes to hours and even days), 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 can use that request token to check the status of the request execution at any time. Another obvious implication of the long lasting nature of large-scale replication is the need for automatic monitoring and recovery from any transient failures.

8. Multi-Phase Transfer Request Management

The tasks that BDM performs to accomplish 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, the number of parallel streams, and the 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.

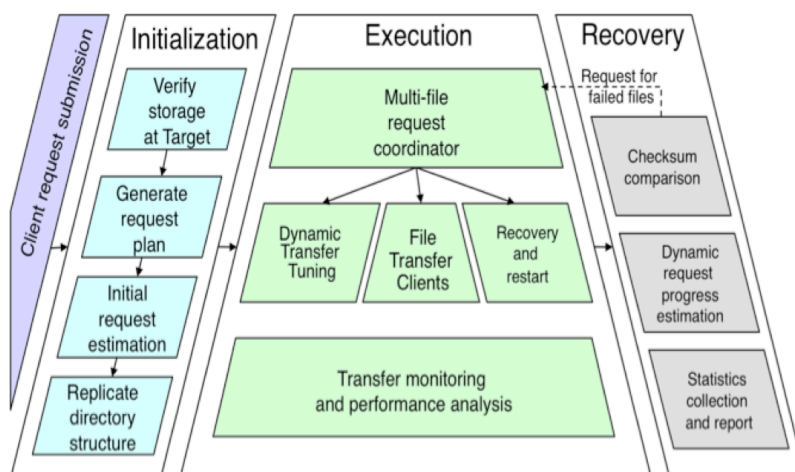


Figure 6: Design of the Bulk Data Mover

The execution phase transfers the requested files, while monitoring and analyzing transfer performance for dynamic adjustment of 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 performs the transfers – it can support any transfer protocols or services preferred by the virtual organization, including GridFTP, HTTPS, secure copy (SCP), and secure FTP (SFTP); 3) the recovery and restart module continuously monitors the health of the system and the files being transferred, and attempt to re-transfer files that failed to transfer properly; and 4) the module responsible for monitoring and adjusting concurrency collects dynamic transfer performance, and if significant discrepancies from the estimated performance are noticed, 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 estimations on-demand when a client asks for request progress status, which means that this module needs information on file transfers that completed, are in-progress, or are pending, as well as bandwidth usage statistics and estimation; and 3) it validates files, so this 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 are 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

ESG is intended to serve information products to users of widely varying sophistication—from numerical modelers, who want access to “raw” model output files and verbatim subsets of model output; to climate impacts investigators, who often want rapid access to these small subsets of data without the complexities of model-specific coordinate systems; and to users who only want to quickly visualize the overall behaviours of models. Information products such as scientific visualizations, tables and results from simple statistical calculations are provided by the Live Access Server (LAS) developed at PMEL. Highlights of the enhancements to LAS made through ESG support are:

- **LAS security integration.** Ongoing work focussed on integrating security solutions with LAS will result in the LAS deployments leveraging ESG-CET security infrastructure for authentication and authorization of user access.
- **Data intercomparison capabilities through the vizGal interface.** This interface provides the ability to assess differences both qualitatively (via visual inspection) and quantitatively (via regridding and differencing) between fields. Intercomparison tools may be utilized along slices in any dimension, including comparison of maps, time-series, Hoffmuller plots, and vertical sections (see **Figure 7**).
- **THREDDS catalog “cleaning tools.”** An important part of the ESG-CET science mission is to enable 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 observational 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 that can be used to further their scientific goals.

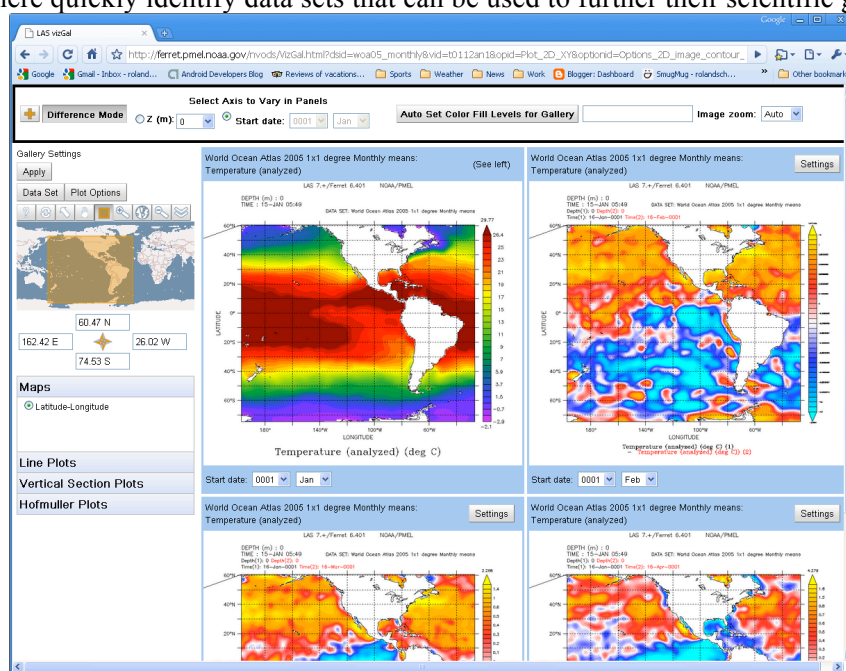


Figure 7: Screen image of the vizGal interface showing differences between temperature fields.

10. Conclusions

The climate research community as a vital infrastructure component relies heavily upon ESG. By unifying computational and analytical climate resources under one unique and powerful knowledge discovery enterprise system, it has enabled research that would otherwise have been performed only with great difficulty or perhaps not at all. ESG 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.

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>.
- Williams, D. N., et al, 2009: "The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data," *Bulletin of the American Meteorological Society*, January 2009.
- Allcock, B., Bresnahan, J., Kettimuthu, R., Link, M., Dumitrescu, C., Raicu, I. and Foster, I., 2005, "The Globus Striped GridFTP Framework and Server." SC'2005.MyProxy Credential Management Service, <http://grid.ncsa.uiuc.edu/myproxy/>
- OpenID, <http://openid.net/>
- Security Assertion Markup Language (SAML), OASIS Security Services (SAML) TC, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security
- Bulk Data Mover, <http://sdm.lbl.gov/bdm/>
- DataMover-Lite, <http://sdm.lbl.gov/dml/>
- Siebenlist, F., Ananthkrishnan, R., Bernholdt, D.E., Cinquini, L., Foster, I., Middleton, D., Miller, N., and Williams, D.N., 2009. "Enhancing the earth system grid security infrastructure through single sign-on and autoprovisioning," Workshop on Grid Computing Environments.