3rd Annual
Earth System Grid Federation and Ultrascale Visualization Climate Data Analysis Tools
Face-to-Face Meeting Report

December 2013

A global consortium of government agencies, educational institutions, and companies dedicated to delivering robust distributed data, computing libraries, applications, and computational platforms for the novel examination of extreme-scale scientific data.

LLNL-TR-650500

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

LLNL-TR-650500

Contents

1 Abstract

The climate and weather data science community gathered December 3–5, 2013, at Lawrence Livermore National Laboratory, in Livermore, California, for the third annual Earth System Grid Federation (ESGF) and Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) Face-to-Face (F2F) Meeting, which was hosted by the Department of Energy, National Aeronautics and Space Administration, National Oceanic and Atmospheric Administration, the European Infrastructure for the European Network of Earth System Modelling, and the Australian Department of Education. Both ESGF and UV-CDAT are global collaborations designed to develop a new generation of open-source software infrastructure that provides distributed access and analysis to observed and simulated data from the climate and weather communities. The tools and infrastructure developed under these international multi-agency collaborations are critical to understanding extreme weather conditions and long-term climate change, while the F2F meetings help to build a stronger climate and weather data science community and stronger federated software infrastructure. The 2013 F2F meeting determined requirements for existing and impending national and international community projects; enhancements needed for data distribution, analysis, and visualization infrastructure; and standards and resources needed for better collaborations.

2 Executive Summary

The Department of Energy (DOE) Office of Biological and Environmental Research (BER) is the primary sponsor for the development and operation of the Earth System Grid Federation (ESGF) software system and coordinator for implementation of the ESGF software stack at numerous sites worldwide. As noted on the ESGF website (http://esgf.org), other ESGF development sponsors include the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), Infrastructure for the European Network of Earth System Modelling 2 (IS-ENES2), and Australia's Department of Education. BER is also the primary sponsor for the development of the Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT). As with ESGF, UV-CDAT's website (http://uv-cdat.org) also includes a list of additional development sponsors, such as NASA, NOAA, and the Australian Department of Education. Both ESGF and UV-CDAT are award-winning software products used both for current climate model and observation assessment studies and as part of a larger integrated workflow and test bed for accelerating climate energy research.

ESGF was originally conceived to provide data federation, search, discovery, and secure access for climate model output from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project (CMIP). ESGF has since become the de facto standard for delivering a wide variety of simulation and observational data to the climate community: CORDEX, TAMIP, LUCID, ACME, GeoMIP, MERRA, and CFSR, to name only a few. As the leading source for current climate data holdings, ESGF's international multi-agency mission is to lead climate data discovery and knowledge integration. Currently, ESGF software is used by over 40 international and national projects, and more projects are expecting to come online in the near future.

The third annual ESGF and UV-CDAT Face-to-Face Meeting, held December 3–5, 2013, provided the climate community an opportunity to discuss and reach agreement upon necessary short-term and long-term goals for developing software components. To maintain operations and integration leadership, ESGF tasks for the upcoming year will include the application and integration of community-driven activities, such as DOE's Accelerated Climate Modeling for Energy (ACME) end-to-end model test bed; DOE's UV-CDAT remote processing services and visualization; DOE's Globus Online high-speed, reliable data transfers among node sites and from node sites to user; DOE's ProvEn workflow and provenance integration; NOAA's and NSF's new CoG-based ESGF user interface for project distinction, larger integration, greater flexibility, and ease of use; IS-ENES2 security model improvements; and NASA's improved data publication services, which include archiving images and other forms of data products. Other funded development activities from the international multi-agency partnership (such as Australia and Asia) will be included to further the mission of this software development collaboration.

As part of the funded mandate by BER, the Analytics and Informatics Management Systems at Lawrence Livermore National Laboratory (LLNL) will support the ESGF software stack and system for CMIP3 and CMIP5. For better maintenance, compatibility, and long-term use, we will migrate CMIP3 from its legacy ESG-II software stack repository to the latest ESGF distributed archive that also hosts CMIP5 and over three dozen other climate data products. To distinguish between CMIP3, CMIP5, and the other climate data projects, a new user CoG-based front-end will be upgraded to become the reference ESGF web user interface. CoG is a NOAA/NSF product that combines a content management system, scientific online collaboration environment, interfaces to data and metadata services, and facilities for formal project governance. The CoG web front-end is specifically designed for multi-project distributed organizations and is, for the most part, integrated with ESGF back-end data services.

The ability to process and visualize data in situ (i.e., where the data is stored) becomes increasingly important as the size of the data grows. To enable this functionality, the ESGF and UV-CDAT teams will work to integrate the wide variety of industry standard scientific computing libraries (e.g., GO, THREDDS, OPeNDAP, VisTrails, R, VTK, ESMF, ParaView, etc.) and applications for ESGF remote processing and visualization services. Ongoing CMIP operations will take advantage of work being done by the UV-CDAT team to bring in diagnostics, model metrics, and uncertainty quantification ensembles. We will also modify both ESGF's back-end and the new CoG front-end to accommodate the integration of remote services.

While successful, the Wget file transfer is not sustainable for long-term use due to its modest performance and user overheads. Therefore, we are moving away from Wget to Globus Online (GO), which provides supported, reliable, and managed file-transfer capabilities over GridFTP. GO will allow users to store data to the ESGF's data cart and will be integrated with remote services to offer the transfer of both users' defined subsets and comprehensive data set requests. For scientific reproducibility and collaboration, remote services, including analysis and visualization, will be connected through a provenance Application Programming Interface (API) to automatically capture meaningful history and workflow. For instance, this captured information may be shared with others or used to reproduce a set of events at a later date.

To broaden data publication beyond the standard netCDF/Climate and Forecast (CF) file type (e.g., other NetCDF files or images), we will extend the ESGF Publisher to include custom, extensible handlers that can be plugged in to parse data in other formats. In addition, the data publication process, the entire new CoG interface, and ESGF user environment will be documented as web tutorials and in written text.

This report lists ESGF, UV-CDAT, and accompanying development software tasks for January 1, 2014 through December 31, 2014. **Figure 1** shows the new ESGF architecture with newly integrated components such as CoG and UV-CDAT.

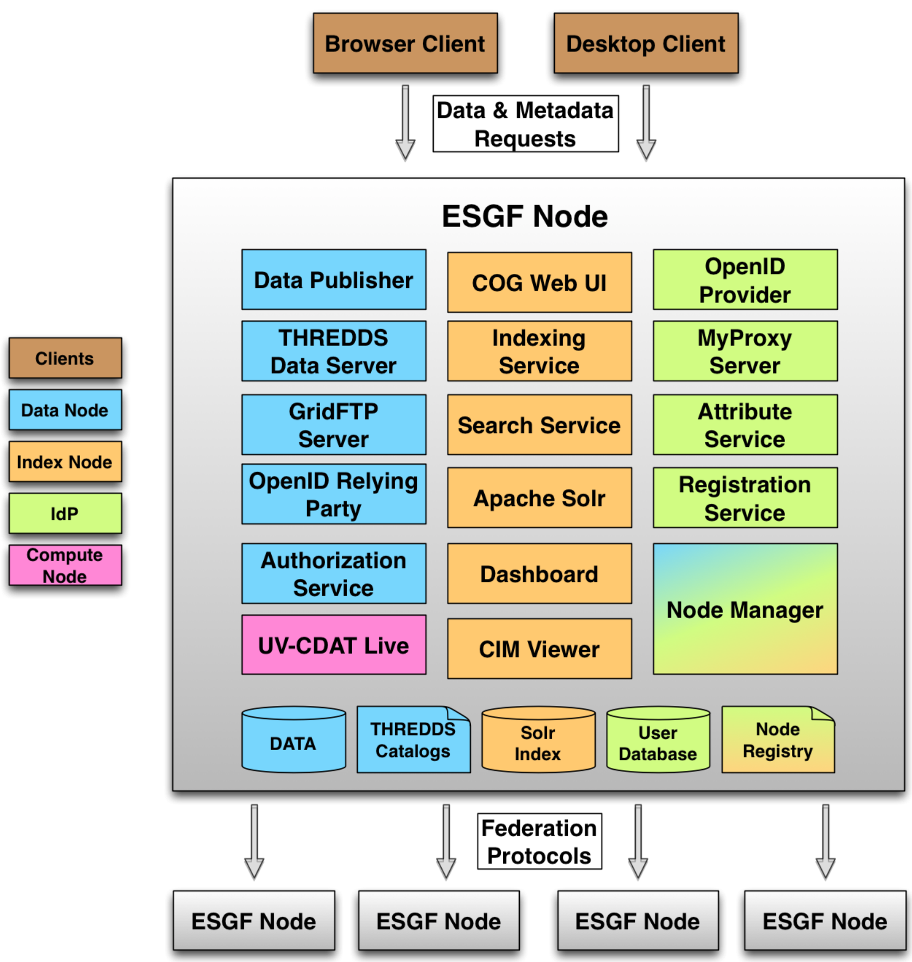


Figure 1. *Next-generation ESGF architecture.*

3 Organization of ESGF and UV-CDAT

3.1 Organizations

With LLNL as a leading partner, international institutions such as the British Atmospheric Data Center (BADC), German Climate Computing Center (DKRZ), and the Australia National University (ANU) are undertaking the software development and project management of ESGF. For UV-CDAT, most of the software developer and project management team reside in the United States (at DOE/LLNL, New York University [NYU], NASA, NOAA, and many other institutions). The sponsors of ESGF and UV-CDAT are represented at the top of **Figures 2 and 3** and include DOE (the primary funder), NASA, NOAA, the European Commission, the Australian government, and others. For ESGF and UV-CDAT, there are dozens of data centers around the globe that use these products for data projects; these projects are represented in the figures as blue and green circles. In the figures, CMIP5 is shown as a green circle to show that it the primary project servicing data and analysis for the community. ESGF provided CMIP5 data to the climate experts that generated the 2013 Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) report, a document on the basis of which political decisions are made. Given the past growth of ESGF and UV-CDAT, it is reasonable to assume that the number of future data projects (depicted as blue circles) will increase as time goes on. Likewise, these projects will continue to deliver more climate and weather data for important reports, which will help our understanding of future climate change.

Users with specific queries about these data holdings can consult climate specialists. Entertaining scientific queries along with technical issues is currently a large part of the user support services provided by ESGF and UV-CDAT. Depicted as an elongated disk in **Figures 2 and 3,** the data holdings, analyses, and visualizations are made available to users worldwide. For more information on ESGF, visit the reference [Cinquini, 2013] and for more information on UV-CDAT, visit the reference [Williams, 2013].

The executive components of ESGF and UV-CDAT are represented in the figures as rectangles. The ESGF and UV-CDAT development and maintenance teams, also known as administrative bodies, are depicted as diamonds. Development and maintenance teams, led by team heads, currently exist in Asia, Australia, Europe, and North America, with South American and African teams joining soon. Each of these teams manages and maintains the nodes within their region. These nodes collaborate with other nodes to form a system that is not only technically federated but also socially, institutionally, and administratively. UV-CDAT will be an integral part of the compute component of the ESGF software stack, as shown in **Figure 1**.

It is worth noting that the labels attached to each team or administrative body in the **Figure 2** with letters D, G, S and C indicate that a node may have four different roles or flavors within itself [Cinquini, 2013]. The letter "D" stands for data node, which means that data holdings are hosted by an administrative body, while the associated number shows how many data nodes are being hosted and maintained by a single administrative body. For example, a label "3D" would indicate that there are three data nodes, which are managed by a particular administrative body. Every node has at minimum a data node and additionally may have a gateway (G), security set-up (S), and/or compute facility (C) node. A gateway node, also known as an index node, is responsible for representing the data sets available in ESGF system to a user via user interface (UI).

A security node, also known as an identity provider (IdP), is responsible for user identification and authorization. The security node ensures that each user is a valid entity who is entitled to access data sets available in ESGF system via its UI gateway from any ESGF node via Single Sign-On. A compute node is responsible for computing and visualization in ESGF. It is also where the UV-CDAT software stack is integrated for analysis, diagnostics, model metrics, and visualization.

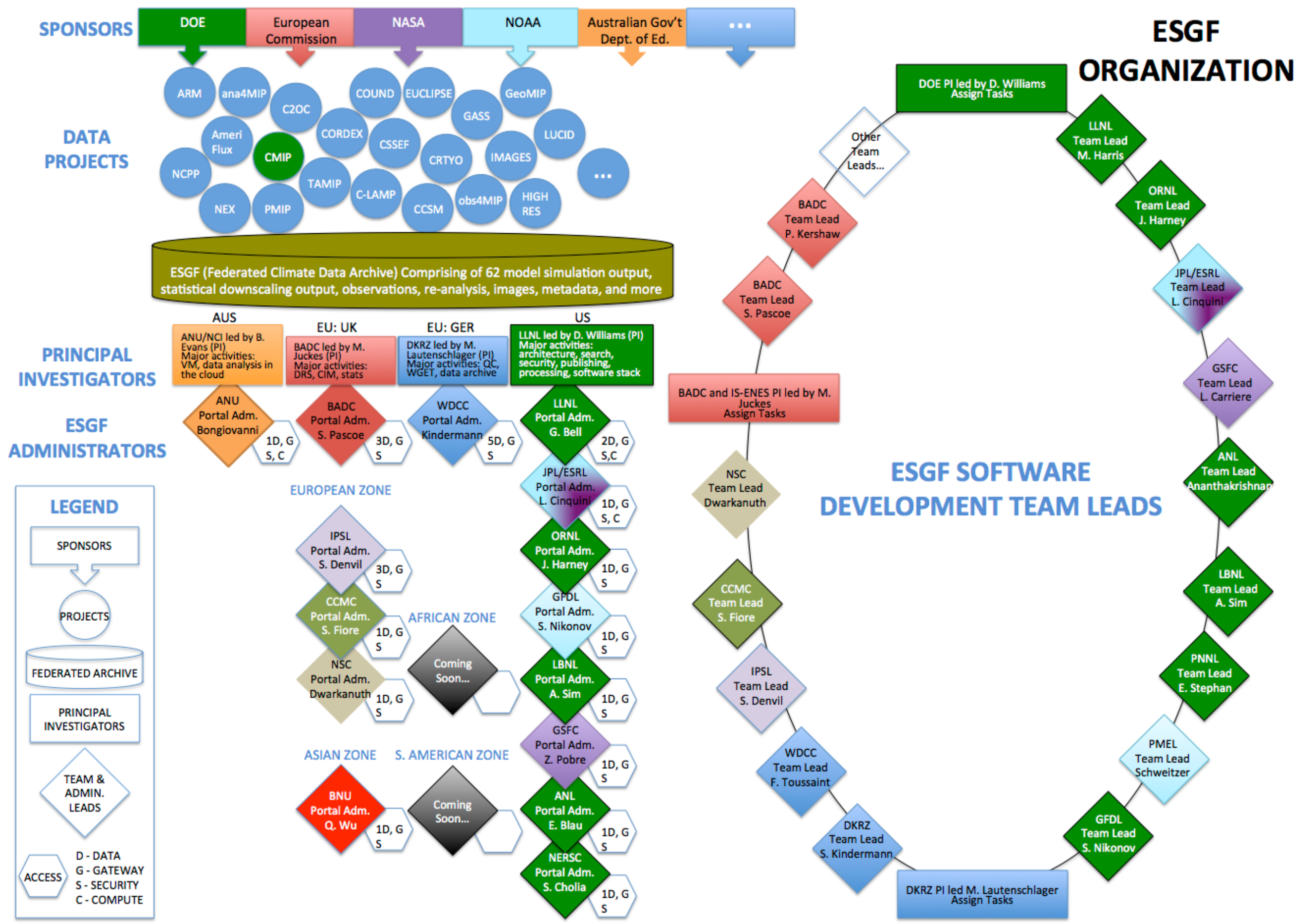


Figure 2. *ESGF's international multi-agency organization of physical and human resources. Note: different colors show different entities and organizations.*

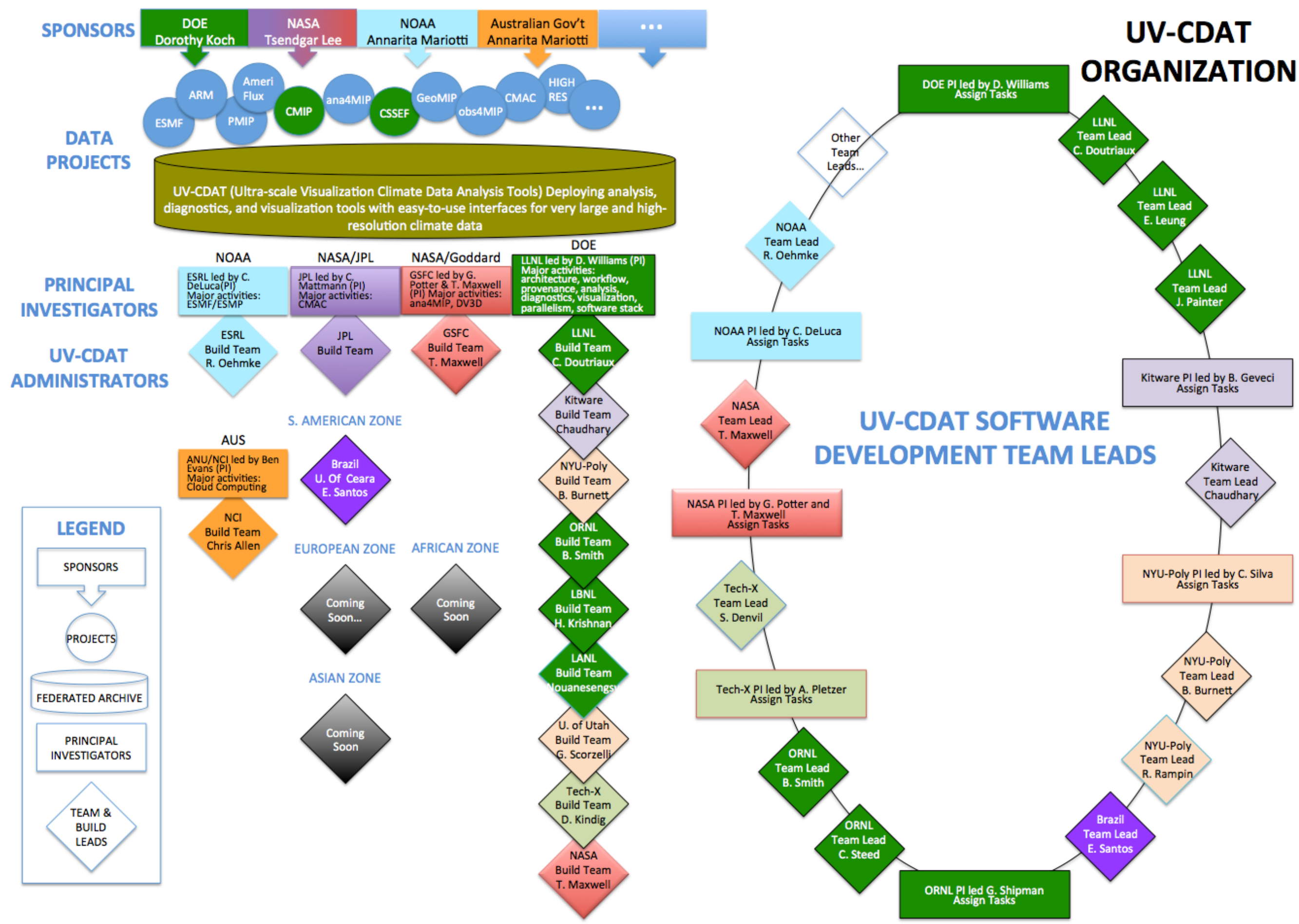


Figure 3. *UV-CDAT's U.S. national organization of physical and human resources. Note: different colors show different U.S. entities and organizations.*

ESGF subsystems or system components can be anywhere in the world. In addition, the network topology of the ESGF node network directly influences its management and administration structure. **Figure 4** shows the international network governance of ESGF, known as the International Climate Network Working Group (ICNWG).

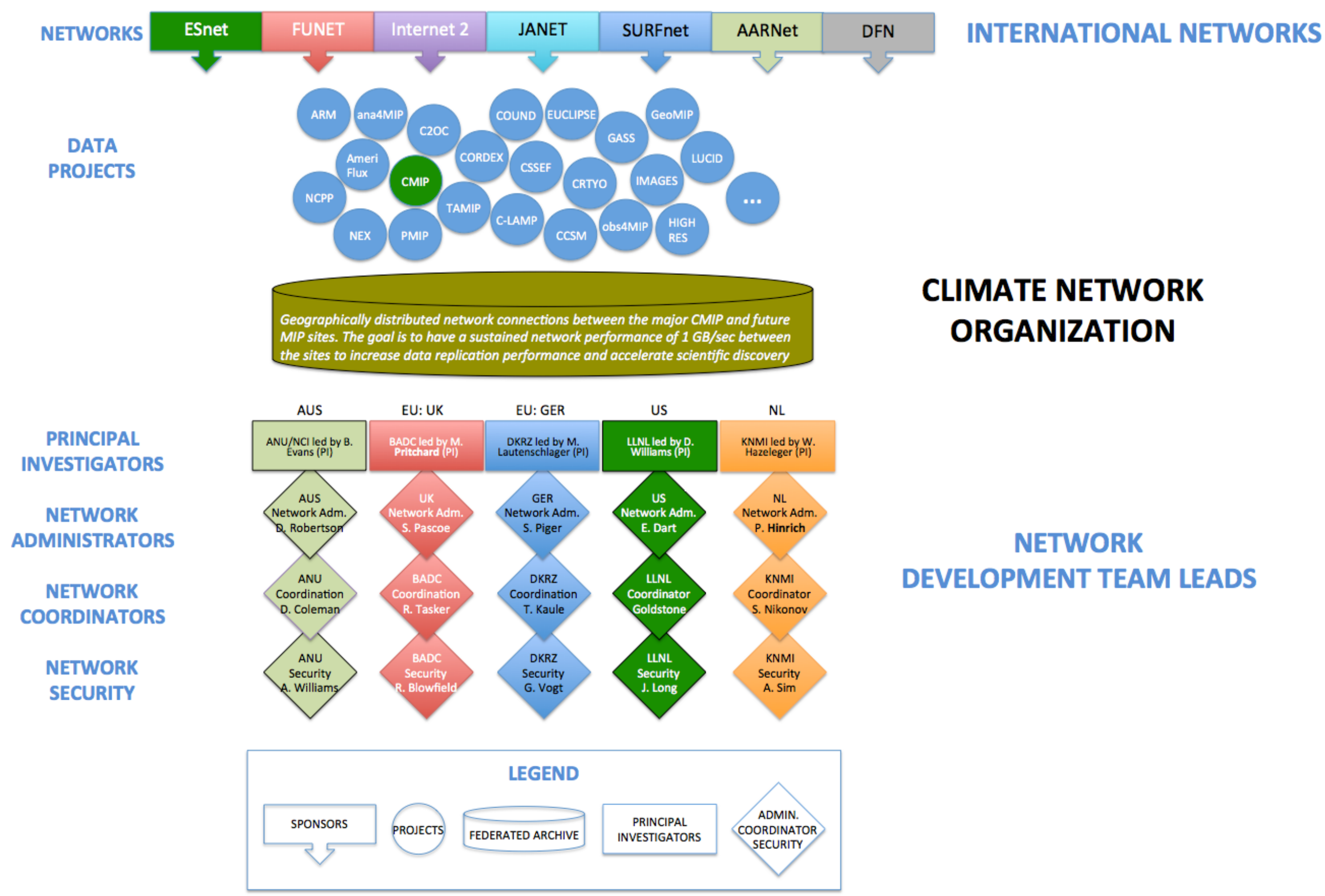


Figure 4. *ESGF started a new working group called the International Climate Network Working Group. This working group is dedicated to helping all ESGF climate data sites set up, optimize, and/or troubleshoot their network infrastructures for international climate data transfers, so that petabytes of data can traverse international networks from end-to-end at the high performance levels required for large-scale data analysis. By 2016, the group aims to achieve 2 GB/sec for data transfer throughput between at least five ESGF sites.*

The number of administrative bodies, data holdings, users, and staffs participating in these systems are likely to increase. Additionally, the role of an administrative body or a node itself may change, and thus the whole arrangement is complex and dynamic. These systems demand continuous architecture redesign activities, software development, hardware changes, data publishing, data curation, data quality check activities, analysis, diagnostics, model metrics, and visualization. A dynamic and ever-evolving infrastructure also needs a dynamic user support "service desk." Both computationally intensive science carried out in highly distributed network environments, and as in our case, science that uses immense data sets requiring grid computing are likely to be confronted with demanding user support issues that are quite similar to data curation and software development requirements.

Table 1 lists ESGF and UV-CDAT of the recommendations for the coming year by members attending the meeting.

Table 1. *Recommendations for ESGF and UV-CDAT by meeting members.*

Recommendation	Task
Comprehensive improvements for ESGF	Identify critical shortcomings of the current ESGF infrastructure and provide major improvement in the overall usability, and reliability of the system to better support CMIP6 and beyond.
Short-term ESGF efforts	In the next few years, ESGF will focus most of its resources towards improving the current infrastructure, rather than researching revolutionary new technologies or architectures.
Comprehensive improvements for UV-CDAT	Identify the critical shortcomings of the current UV-CDAT infrastructure and provide usability and reliability improvements. These will include porting to UV-CDAT to other platforms, faster and lighter builds, and core architecture redesign for greater integration into the overall ESGF, UV-CDAT, and test bed system.
Meeting report should formalize work plan for the year	One of the major outcomes of the face-to-face meeting is to formalize a work plan (including responsibilities and timelines) for all major software components addressing critical problems.

4 Science Drivers and Challenges

4.1 Projects

Climate science is a discipline in which scientific progress is critically dependent on the availability of a reliable infrastructure for managing and accessing large and heterogeneous data sets. It is an inherently collaborative and multi-disciplinary effort that requires sophisticated modeling of the physical processes and exchange mechanisms between multiple Earth realms (atmosphere, land, ocean, and sea ice) and comparison and validation of these simulations with observational data from various sources, sometimes collected over long periods of time.

The climate community has worked for the past decades on concerted worldwide modeling activities led by the Working Group on Coupled Modeling, sponsored by the WCRP, and leading to successive reports by the IPCC. The most recent, AR5, was released in September 2013. For these activities, dozens of modeling groups in as many countries run the same prescribed set of climate change scenarios on the world's most advanced supercomputers and produce several petabytes of output containing hundreds of physical variables spanning tens and hundreds of years.

These data sets are held at distributed locations around the globe but must be discovered, downloaded, and analyzed as if they were stored in a single archive, with efficient and reliable access mechanisms that can span political and institutional boundaries. The same infrastructure must also allow scientists to access and compare observational data sets from multiple sources, including, for example, Earth Observing System satellites and Atmospheric Radiation Measurement sites. These observations, often collected and made available in real-time or near real-time, are typically stored in different formats and must be converted, post-process and on demand, to a format that allows easy comparison with models. Finally, science results must be applied at multiple scales (global, regional, and local) and made available to different communities (scientists, policy makers, instructors, farmers, and industry). Because of its high visibility and direct impact on political decisions that govern human activities, the end-to-end scientific investigation must be completely transparent, collaborative, and reproducible. Scientists must be given the environment and tools for exchanging ideas and verifying results with colleagues in different time zones, investigating metadata, tracking provenance, annotating results, and collaboratively developing analysis applications and algorithms.

DOE climate modeling and data centers, such as Argonne's and Oak Ridge's Leadership Computing Facilities and the National Energy Research Scientific Computing Center, provide the community with high-performance computing (HPC) systems, clusters, short- and long-term storage, networking, and coordinated software infrastructure resources. In addition to running and processing state-of-the-science climate models at these DOE computing facilities, the community must rely on several types and levels of services to effectively produce and manage distributed climate data from many sources:

- Domain-Specific Distributed Data Services: Captures the set of unique requirements and needed services for each unique climate project.
- Common Data Services: Shared across all climate projects, such as movement, curation, discovery, exploration, etc.
- Data Systems Software Layers: Includes lower layers of software services such as metadata, directory structures, provenance, and workflow.
- Data System Hardware: Includes HPCs, clusters, clouds, and large storage, for modeling, for in situ data analysis, and for post-hoc large-scale computational data analysis. This also includes in-transit processing to enable extreme-scale climate analysis.
- Networks: Binds the collection of disparate hardware, networks, and software resources for community use. Networks are necessary also to replicate and move large data holdings at storage facilities and to federate connectivity. ESnet's 100 Gbps network is of particular interest.

If the climate community is to optimize its investments in data, it must ensure that a common open architecture is in place and a significant fraction of that architecture is shared among the different climate activities, rather than having specific domain architecture for each project.

Similar infrastructures are supported at other ESGF nodes. For example, the National Computational Infrastructure (NCI) facility at the Australian National University provides high-performance cloud computing and storage and community software stack support, which provides researchers with an increasingly powerful and ultimately standardized environment for researchers across ESGF.

Table 2 lists some of the recommendations for the coming year by members attending the meeting.

Table 2. *Specific recommendations from the face-to-face meeting.*

Recommendation	Task
Comprehensive long-term support	Continue to develop and support a comprehensive, long-term, sustainable solution for empowering distributed data services, data systems software layers, next-generation HPC and storage, and next-generation networks accessing national and international large-scale data sets.
Community standards	Continue to develop community-established standards and protocols that are needed for distributed data and service interoperability of independently developed data systems and services.
Reference model	Continue to develop a reference model and support application programming interface standards, which are essential for enabling collaborations and facilitating extensibility so that similar, customized services can be developed across science domains. This framework must support the ability for resources contained at every level to transfer information within and across multiple software layers.
User feedbacks	Expand opportunities for users to provide feedback on ESGF and UV-CDAT user interfaces, capabilities, and support services, and enable this feedback to guide development priorities and improvements to operations.

4.2 *Operations*

ESGF is an open consortium of institutions, laboratories, and centers around the world that are dedicated to supporting research on climate change and its environmental and societal impacts. As stated, ESGF was a key contributor to the success of CMIP5 and the IPCC's Fifth Assessment Report. To continue support for the IPCC community, we have operation activities divided up into following categories:

- Data transport—the management of data from modeling centers to LLNL and from LLNL to data centers;
- Data publication and replication of data between LLNL and designated data centers;
- Network maintenance to optimize data delivery and to correct subpar performance of the network;
- Hardware maintenance—maintaining and upgrading a two petabyte storage system, front-end servers, and computer cluster for derived variables, using UV-CDAT; and
- User interface maintenance and compatibility development for analysis, diagnostics, and model metrics capabilities—again using UV-CDAT.

Because a number of new CMIP-related projects are coming online, ESGF must continue to help with installation and use of the ESGF software stack and system. In 2014, this effort will include improvements and system installation on hardware platforms and virtual machines, as well as the development of administration instructions and the documentation associated with becoming an ESGF node administrator. Operational activities recommended for the year are shown in **Table 3**.

Table 3. *Recommended operational workforce efforts for the coming year.*

Operational Task	Work Description
Replicating data	Transfer and publish CMIP5 data sets for centers without data nodes (and replicate high priority data).
Enhancing publisher	Enable publisher to publish to multiple projects (e.g., CMIP and PMIP). This includes simplifying the publisher.
User statistics	Provide download statistics: How many times has each file in CMIP5 been downloaded? This includes the new Node Manager to make sure the numbers are consistent across the federation.
Enabling OPeNDAP and GridFTP across the federation	Assist all CMIP ESGF data nodes in implementing OPeNDAP and GridFTP and republishing data.
Providing user help	Provide ESGF help desk support and maintain ESGF documentation and website for node installation, publishing, node updates, certificates, authentication, security issues, etc. Maintain Askbot.
New UI	When ready, transition to the new CoG web interface.
Convention support	Continue to be involved in the CF convention website and documentation support.
Server-side analysis	Develop a LLNL-hosted web utility to produce graphs based on output from the model metrics package provided by modeling centers.

4.3 Modeling and Data Centers

With input from representatives around the world, the modeling and data centers provided clear goals and direction for improving climate and weather data discovery, access, efficiency, and sustainability across all aspects of their sites. As shown in **Figure 5**, each center touched upon the major components necessary for large-scale data accessibility, including websites, portal services, frameworks and tools, core application programming interfaces, collaborative services, facilities, and networks. Representatives outlined current best practices and a roadmap for future climate and weather virtual laboratory centers. For example, the U.S. National Center for Atmospheric Research (NCAR) provided a detailed description of the requirements necessary for operators to benchmark their current performance using an aggregation of use metrics from distributed systems, thereby determining the need for data access control and workflow adjustments. Other requirements such as remote publication of geographically distributed products and integration with legacy and tape storage systems will help to achieve greater federation efficiency and sustainable data delivery, both today and into the future.

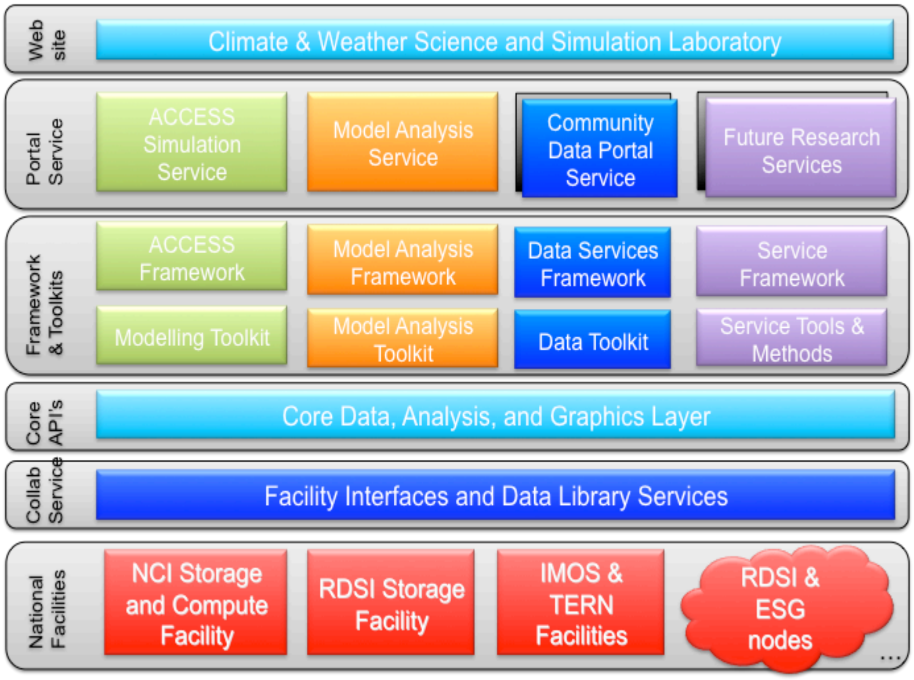


Figure 5. *Modeling accessibility of a climate and weather virtual laboratory provided by the Australian National Computational Infrastructure.*

The ESGF software stack must operate within the financial, time, and resources constraints of centers and their prospective projects. These content areas align with the way many organizations are structured, which means that centers can easily separate out the ESGF model to meet their needs and resources by only enabling one or more of the ESGF software components (i.e., data node, index node, IdP node, and/or compute node). Since the climate modeling and data centers rely on ESGF to distribute project data, they may want to assemble an international task force to lead the effort of crafting a standard ESGF use model for future users. This task force would solicit input from a variety of institutions throughout the federation, capturing project data needs and requirements by the community at large. As indicated in **Table 4**, the task force would seek out guidance, ideas, and improvements from the projects, asking them to pay particular attention to their areas of needs. This would help ESGF move into the future and capture the growing needs of the community it services.

Table 4. *Modeling and data center recommendations.*

Recommendation	Work Description
Task force	Establish a governance board to provide guidance, ideas, and improvements to the ESGF and UV-CDAT governing bodies for better serving the community of users and projects.
Documentation	Provide better installation, publication, data management, analysis, visualization, and use documentation for the distributed software.

5 User's Perspective

A share of our research is focused on the improvement of the user support process. Therefore, in 2013, a survey was conducted with working members who do user support and use ESGF. That is, the population of interest in this research was not only working members who provide support to users but also the many users on projects that access ESGF data holdings. The goal of the research was to improve the user support process and better serve supported ESGF data projects. Based on survey results, a part of the current user support scenario was uncovered and is shown in **Table 5**. It is anticipated that the recommendations from support staffs (and other stakeholders) pave the way towards optimizing the user support process within ESGF. Consequently, these recommendations can be implemented in other ESGF-like climate e-Science infrastructures (such as UV-CDAT), which can in turn be accepted as best practices to streamline user support processes. The survey questionnaire includes 25 recorded responses to 43 questions. The data collected from these questions focused on:

- Basic biographical data of employees, with self-rating of their expertise;
- Support structure;
- Time related factors in user support;
- Communications factors in user support;
- User requests; and
- User-related factors in user support.

Interviews conducted after the survey with the stakeholders (mainly ESGF users) identified the following weaknesses in the ESGF user support process.

Table 5. *User's perspective on ESGF.*

Survey	Findings
Funding	Support activities in e-Science are normally not explicitly funded, admired, or funded at all.
Staffing	Understaffing and lack of time with support staffs to respond and solve user problems or needed project features.
Training	Insufficient training of the user support staff and users. More time and resources needed for investment in training.
Scientific experts	Weak or missing link to the scientific experts. Better connections are needed with scientific projects and their members.
Support activities	At times, user support activities are left to the good will or mercy of people.
Documenting project or user requests	There is no standardized way for referencing and documenting incoming user or project requests.
Single point-of-contact	No single point of contact. There are a number of projects that all have very different or very similar points of views on how ESGF should work.
Support mechanisms	Parallel support mechanisms (RSS and Mailing Lists) in place, but both offer limited support due to their disadvantages.
Measured support	Support activities cannot be measured.
Response time	Response time is not optimal or at times longer especially when a particular support person (especially a specialist) is unavailable to provide a solution or to answer the user query.
Missing information	Missing support information on the supported webpages. Infrequent updates of support webpages and missing road maps on the future direction.

It should be noted that ESGF support staff are "to some extent satisfied" but not "completely satisfied" with the support process. Please refer to **Figure 6**. Most of the respondents response range from "satisfied" (23%) to "somewhat satisfied" (36%). 23% of the respondents remain neutral. However, 9% of the responses of the support staff range from "somewhat dissatisfied" to "dissatisfied" (5%), and 5% are "extremely dissatisfied."

A full user survey report will be coming out in 2014 and next year's face-to-face meeting will cover the users more broadly throughout the federation of ESGF data holdings.

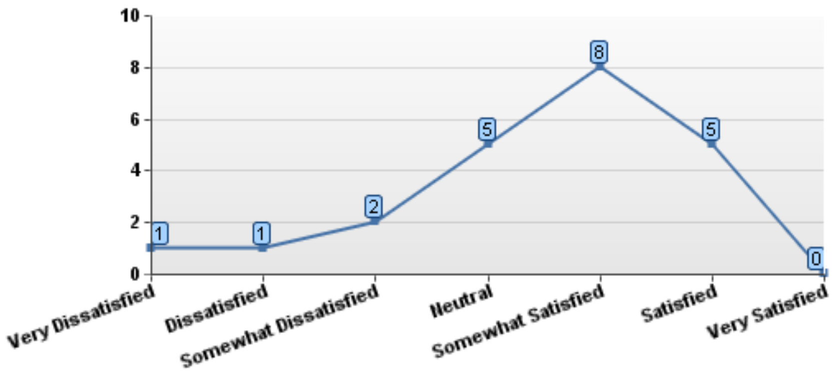


Figure 6. *The responses of the support staff "satisfaction factor" indicate that the support staff is concerned with the level of support.*

Table 6. *Expected milestones for the year.*

Expected Time of Achievement	Milestones
January 2014	Analysis of survey of ESGF employees involved in user support.
February 2014	Field study report of the user support practices.
March 2014	Detailed Analysis of interviews with the stakeholders including users of ESGF.
April 2014	Analysis of communication of users and ESGF support staff via esgf-user mail-list.
May 2014	Survey questionnaire aimed at ESGF users to assess ESGF usage, usability and user satisfaction.
June 2014	Recommendation report to improve the user support process and practices in ESGF.
July 2014	Validation analysis of research results via focus groups including their feedback and further suggestions of improvement.
December 2014	Present finding at the 4^{th} Annual ESGF & UV-CDAT Face-to-Face Meeting.

6 Innovation Technology

6.1 Earth System Grid Federation

6.1.1 Improved Publishing Services

The capability to publish data to ESGF, and make them available throughout the federation, will be improved with several goals in mind:

1. Making it easier for data providers to push data and metadata to ESGF nodes;
2. Improving the richness and quality of the published metadata; and
3. Enabling the publishing of new resource types to the system.

To this end, we will start by providing facilities to validate the syntax and semantics of the metadata records indexed on the server, using a hierarchy of schemas that specify required and optional metadata depending on the project. We will also collaborate with the European IS-ENES2 project to generate and evolve Controlled Vocabularies (CVs) for multiple fields and domains—for example, in support of climate CMIPs. To guarantee the highest quality of data, we will set up and enforce an access control mechanism such that only specific index nodes are authorized to publish certain data sets for certain projects. Finally, we will develop the ESGF Publishing Client to be easily configurable with a set of predefined handlers that is able to parse arbitrary directory structures and publish data, images, documents, and other resources to the system, while harvesting descriptive metadata from the available sources.

Table 7. *Publisher achievable milestones for this coming year.*

Expected Time of Achievement	Milestones
February 2014	Validate metadata validation according to a hierarchy of schemas.
April 2014	Establish a formal development process and upgrade the CV metadata.
June 2014	Develop an access control mechanism for publishing project data from an index node.
August 2014	Release and use the first version of CV metadata.
October 2014	Enable the publishing client to handle arbitrary directory structures and resource types.

6.1.2 New User Interface

We will use the existing CoG web application as the basis for the reference user interface for the underlying ESGF data and metadata services. CoG is already a mature software product that is used operationally to host a variety of scientific projects, including climate modeling software, data downscaling, reanalysis data, and several model intercomparison projects. Currently, CoG is deployed as a single, standalone web portal hosted on a dedicated NOAA server, serving a growing community of hundreds of scientists. Before it can be deployed at ESGF sites, CoG will need to have federation capabilities for searching, security, and project management. Some critical ESGF functionality (like the data cart and integration with Globus Online) will have to be developed. Once integrated with ESGF, CoG will provide a superior interface to ESGF users and data managers in terms of usability, content management, CMIPs and multi-project support, and online collaboration tools. The CoG software stack is based on Python and Django, and will therefore open new possibilities for integration with other Python-based software stacks such as PyWPS (for data processing) and UV-CDAT (for data analysis and visualization).

Table 8. *CoG's milestones for the year.*

Expected Time of Achievement	Milestones
January 2014	Add CoG support for HTTPS protocol. Switch CoG to use Postgres database back-end.
March 2014	Develop CoG data cart. Support Wget downloads.
May 2014	Improve CoG search services (support for replicas and versions).
July 2014	Provide CoG support for OpenID authentication.
August 2014	Provide CoG support for project federation.
October 2014	Integrate with ESGF security infrastructure.
November 2014	Release candidate of CoG for ESGF usage.

6.1.3 Data Transfer and Sharing

Data transfer is a key capability in the ESGF stack, both for replicating data across sites and for users to move data to their machines or to other analysis clusters. As mentioned previously, the Wget solution provided today has limitations in terms of usability, scalability, and the need to always download to the machine where the Wget script is executed.

GO provides reliable, secure, managed data transfer. It leverages GridFTP, a proven protocol for high performance data transfer, and provides platform capabilities allowing clean integration with the ESGF stack. There has been some prototype work done to demonstrate the usefulness of Globus Online in the ESGF stack. This initial version, though successful, could be improved with a better user interface, cleaner single sign-on solutions, and better handling. With the introduction of the CoG user interface, we acknowledge we must have a better integration path of Globus Online functions into the new interface framework.

In addition to the integration of GO as a download option in the ESGF data cart, we plan to provide transfer and sharing interfaces from within the ESGF interface, directly against the files in the various ESGF sites. This functionality will allow power users to have access to the data directories to download the files directly and also share files/directories with their collaborators.

Table 9. *Globus Online milestones for the year.*

Expected Time of Achievement	Milestones
April 2014	Integrate transfer with data cart in ACME UI.
May 2014	Design integration with CoG UI and integrate direct transfer and sharing with ACME UI.
August 2014	Integrate transfer with CoG interface (data cart).
October 2014	Integrate transfer and sharing directly off CoG interface.
October 2014	Provide support for user delegation and OpenID Connect.

6.1.4 Identity, Entitlement, and Access Management System Enhancements

Several enhancements are proposed for ESGF's system for Identity, Entitlement, and Access Management that will support new use cases and to improve the user experience. These include rationalizing the trust roots to use the Open Science Grid (OSG) public key infrastructure (PKI) certificate authority (CA), supporting user-managed groups from a portal interface, simplifying the access control for Wget scripts, and adding functionality for user delegation using OAuth. We list all the proposed enhancements below, in order of priority. In the subsequent roadmap, we propose a schedule for their implementation.

1. PKI Trust roots: To improve the level of trust for the root CA, and to decrease the operational overhead of maintaining a CA for host certificates, we will transition to use the OSG PKI CA.
2. Support for user delegation: A number of use cases have arisen recently for projects using ESGF in which third-party services needed access to resources on behalf of a user. Since this requires a means of user delegation, we will implement a mechanism that uses the OAuth protocol to allow services to obtain short-lived credentials on behalf of users. This will also involve the replacement of MyProxyCA with a new HTTPS-based Online CA service.
3. Attribute registration interface: Users need to be able to register for Virtual Organization (VO)-level attributes such as “CMIP5 Research.” This interface will make registration for access to resources within ESGF easier for users and less error prone.
4. Browser-based sign-in usability: We need to alter the web-based sign-in so that users no longer need to explicitly enter their OpenID but instead simply select their home institution for login. This requires changes to the user interface for the OpenID Relying Party and also the configuration of ESGF OpenID Providers.
5. Wget script improvements: The current system for scriptable HTTP data download relies on users obtaining short-lived X.509 credentials. These can be problematic to set up and use for end-users, so we propose to replace them with a simple cookie-based authentication for single sign-on. This will involve changes to the ESGF OpenID providers and also to the Wget scripts themselves.
6. OpenID Connect: We plan to migrate from the current OpenID single sign-on system to OpenID Connect. The latter is a technology that builds on the OAuth 2.0 protocol and will enable rationalization of single sign-on, attribute release, and user delegation under a single interface. This work builds on items 2 and 4.
7. Identity bridging: Bridging will enable users to link accounts they hold with other federations to ESGF. For example, a user could link their account from a Shibboleth federation to ESGF. This would mean they could use their Shibboleth credentials to log into ESGF.
8. VO attribute distribution: In the current system, services wishing to make access control decisions query a central VO-level attribute service. This presents a single point of failure for the system. As a solution, we propose to implement a system to push VO-level attributes to individual IdPs. IdPs would then push these to services on user login, thus eliminating the need for these services to poll a centralized attribute service.

The above prioritization has been determined based on a combination of beneficial impact and ease of implementation. Some examples illustrate this approach: improvement to the model for VO attribute distribution is an area that could be of great benefit to the operational robustness of ESGF but may require significant implementation effort. This task is left off the roadmap, subject to a later reappraisal based on the progress of the other tasks. Note as well that some tasks are interdependent. For example, the identity bridging work is best tackled once the OpenID Connect work is completed.

Table 10. *Milestones for security enhancements for the year.*

Expected Time of Achievement	Milestones
May 2014	Implement PKI trust roots and new OnlineCA service (replace MyProxy).
May 2014	Implement browser-based sign-in usability.
June 2014	Improve Wget script.
July 2014	Create attribute registration interface.
October 2014	Provide support for user delegation and OpenID Connect.

December 2014	Implement identity bridging.

6.1.5 Monitoring the Earth System Grid Federation (Dashboard)

The capability to monitor the ESGF infrastructure with a comprehensive set of metrics will be a primary goal of the Dashboard module. Most of the planned activities will be strongly related to the back-end of the system to address a flexible and easy management of multiple metrics.

The most relevant future activities will be related to:

1. Programmatic support for developers in terms of new REST-like APIs;
2. A configurable and extensible framework for monitoring sensors;
3. Long-term metrics management support (both in terms of storage model and clients);
4. New sensors to fully address key user requirements (e.g. need for project-based data download metrics);
5. Stronger support for real-time monitoring capabilities; and
6. Desktop based environment displaying in the same views multiple gadgets/metrics at the same time;

To this end, we will start working on the back-end of the system to make it more configurable and extensible. A framework for long-term metrics management based on files rather than relational databases will be designed and implemented. This will allow us to keep track of long time series in the monitored metrics. A set of clients to subset or aggregate the stored data will be also needed.

The framework will be plugin-based and extremely configurable and extensible. This way, adding new sensors (e.g. for site level needs) will be straightforward. Changing sensors settings will be possible just through a configuration file.

The activity will be strongly connected to the European IS-ENES2 project, as it is also part of a Work Package regarding the data management infrastructure. In this regard, a report on the design of the new Dashboard will be delivered by the end of March. It is important to note that relevant inputs and feedback are expected this year by the CORDEX project and the impact community.

A solution to manage network-based metrics will be also explored. A set of algorithms to get this information will be analyzed to identify the best and most stable one to be adopted at the federation level.

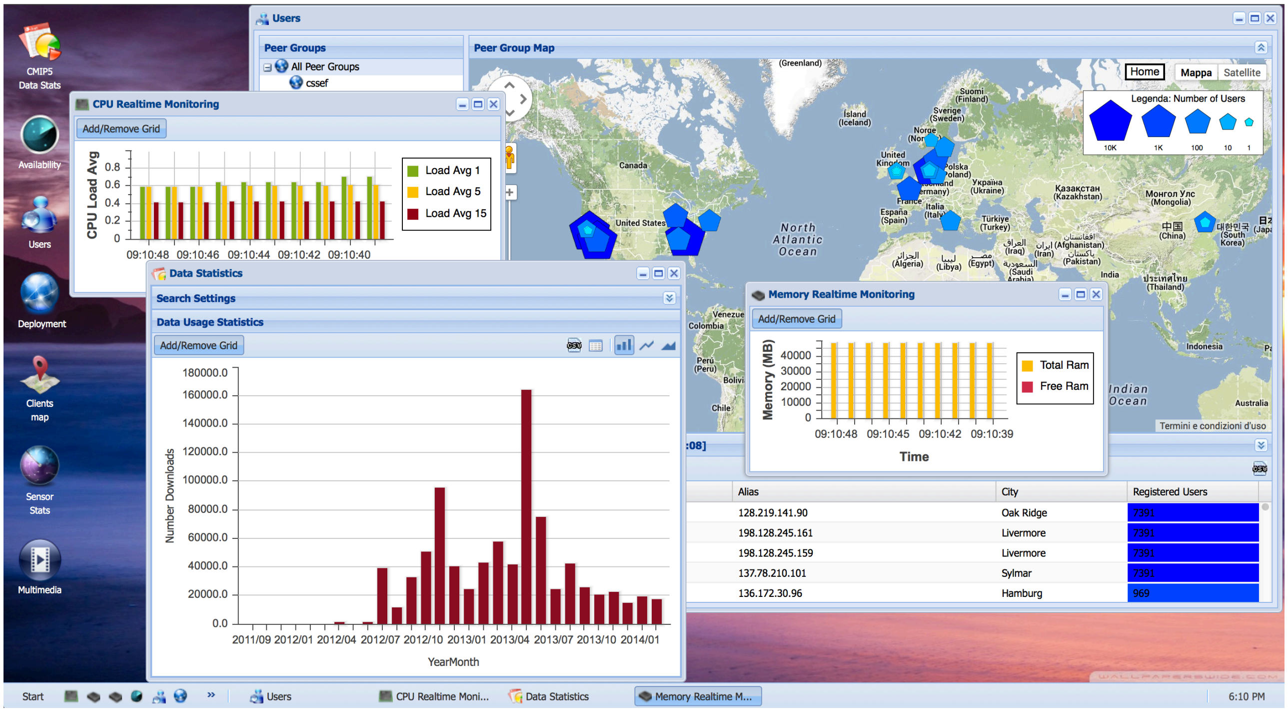


Figure 7. *A snapshot from the current release of the Dashboard.*

Several enhancements are proposed for extending the dashboard and improve the monitoring of the Earth System Grid Federation. These will support new use cases and communities (e.g. impact community), and thanks to new web interfaces, will try to strongly improve the user experience. We summarize these below, in order of priority. In the roadmap, which follows, we propose a schedule for their implementation.

Table 11. *Dashboard achievable milestones for the year.*

Expected Time of Achievement	Milestones
March 2014	Design of the dashboard framework (back-end and front-end). Documentation on the wiki about the existing system.
June 2014	New configurable back-end of the system for long-term metrics + a set of sensors. Documentation updated and tutorials created.
August 2014	REST-like APIs for programmatic access to the metrics.
October 2014	New set of metrics for network and download statistics .
December 2014	Real time metrics and web interface improvements/extensions.

6.1.6 UV-CDAT Integration

DOE-sponsored UV-CDAT will be the basis for the much-needed remote processing and visualization within the ESGF's enterprise system. The integration will key on remote processes needed for large-scale data processing for the CMIP and ACME projects, which will include model metrics, diagnostics, and comparative visualization and statistical analyses. These and other CMIP developments will heavily leverage the work of the core UV-CDAT team, such as parallel framework development and remote web informatics visualization. Additional work will be needed to interface the new ESGF CoG interface and to integrate the ESGF's workflow and provenance capabilities with UV-CDAT's VisTrails workflow and provenance framework.

Table 12. *UV-CDAT integration into ESGF milestones.*

Expected Time of Achievement	Milestones
January 2014	Update UV-CDAT's thick client interface to the latest release of ESGF.
February 2014	Prototype the integration of UV-CDAT server-side capabilities with diagnostics.
March 2014	Prototype the integration of UV-CDAT Live (smart client), diagnostics, and ESGF for remote processing services and visualizations.
July 2014	Release the beta version of UV-CDAT and UV-CDAT Live to the ACME community.
August 2014	Release the official version of the UV-CDAT and UV-CDAT Live that is fully integrated with the ESGF software stack.
September 2014	Document how to use UV-CDAT and ESGF together.
October and November 2014	Add model metrics and additional diagnostics to the UV-CDAT and ESGF remote processing services.

6.1.7 Tutorials and Documentation

To support both the user and developer communities, the core ESGF team—LLNL, Oak Ridge and Argonne national laboratories, NASA, NOAA, NSF, IS-ENES2, and NCI—will apply progressive open-source documentation and proven tutorial tools (i.e., Github, Sphinx, Docutils, MaxMind, YouTube, and others) to deliver a robust user and developer support system. A formal written document on ESGF use is in progress. In addition to documentation and training, we will write book and report content. The ESGF team will also continue to publish journal articles about ESGF; they currently have several under peer review (recent publications can be found on our ESGF website). Further, we will support outreach activities and present at major conferences, such as the American Meteorological Society and American Geophysical Union.

We expect to build an online library of tutorials by archiving a bi-monthly online webinars, for hands-on classes and remote learning. Our initial tutorials will be designed to help users get started with ESGF, while many additional tutorials will be created later in the year, with community input. Viewing online will require only a web browser and Internet connection, will be entered on the ESGF website. See the ESGF website for first/initial webinars as we gear up to produce many more.

Table 13. *ESGF online tutorials and documentation milestones for the year.*

Expected Time of Achievement	Milestones
May 2014	Begin video tutorials on installing the systems at modeling and data centers. Begin video tutorials for basic user features.
June 2014	Begin documentation based on videos and start Frequently Asked Question (FAQ) list for the website. Begin videos of advanced user features, including "help."
August 2014	Document advanced user features.
September 2014	Complete web documentation of basic features and prepare report of progress.
October and November 2014	Complete web documentation of advanced features in preparation for next face-to-face meeting.

6.2 Ultra-scale Visualization Climate Data Analysis Tools

6.2.1 Parallel Visualization and Analysis

UV-CDAT provides a number of innovative mechanisms for exploiting high-performance computing resources to analyze and visualize data sets using distributed memory parallelism. Parallel visualization within UV-CDAT is supported through VisIt and ParaView. Data reduction of large-scale data sets to generate summary statistics and climatologies is supported through ParCAT, as shown in **Figure 8**. Temporal–spatial parallelization is supported through both ParaView and ParCAT. Distributed array-based parallelism is supported through MPI-2 one-sided operations coupled with a global address space paradigm within UV-CDAT. These complimentary parallel analysis and visualization mechanisms cover a wide range of common use cases in ultra-scale climate data processing. These techniques are increasingly required to handle the large-scale climate model data sets that are hundreds of gigabytes to several terabytes in size and larger.

Diagnostics, model intercomparison, and visualization are required capabilities for post-processing and effective manipulation of this data by scientists. Fortunately, many of these post-processing tasks are “pleasantly parallel,” including statistical summaries and single-task, multiple-data workloads. The technologies that enable UV-CDAT to provide these capabilities and more include Distributed Arrays, Paraview, the Spatio–Temporal pipeline, ParCAT, and TaskFarmer. Each of these tools is designed to handle a subset of the use cases we have identified to be of relevance to the climate science community. Our current work is focused on integration of these techniques more directly within the UV-CDAT core infrastructure and consolidation of these approaches where applicable. This plan and the overall design is captured in our “UV-CDAT Parallel Design Efforts for Project Enterprise—Design Document.”

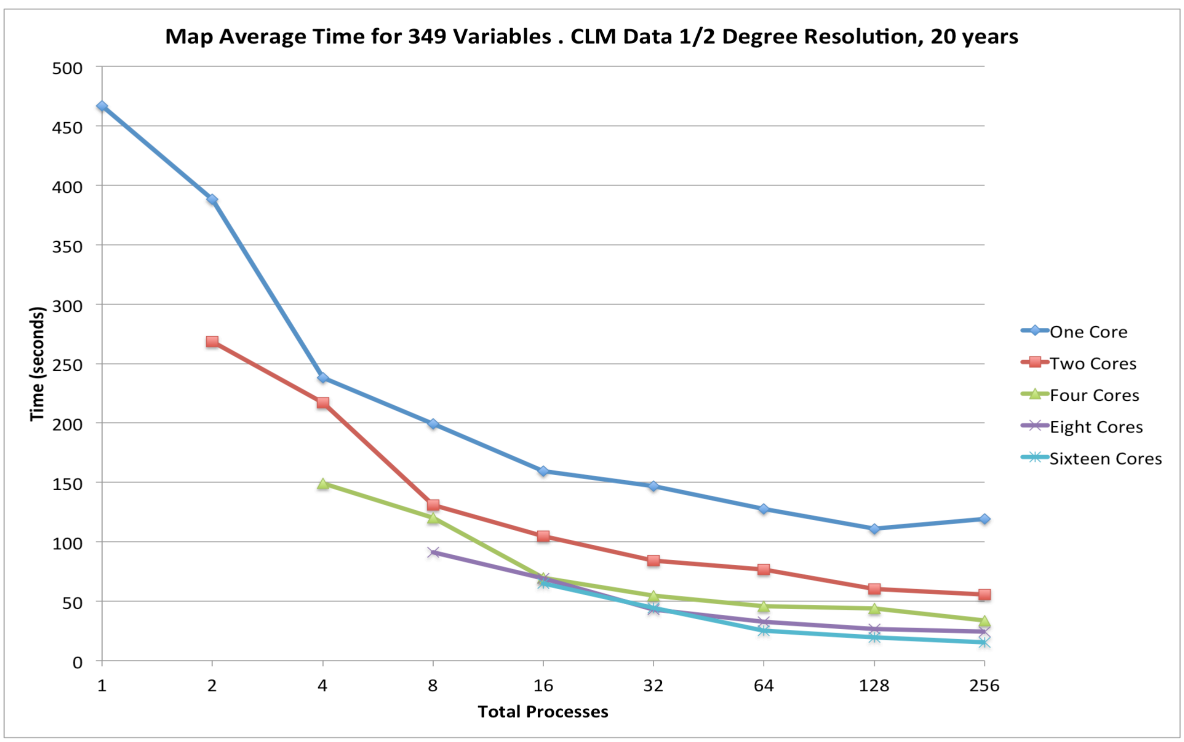


Figure 8. *Parallel speedup of map average routing.*

Table 14. *Parallel visualization and analysis expected milestones.*

Expected Time of Achievement	Milestones
January 2014	Complete the design document and planning for parallel visualization and analysis.
May 2014	Complete a detailed design and implementation plan for parallelization of cdutil.averager and similar functions.
September 2014	Prepare initial prototype of parallel cdutil.averager.
October 2014	Parallelize additional similar functions.
November 2015	Prepare a detailed design and implementation plan for pipelining CDMS operations and integrated parallel I/O with pnetCDF.

6.2.2 Web-based Analysis and Visualization

At present, the majority of the climate science community still relies heavily on primitive analysis and visualization tools that are based on the thick (or fat) client application concept, meaning that the user must download software to appropriate machines or hardware where the data resides (e.g., laptops, desktops, or HPC machines). In such cases, users can become frustrated by installation challenges, such as finding the right prerequisite software packages, software versions, and currently supported hardware and operating systems.

Analysis and visualization tools have thus begun moving toward the thin client application concept, where users install very little software—in most cases, only a web browser is needed. In this case, the analysis and visualization software is deployed on a central sever rather than each individual system, eliminating users' installation and operating system requirement challenges. Thin clients are well suited for environments in which the same information is going to be accessed by a general group of users, but the tradeoff is that thick clients provide users with more features, analysis and visualization, interaction, and other choices that make the software more customizable. In contrast, the thin clients appear to be a bit more primitive and have fewer features.

For our web-based analysis and visualization system, UV-CDAT Live, we consider it to be neither a thick nor a thin client, but rather a smart client. That is, it is based more on the traditional client–server architecture concept within the web-based model, as shown in **Figure 9**. It is more similar to the thick client concept in that the UV-CDAT Live smart clients are Internet-connected devices that allow user's local applications to interact with server-based applications through the use of web services. This allows for more analysis and visualization interaction and software customization, just like thick clients. On the other hand, like thin clients, the user download minimal software. In most cases, only the software applications needed are the appropriate web browser and associated runtime libraries, such as Flash or Java. With the UV-CDAT Live, user will be able to work offline and access the appropriate local data or connect to the Internet and access distributed data via ESGF. With this concept, UV-CDAT Live will be have the capability to be deployed and updated in real-time over the network from a centralized server without the notices of users; support multiple platforms and operating systems; and run on almost any device, such as mobile phones, notebooks, tablet PCs, laptops, desktops, and HPC machines.

The main goal of this design is to facilitate collaboration among distributed users by connecting them through a user-friendly web-based system to conduct visualization and analysis on both real-time simulation and archived data sets with intuitive visual output and online steering capability. The users access the UV-CDAT Live system, shown in **Figure 10**, solely through a web interface, hence obviating the need to install any third-party software packages to perform daunting tasks such as simulation execution, parameter control, data visualization and analysis, and information sharing. Although the system is standalone, it is part of the large ACME end-to-end system that will allow users to run models and collect workflow and provenance for sharing scientific results and reproducibility. The system is directly connected to the ESGF distributed archive, requiring all users to log in at the beginning of each session. ESGF login accounts are required for authentication and authorization before any data is accessed, but no login is needed for local data or for ESGF data that is open to the public.

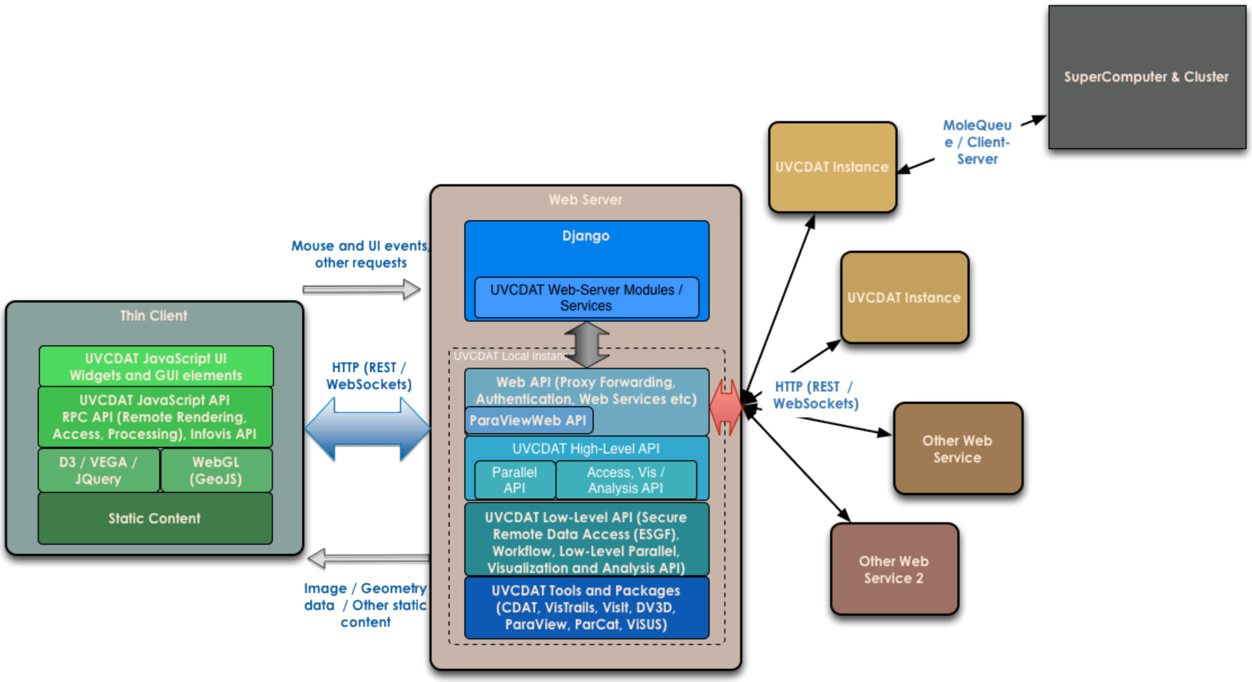


Figure 9. *The UV-CDAT client-server architecture design.*

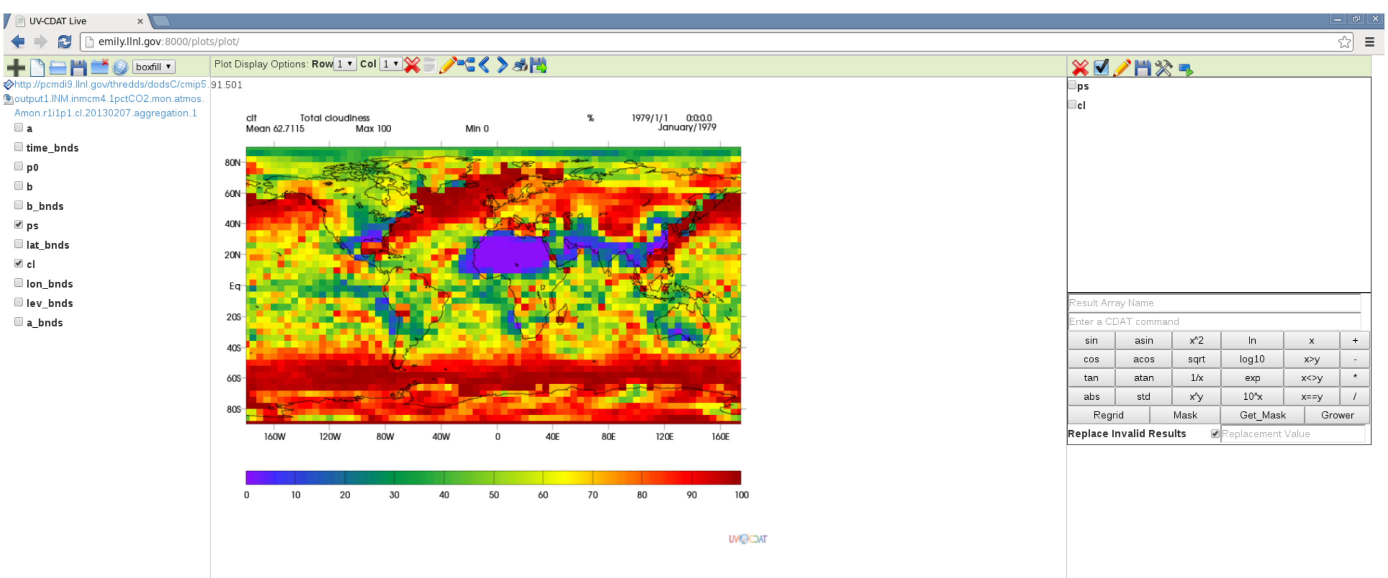


Figure 10. *The interactive UV-CDAT Live web-based analysis and visualization system looks similar to the UV-CDAT thick client, but without the hassle of installation.*

Table 15. *UV-CDAT Live web-based analysis and visualization milestones for 2014.*

Expected Time of Achievement	Milestones
January 2014	Put the UV-CDAT Live repository in Git and on GitHub. Establish a connection to the ESGF distributed archive. Complete the basic web-based template for climate data analysis and visualization.
February 2014	Start the preliminary integration of DV3D visualization into UV-CDAT Live. Test UV-CDAT Live for early-March ACME demonstration.
March 2014	Begin the workflow and provenance representation in UV-CDAT Live.
April–May 2014	Beta test workflow and provenance in UV-CDAT Live.
July 2014	Release official version of UV-CDAT Live to the community.
October and November 2014	Integrate more diagnostics and model metrics into UV-CDAT Live.

6.2.3 Server-Side Analysis

One of the important components of UV-CDAT Live will be the server-side analysis and visualization. Server-side computation is necessary as the increase in data size and complexity of algorithms lead to data-intensive, compute-intensive challenges for climate data analysis and visualization. In this architecture, a web-server receives request for data processing, analysis, or remote rendering from a thin client such as web browser. A prerequisite to this is the user authentication, which involves the user providing a username/password to the server. If the prerequisite is satisfied, then upon receiving the request, the server may create a process specific to the user's session (per session ID). A separate process per session strengthens the security and offers solutions for scalability if needed. Any subsequent requests for data processing are then forwarded to this process until the session ends, at which time the process gets deleted on the server side. Certain requests that do not require intensive computing are served directly by the web-server. For instance, accessing pre-computed climatology for diagnostics is served directly by the server. The server delivers most of the static content to thin clients as needed.

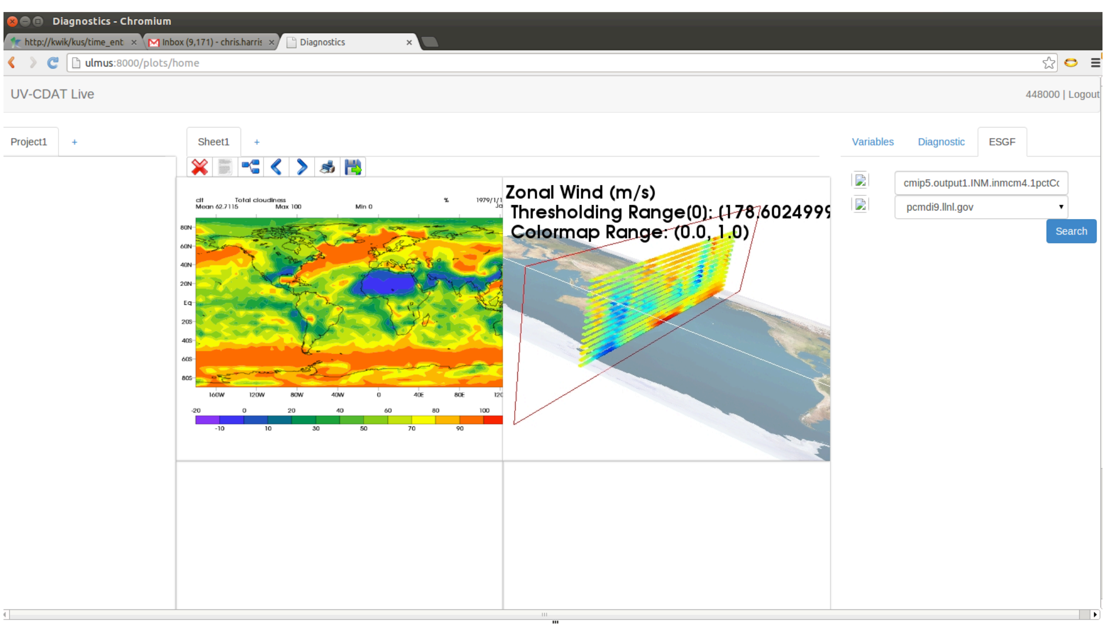


Figure 11. *A 3D visualization is shown in UV-CDAT live interface. Processing and rendering is performed on the server side at interactive frame rates.*

In the design of server-side analysis, we have adopted standard protocols and communication channels between the client and the server. Most of the static content and some dynamic content for exploratory analysis are served in a RESTful manner. The client can access this content by invoking an AJAX call over the web. For interactive visualization and analysis, data is accessed via WebSockets, as it provides the required flexibility and interactivity. The UV-CDAT live server-side analysis is built on top of an open-source parallel remote data processing framework known as ParaViewWeb. ParaViewWeb is a collection of components that enables the use of parallel visualization and data analysis capabilities within web applications.

Using the latest HTML 5.0-based technologies, such as WebSockets and WebGL, ParaViewWeb enables communication with a ParaView server running on a remote visualization node or cluster using a lightweight JavaScript API. Using this API, web applications can easily embed interactive 3D visualization components. Application developers can write simple Python scripts to extend the server capabilities, including creating custom visualization pipelines.

We have chosen Python as the binding language for the server-side analysis. Python was the natural choice for the framework because of its support in scientific computing community and widespread use in almost every field of computer science. Also, a Python-based platform enabled us to integrate existing UV-CDAT source code on the back end with minimum effort.

Table 16. *Server-side analysis and visualization milestones for UV-CDAT.*

Expected Time of Achievement	Milestones
January 2014	Prototype basic analysis and 2D plotting capabilities.
February 2014	Provide 3D plotting capabilities and prototype exploratory analysis.
March 2014	Second iteration. Develop JavaScript for accessing server-side capabilities. Integrate with the industrial-strength server.
April 2014	Third iteration. Develop Python API for processing and distributing computing tasks for analysis and remote rendering.
July 2014	Deliver fully functional API for front and back end.

6.2.4 Diagnostics

The UV-CDAT flexible diagnostics framework is being developed to identify problems within the DOE ACME coupled model running on DOE leadership-class facility computers. These diagnostics test the individual sub-model components (i.e., atmosphere, land, ocean, sea ice, and land ice) of the larger system for issues and help to inform model developers and users of potential problems or inconsistences in model run. The comprehensive diagnostics framework will incorporate the following standard diagnostic packages:

- The Atmosphere Model Working Group (AMWG) for atmosphere model testing;
- The Ocean Model Working Group (OMWG) for ocean model testing;
- The Land Model Working Group (LMWG) for land model testing; and
- The Model for Prediction Across Scales (MPAS) diagnostic suite for land ice and sea ice model testing.

In past years, some of these diagnostic packages, such as the AMWG diagnostics, used for atmosphere model testing were done exclusively in batch mode via the NCAR Command Language. With the increase in technology, ACME model runs will generate climatology and departure data files and store them directly in the ESGF distributed archive for interactive or batch UV-CDAT diagnostics production. Within the ACME test bed design, these diagnostics can be generated while the model is running for immediate feedback on model performance. Whether running the diagnostics interactively, batch, or in situ, users will need little or no training to use the diagnostics suite.

The DOE laboratories that are responsible for providing the sub-model component diagnostics and their training documents are:

- LLNL, responsible for the overall diagnostics framework and the AMWG diagnostics;
- ORNL, responsible for LMWG; and
- LANL, responsible for OMWG and the MPAS diagnostic suite.

As stated above, this wide range of diagnostics tools are necessary to diagnose the many detailed levels of the model components. As the ACME project progresses, more diagnostics will be added to each component diagnostics suite and additional diagnostics will be develop to cover additional model behavior.

The diagnostics software is integrated into UV-CDAT in three ways. The first uses the UV-CDAT GUI front-end thick-client interface, as shown in **Figure 12**. In this instance, the GUI and the diagnostics computations both run on the same computer. The second way to run the diagnostics is via UV-CDAT Live, as shown in **Figure 13**. Here, the UV-CDAT Live runs in the user's web browser and the diagnostics computations run on a remote machine where the data may be co-located, therefore limiting large-scale data movement. In both of these cases, simple menus let users quickly specify plots and UV-CDAT interfaces allow the users to manipulate the plots interactively in new ways, previously unavailable with static diagnostic scripts. The third method, shown in **Figure 14**, uses a dynamically generated tree view inside a web browser. The user can select data sets and diagnostic sets and then variables and climatologies of interest. Clicking on a leaf node (e.g., my_dataset->lmwg->set1->TG->MAR) will generate that specific plot (if it is not already cached). This is meant as a more dynamic and better-organized replacement for the static web pages generated by the legacy NCAR diagnostic scripts. Users can compare just the plots needed without having to search for them in a flat, text-based web page.

The web-based diagnostic options (UV-CDAT-Live and the tree viewer) will make the diagnostic software (and results) available for users without downloading additional software. Only a web browser is needed. With minimal web searching, nearly any user can find a wide range of climatology data set from within ESGF. This offering will speed up model diagnosis and reduce the transmission of data.

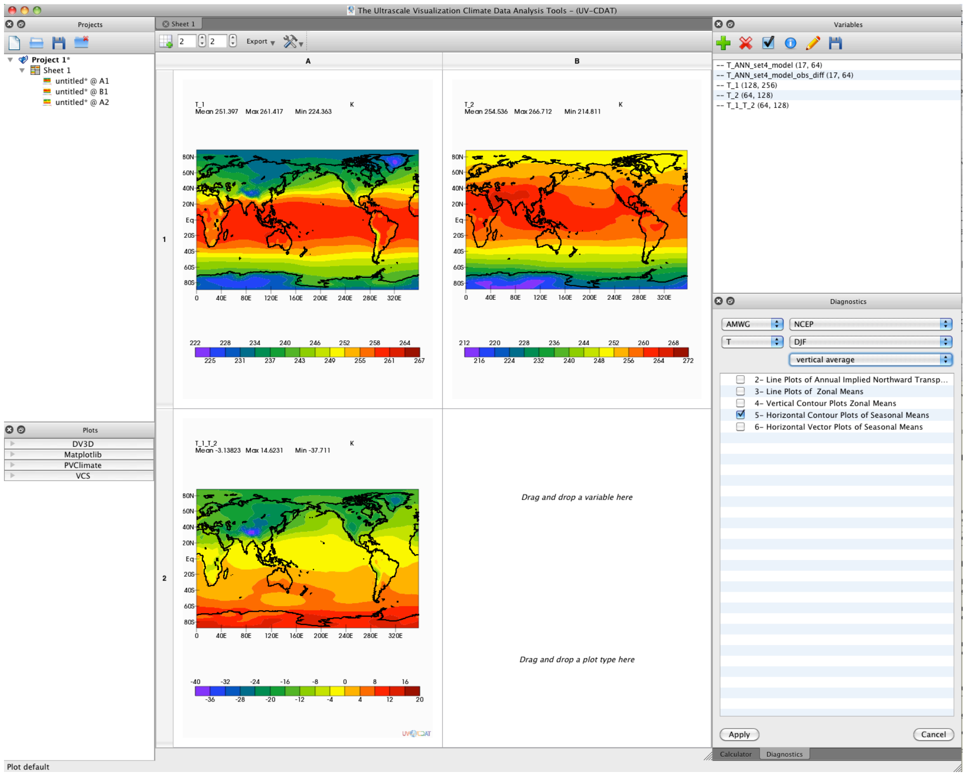


Figure 12. *UV-CDAT thick-client shows an AMWG diagnostics. The user selects the type of diagnostics (e.g., AMWG, LMWG, OMWG, MPAS), the type of observation data set, the variable, the season, and the specific diagnostics. UV-CDAT finds the appropriate climatology file in the distributed archive (i.e., ESGF) and generates the appropriate plots and diagnostics for further interactive exploration.*

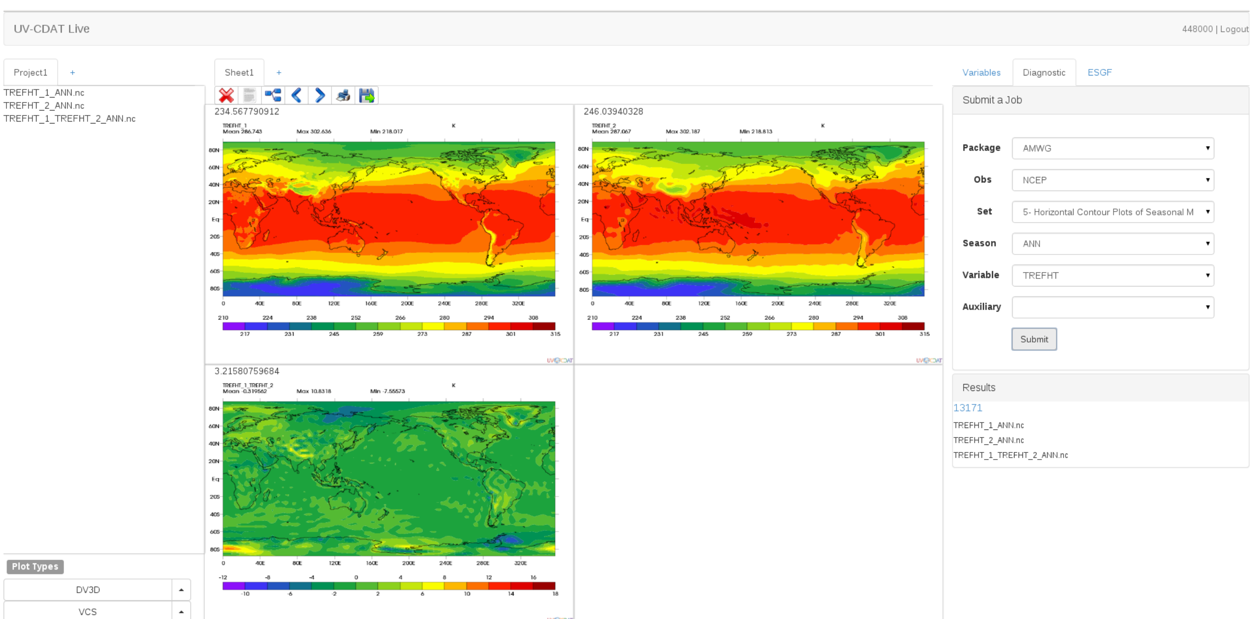


Figure 13. *UV-CDAT Live is a thin client of UV-CDAT that runs on any give web browser. As with the thick client, the user selects the type of diagnostics, observation data set, variable, season, and specific diagnostic for plot generation. As with the thick client, once the diagnostics are generated, the user can interactively manipulate the plot and diagnostic variables for further exploration.*

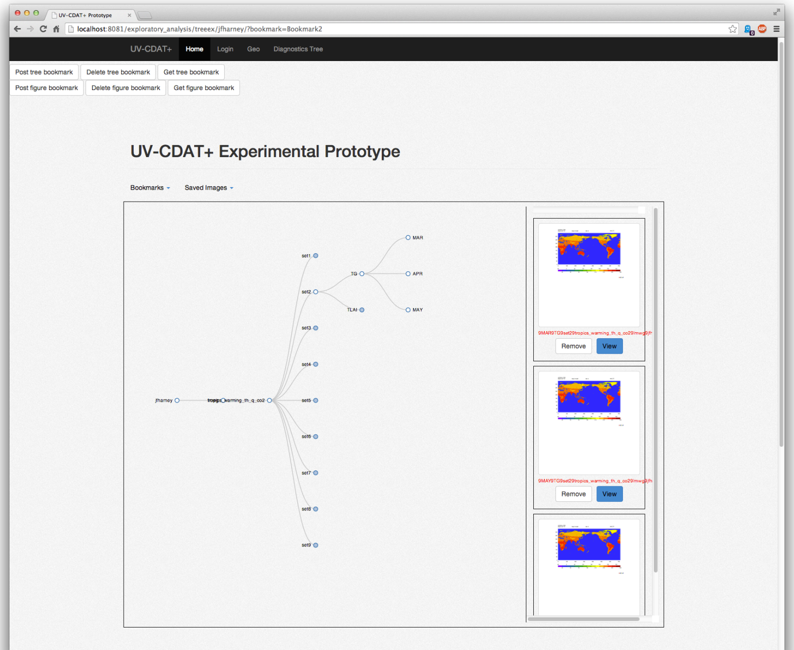


Figure 14. *The tree-view diagnostics, where the user traverse a tree branches to pick the leaf (or diagnostics) of interests. The diagnostics of interests are shown on the far right.*

Table 17. *Diagnostics framework milestones for 2014.*

Expected Time of Achievement	Milestones
January 2014	Ensure that the basic diagnostics infrastructure is in place. Implement AMWG and LMWG diagnostics.
February 2014	Merge the AMWG and LMWG diagnostics and integrate with UV-CDAT GUI and UV-CDAT Live. Integrate with the visual informatics back end.
March 2014	Improve labels and other plot features.
April 2014	Implement more diagnostics as needed (a continuing project, especially for the next six months).
May 2014	For batch use, add a capability to write plots as graphics files.
June 2014	Enhance the present batch script for climatologies, to do other data reductions as needed.
July 2014	Integrate the scripts for data reduction and graphics files with the overall workflow, so they can be invoked during or after a model run.
August 2014	Continue to add additional diagnostics to framework (OMWG).
September 2014	Test implementation of new addition (OMWG) and get user feedback.
October 2014	Continue to add additional diagnostics to the framework (MPAS).
November 2015	Test implementation of the new addition (OMWG) and get user feedback.

6.2.5 Exploratory Analysis

The UV-CDAT Exploratory analysis framework, shown in **Figures 15–19**, aims to provide climate scientists, model developers, and other stakeholders with a rich visual analytics interface for data exploration and analysis. One of the most challenging tasks in exploring multivariate data is the identification and quantification of associations between a set of interrelated variables. For climate model data, this task is even more daunting due to the increasing complexity, numerical fidelity, and number of variables that are considered driven by technological advances. Often times, the scientist has some understanding of the expected relationships, based on experience and background knowledge, but unexpected discoveries are nearly impossible with conventional climate analysis tools. Ironically, the size of the data is both the root of most difficulties with climate data analysis and the inspiration for delivering the next round of model improvements and scientific insight.

Based on prior work on the Exploratory Data Analysis Environment to provide an efficient framework for interactive knowledge discovery and hypothesis testing for climate model analysis, we have designed a new web-based visual analytics framework for UV-CDAT

written in d3.js and to harness the high-bandwidth human visual channel with interactive coordinated views that guide the scientist to the significant associations in the data. Our work [Steed, 2013] was conducted with the climate model scientists who corroborate the notion that an interactive visual analytics framework can provide a more efficient environment for climate model analysis. Furthermore, our research addresses an important point brought out in the NIH/NSF Visualization Challenges Report, which encourages visualization researchers “to collaborate closely with domain experts who have driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs.” The exploratory data analysis prototype permits investigation of time + space dimensions of variables from full model simulations in a client-side web browser.

We also have developed a prototype multi-variate visualization tool in d3.js for exploratory analysis of model data taking a more abstract, information visualization approach, as well as a level-of-detail algorithm for an interactive correlation-mining tool for model data that automatically streams in additional levels of detail as the scientist zooms into the graphical correlation matrix. This tool is also runs in the user’s web browser.

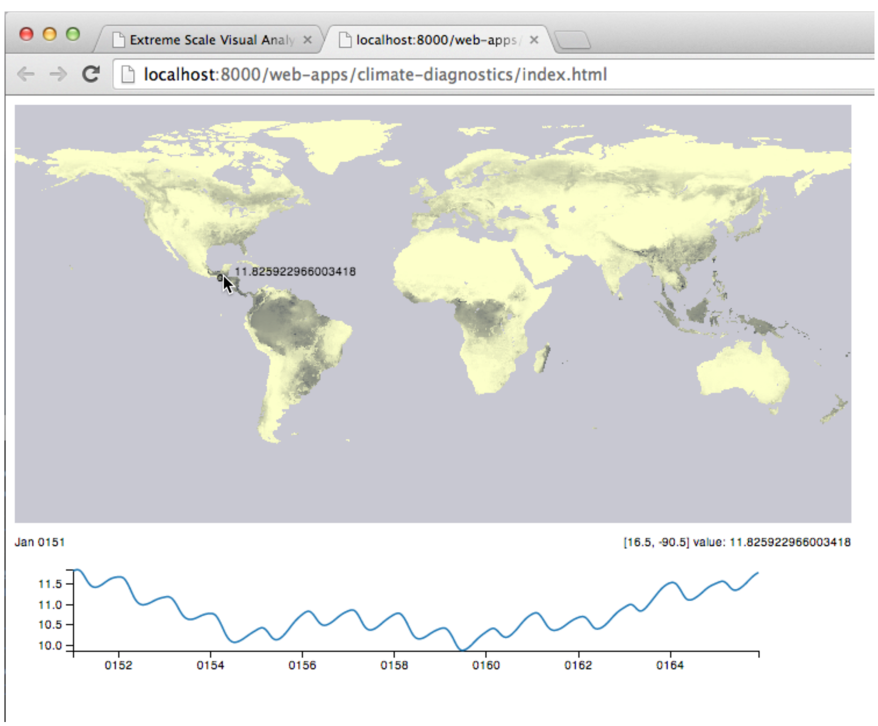


Figure 15. *Coordinated map view with region-of-interest selector.*

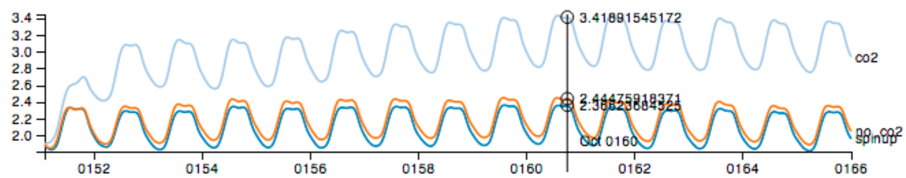


Figure 16. *Multiple trend lines view with point of reference value display.*

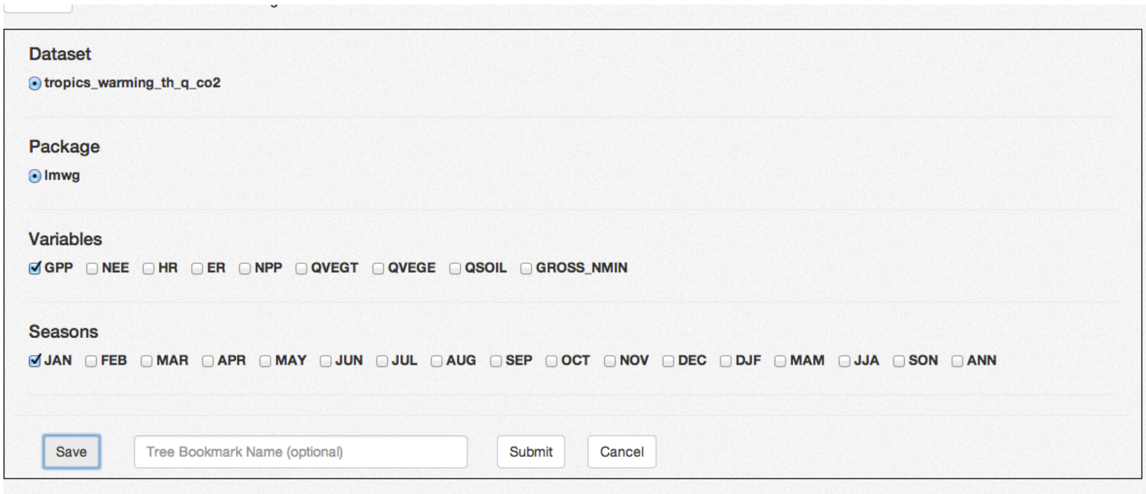


Figure 17. *Data set selector interface.*

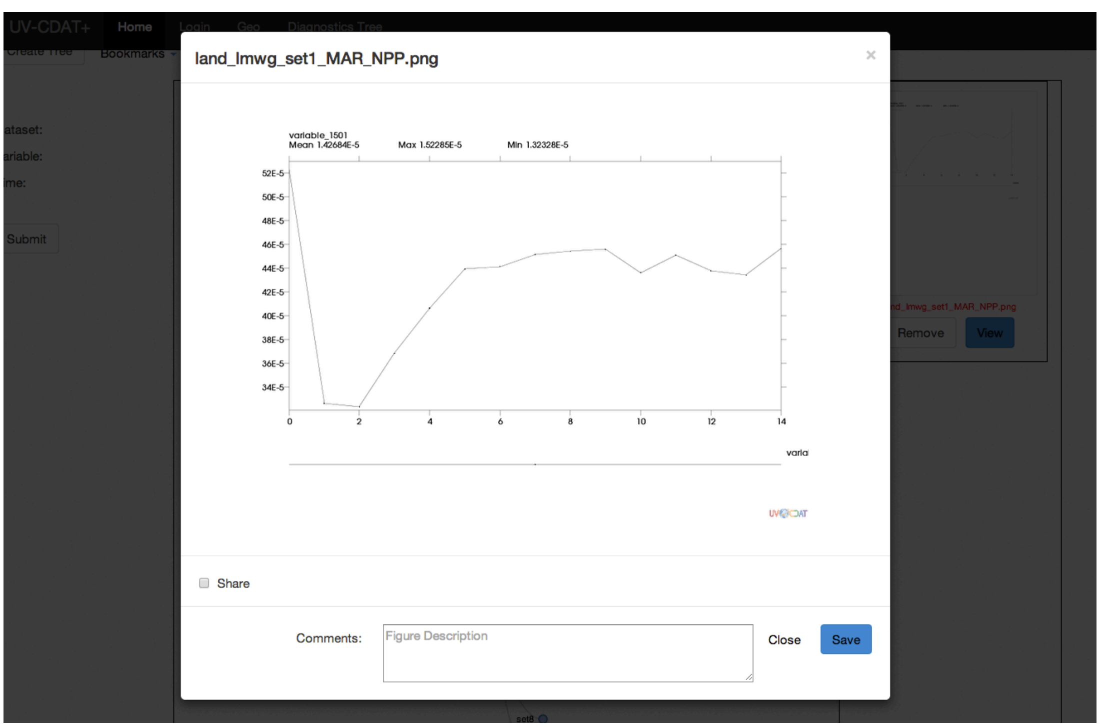


Figure 18. *Visualization bookmarks.*

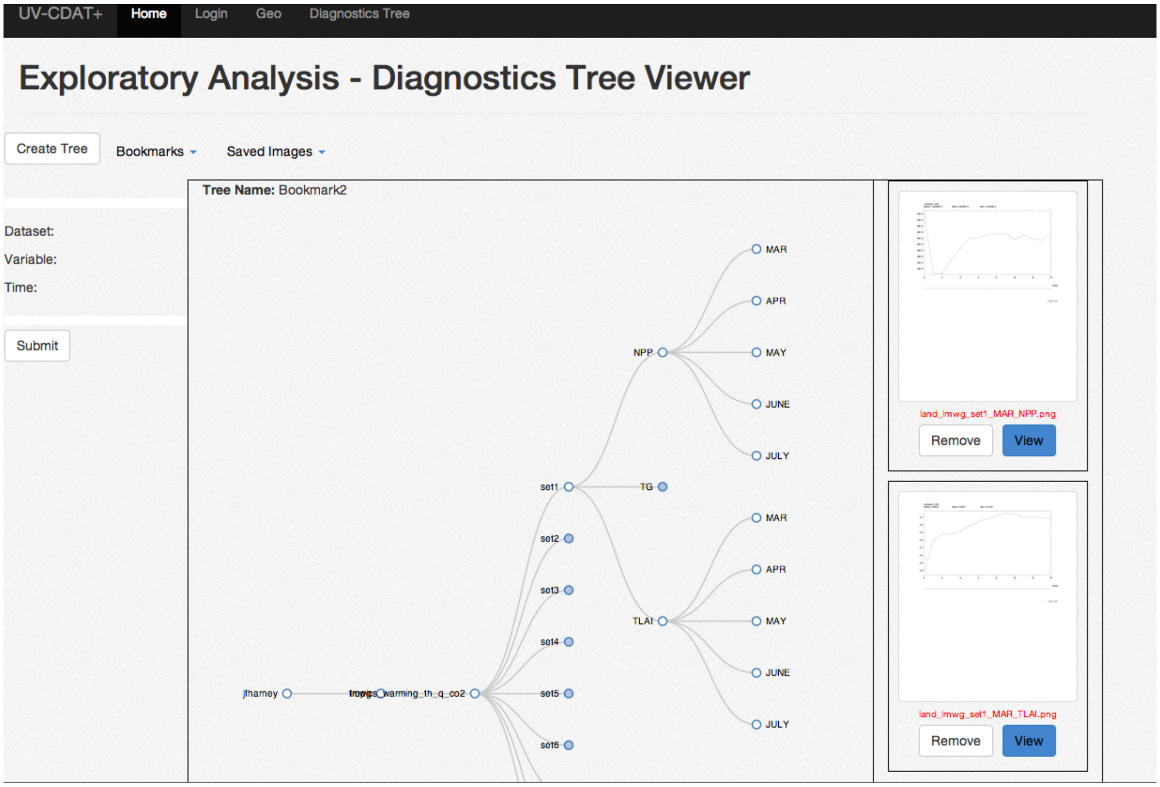


Figure 19. *Tree viewer allowing users to easily navigate complex diagnostics.*

Table 18. *Exploratory analysis expected framework milestones for UV-CDAT.*

Expected Time of Achievement	Milestones
January 2014	Ensure initial prototype is functional within a d3.js infrastructure.
March 2014	Integrate front-end view with the Django back end server infrastructure.
June 2014	Integrate parallel coordinate views and additional coordinated views, including heat maps.
October 2014	Prototype the climate notebooks feature, allowing users to save and recall a collection of defined plots and visualizations.

6.2.6 1D and 2D Climate Data Visualization

The Visualization and Control System (VCS) is currently the de facto standard 1D and 2D graphics package for UV-CDAT. Especially designed for climate change research and conforming to the netCDF CF convention, VCS allows users to have complete control over their graphics. That is, by specifying the desired netCDF CF data, graphics method, and picture template, the user can control the appearance of the displayed data, associated text, and animation. Within UV-CDAT, these can be modified and controlled interactively, command line, in batch mode, or in a Python application.

Because it is so flexible, wide ranges of graphical displays are predefined and users can create and share new ones. For the climate community (including CMIP, ACME, and other projects), users have predefined graphical output specific to their diagnostic climate needs. For these projects, VCS provides ten methods for graphically representing 2D data: boxfill, isofill, isoline, outline, outfill, meshfill, scatter plots, vectors, taylordiagram, and continents. For 1D data, VCS provides three graphics methods: Xyvsy, Yxvsx, and XvsY. There is a graphics method table for each 1D and 2D method that contains predefined and custom graphics methods. An entry in the graphics method table contains: the name of a method for projecting the graphic representation of data onto a display in the UV-CDAT graphics spreadsheet canvas. For example, **Figure 20** shows the "default" graphics methods for each 1D and 2D plot type represented in VCS.

The graphics method defines the how to display the netCDF CF data in the UV-CDAT graphics spreadsheet canvas, and the picture template determines the location of each picture segment, the space to be allocated to it, and the related properties relevant to its display. More specifically, the picture template attributes describe where and how segments of the picture will be displayed, if at all. On the plot, segments are graphical representations of: mean, maximum, and minimum data values; axes; tick marks; labels; boxes; lines; and a legend that is associated with specific graphics methods. Picture templates describe where to display all segments, including the data. VCS also allows for control of secondary objects, such as color maps, fill areas, lines, markers, and texts.

In the process of updating VCS to make it more compatible with more modern software within the UV-CDAT software stack, we have decided to replace the XGKS and Cairo VCS backend graphics with VTK. This conversion to VTK will not only allow better performance and compatibility with other visualization tools but also allow us to port UV-CDAT to previously restricted platforms (such as Windows and other non-Unix/Linux platforms) with relative ease. This conversion will allow VCS users more control over their 1D and 2D plots and the VCS developers will not have to worry about hardware and operating systems specific details. Currently, DV3D, VisIt, and ParaView all use VTK as their back-end visualization package.

In the process of updating VCS to make it more compatible with more modern software within the UV-CDAT software stack, we have decided to replace the XGKS and Cairo VCS back-end graphics with VTK. This conversion to VTK will not only allow better performance and compatibility with other visualization tools but also allow us to eventually extend VCS to support 3D graphics. It will also enable more control over VCS 1D and 2D plots and the VCS developers will not have to worry about hardware and operating systems specific details. Currently, DV3D, VisIt, and ParaView all use VTK as their back-end visualization package.

We plan to:

- Port all existing graphics methods and objects to VTK and make 1D more user friendly;
- Make sure the new VCS objects have hooks to be further used within the VTK framework by the users; and
- Create new VCS graphics methods, both 2D and 3D (probably by merging with DV3D).

The VCS development team will maintain a healthy VCS/VTK development branch as well as the current VCS/XGKS and VCS/Cairo backend graphics capabilities until all features are included in the VCS/VTK branch. Users will be unaware of the back-end visualization change other than the ability to use UV-CDAT on machines that were previously restricted.

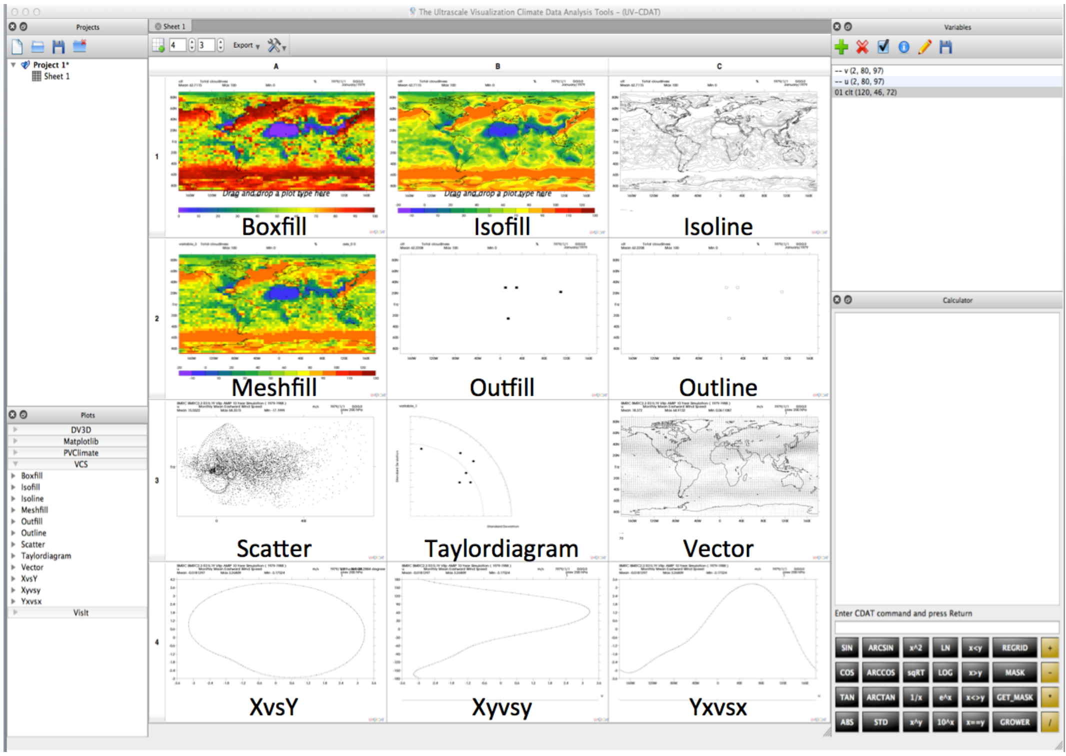


Figure 20. *The image shows 12 of the 13 VCS plot types. From left-to-right, top-to-bottom, the visualization spreadsheet shows the following VCS plots: boxfill, isofill, isoline, meshfill, outfill, outline, scatter, taylordiagram, vector, xvsy, xyvsy, and yxvsx. A picture template defines each VCS image in visualization spreadsheet, plot type (boxfill, isofill, etc.), and data.*

Table 19. *VCS 1D and 2D visualization milestones for 2014.*

Expected Time of Achievement	Milestones
January 2014	Establish proof of concept for existing VCS capabilities.
March 2014	Isolate the XGKS and Cario graphics back end from VCS.
April 2014	Integrate VTK into VCS.
May 2014	Test the alpha version of VCS using the VTK visualization back end.
June 2014	Release the beta version of the VCS/VTK software.
July 2014	Release the candidate and official release version of UV-CDAT with the new VCS/VTK software release.
August 2014	Merge DV3D and ParaView 3D capabilities with VCS 1D and 2D interactive capabilities.
September 2014	Release the alpha version of merged DV3D, ParaView, and VCS interactive capabilities.
October 2014	Release the beta version of merged DV3D, ParaView, and VCS interactive capabilities.
November 2015	Release candidate and official release version of UV-CDAT's merged DV3D, ParaView, and VCS interactive capabilities.

6.2.7 Unstructured Climate Data Visualization

The purpose of the new DV3D unstructured grid plotter is to enable scientists to interactively construct 3D visualizations of climate data in its raw (unprocessed) geometry. All of the previously developed DV3D plotters (and virtually all other climate data visualization tools) require the climate simulation data to be laid out on a rectangular grid. If the data is on an irregular or curvilinear grid then it must be resampled and interpolated to a regular array. However, this regridding process is undesirable in many cases because it is very time consuming; it can take a week or more for a large data set. In addition, the regridding process always distorts the data, which can warp and mask the features of interest. For example, computational artifacts at the tile boundaries of a cubed-sphere grid will most likely be hidden or distorted by the transformation to a projected coordinate system.

The new DV3D unstructured grid plotter allows irregular or curvilinear data to be viewed directly, without requiring any preprocessing. It makes no assumptions regarding the geometrical or topological layout of the points, but rather displays the points directly, with each point colored by the value of the variable at that location (**Figure 21**). By selectively varying the opacity of the points, the user can generate a wide range of slices, isosurfaces, and volume renderings. The same display process is applicable to data that is gridded (regular, curvilinear, or unstructured) or ungridded (e.g. level 2 observational data). These disparate data structures are all easily overlaid in a single view. It is also easy to shift the viewing geometry; for instance, points' geographic coordinates can be plotted in a flat projected coordinate system or a spherical global coordinate system. Multi-resolution, multi-core data processing and rendering algorithms facilitate responsiveness during user interactions. The user-friendly interaction controls enable users to drag slices across the data, probe for data values, control point rendering and geometry, run animations, and configure volume renderings, isosurfaces, and color maps (**Figure 22**). The plotter is designed to handle unprocessed NetCDF data from a wide range of climate simulation models. The initial version has been tightly integrated into UV-CDAT and tested with data from the CAM, GEOS5, ECMWF, WRF, and MMF models.

The next phase of development will build upon the current prototype DV3D unstructured grid plotter in response to feedback from the climate science user community. Potential focus areas include:

- Integration into the UV-CDAT web architecture;
- Support for other (as yet untested) simulation model and observational data structures;
- Enriching the visualization methods;
- Integration of an unstructured grid data processing toolkit;
- Promotion of the tool and solicitation feedback from a wider pool of climate scientists,
- Enhancement of the interaction and probe capabilities, and
- Improvement of the data processing performance and memory efficiency.

Additional development prospects are expected to emerge through interaction with the user community.

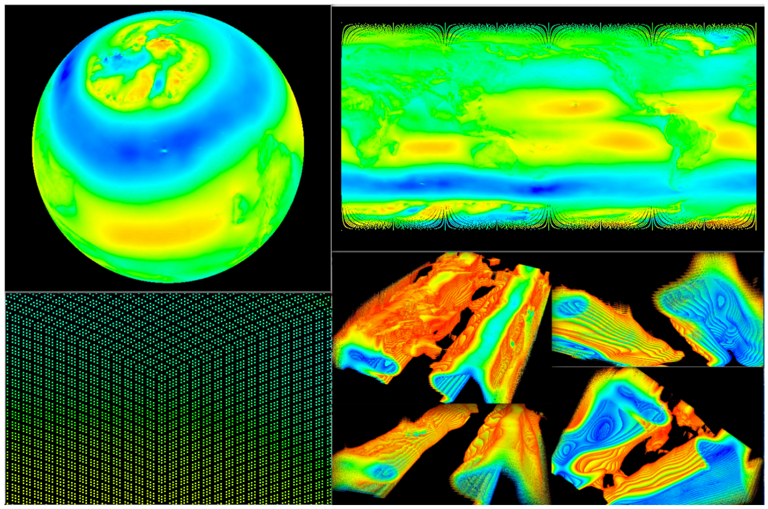


Figure 21. *DV3D unstructured grid plots of cubed-sphere CAM data. Clockwise from upper left: 1) data sliced along constant z-coordinate, in spherical global coordinate system, 2) data sliced along constant z-coordinate, in flat projected coordinate system, 3) data volumes generating by thresholding on a range of variable values, 4) data zoomed to reveal points.*

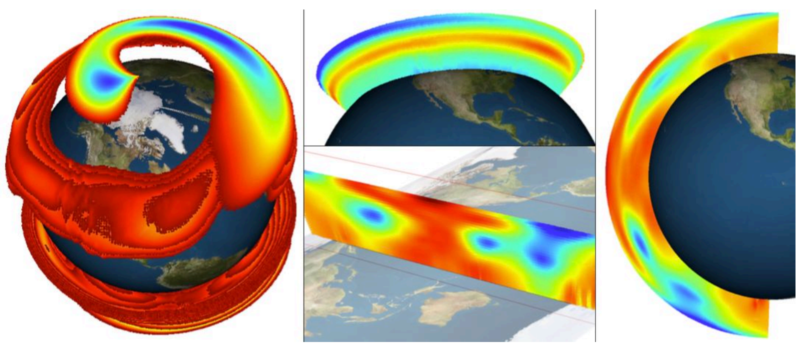


Figure 22. *DV3D unstructured grid plots of cubed-sphere CAM data. Left to right: 1) volume rendering generating by thresholding on a range of variable values in spherical coordinate system, 2) data sliced along constant latitude in spherical coordinate system (above) and data sliced along constant longitude in projected coordinate system (below), 3) data sliced along constant longitude in spherical coordinate system.*

Table 20. *Unstructured grid milestones for 2014.*

Expected Time of Achievement	Milestones
April–August 2014	Integrate 3D visualization with UV-CDAT Live.
June–September 2014	Implement new 3D visualizations for diagnostics and model metrics.
September–November 2014	Implement new 3D visualizations for observational data.

6.2.8 Tutorials and Documentation

Tutorials and training documentations are needed to ensure effective use of the UV-CDAT software application and to inform and cite services needed to support the many analysis and visualization workflow-processing techniques. This work is to include the necessary training for rudimentary analysis procedures as well as more intensive training on manipulation, diagnostics, model metrics, and complex 3D data viewing, to name only a few. It will be designed to cover the diverse needs of users in the community, from novices to old hand, and from scientists to non-scientists.

To expedite the documentation, we will first develop a number of three-minute tutorials that will focus on the most used features of UV-CDAT. These tutorials will demonstrate basic UV-CDAT operations. They will also cover the selection of data; 1D, 2D, and 3D visualization capabilities; workflows and provenance; diagnostic generation; useful climatological calculations (including departures, correlation, variance etc.); regridders; and animations.

Following the development of tutorials, we will prepare an informational document on how to use UV-CDAT. Although technical in nature, it will attempt to make the concepts involved understandable and applicable to the average user. Because of this, some aspects of the material will be simplified or omitted, for the sake of clarity. We will make more complex procedures available via three-minute tutorials at a later date. The tutorials and documentation will be found online under the UV-CDAT website: http://uv-cdat.org. Also found under the tutorials and documentation "Help" section of the website will be: the UV-CDAT help desk, FAQs, supported UV-CDAT mailing lists (such as the uvcdat-support mailing list), and other helpful websites and wiki sites (such as the http://esgf.org site) for UV-CDAT users.

Table 21. *UV-CDAT milestones for 2014 tutorials and documentation.*

Expected Time of Achievement	Milestones
March 2014	Begin video tutorial of basic features.
May 2014	Begin documentation based on videos. Start FAQ list for the web. Begin videos of advanced features, including "help."
August 2014	Begin documentation of advanced features.
September 2014	Complete web documentation of basic features and prepare report.
November 2014	Complete web documentation of advanced features in preparation for the face-to-face meeting. Create a hard-copy alternative to web documentation.

6.3 Community Improvements

6.3.1 International Climate Network Working Group: Large-Scale Data Transfer

In January 2014, ESGF will start a new working group called the ICNWG, as shown in **Figure 23**. This working group will be dedicated to helping all ESGF climate data sites set up, optimize, and/or troubleshoot their network infrastructures for international climate data transfers, so that petabytes of data can traverse international networks from end-to-end at the high performance levels required for large-scale data analysis.

In 2014, ICNWG will start with getting a handful of data centers set up with at least 500 MB/sec (4 Gbps) of data transfer throughput. These sites include LLNL (US), NCI/ANU (AU), BADC (UK), DKRZ (DE), and the Royal Netherlands Meteorological Institute (KNMI, NE). The 500-megabyte-per-second goal will allow these sites to transfer about 43 TB of data per day, or over 1.25 PB per month. Additionally, this project has a stretch goal of 1 GB/sec (or 8 Gbps) of data transfer throughput, which will allow for about 86 TB of data to be transferred per day, or over 2.5 PB/month. ICNWG's milestones and timeline to fully implement and achieve these goals are listed in **Table 21**.

To help guarantee the success of ICNWG, these milestones run in parallel with and will be supported throughout 2014 by the Enlighten Your Research Global (EYR-Global) international networking program award that ESGF received this last November 2013. This award will help get the working group started; the group is led by Dean N. Williams of Lawrence Livermore National Laboratory with the help of Eli Dart, a network engineer from ESnet.

Ideally, the EYR-Global project will help develop network-related best practices so that all ESGF sites will be able to achieve good network performance throughput that will continue to be implemented by ICNWG. Looking further out, by 2016 ICNWG will aim to achieve 2 GB/sec for data transfer throughput between at least five ESGF sites. ICNWG is critical for the future of the ESGF because petascale observation and simulation data sets will be reaching a critical point in the next few years. Petabytes of data will need to be transferred between sites and to end-users. Therefore network and storage infrastructures will need to be engineered in order to allow successful data transfers, thereby allowing the ESGF sites to achieve the data transfer performance required by the science and overall, serve the climate community better.

Table 22. *ICNWG milestones for the 2014–2016 for at least five ESGF sites.*

Expected Time of Achievement	Milestones
March 2014	Deploy 10G perfSONAR test server and 10G data server. Set up perfSONAR tests.
May 2014	Test file system for 10G data servers—target 500MB/sec. Achieve 500MB/sec (4Gbps) network test throughput between perfSONAR test servers.
August 2014	Achieve 500MB/sec (4Gbps) disk-to-disk transfers between data servers.
September 2014	Extra time for resolving issues.
November 2014	Deploy second 10G data server. Configure second 10G data server for striped Globus/GridFTP transfers with first 10G data server.
November 2014	Deploy second 10G data server. Configure second 10G data server for striped Globus/GridFTP transfers with first 10G data server.
March 2015	Test striped Globus/GridFTP transfers with one other center.
June 2015	Test striped Globus/GridFTP transfers with all centers.
August 2015	Demonstrate 1GB/sec (8Gbps) transfers between all centers.
Remainder of 2015	Extra time for schedule slip. Prepare for 2016 stretch goals.
June 2016	Demonstrate 2GB/sec between all centers that are capable (stretch goal).

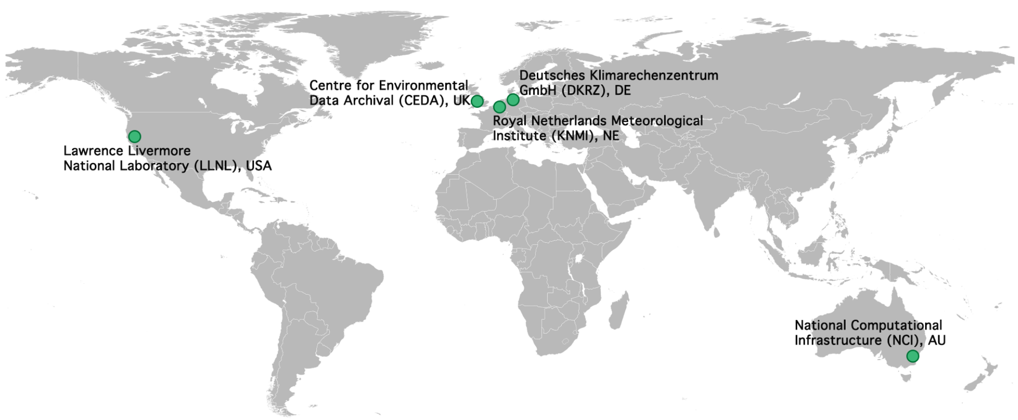


Figure 23. *ICNWG, in collaboration with the EYR-Global project, will set up, optimize, and/or troubleshoot 5 ESGF locations in different countries throughout 2014.*

6.3.2 Earth System Documentation

Earth System Documentation (ES-DOC) (http://es-doc.org) is an international project supplying tools, APIs, and consultancy in support of earth system documentation creation, analysis, and dissemination. It nurtures a sustainable, standards-based, documentation ecosystem that aims to be easily integrated into the next generation of exascale data set archives, for instance, the ESGF system. At the core of this ecosystem is a set of documentation tools built on top of the ES-DOC API. The tools are delivered as a set of embeddable web-browser plugins that support documentation search, viewing, comparison, and visualization. Documentation creation is supported via both a command line client and an online questionnaire.

The ESGF front-end currently embeds the ES-DOC CIM viewer into its search front end. As ESGF migrates to the CoG system, ES-DOC will ensure that its documentation tools continue to be seamlessly embeddable within the CoG front-end at appropriate integration points. It is thus envisaged that several ES-DOC tools will augment the CoG system. ES-DOC will collaborate with ESGF in the efforts to design, develop, and deploy a robust set of controlled vocabulary services. Such services will be leveraged by various components within the ESGF stack. The set of controlled vocabularies include earth system model component hierarchies for the various MIP's, as well as DRS terms, CF names, and the like.

ES-DOC will also exploit the faceted search capabilities of the ESGF search API in order to build graphical archive visualizations so as to help users better understand the archive from multiple viewpoints.

Table 23. *ES-DOC expected milestones for the coming year.*

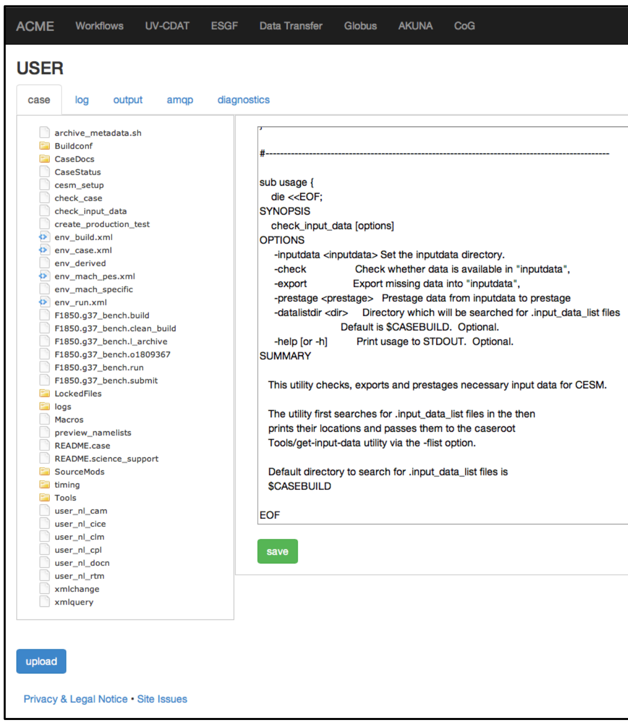
Expected Time of Achievement	Milestones
January 2014	Automate deployment of the ES-DOC stack.
February 2014	Initiate a controlled vocabulary services design.
March 2014	Develop an embeddable simulation comparator.
April 2014	Deliver the PDF documentation viewer.
May 2014	Build visualization data sets upon ESGF search facets.
June 2014	Support real-time simulation documentation publication.
July 2014	Support online model documentation creation.
August 2014	Deliver an embeddable archive visualizer.
September 2014	Supply controlled vocabulary hosting services.
October 2014	Deliver an embeddable simulation comparator.
November 2015	Full support of CoG system.

6.3.3 ACME End-To-End Test Bed

DOE's ACME end-to-end model test bed objective is to strongly accelerate the advanced model development, testing cycle, and execution for the new DOE Earth System Model by automating labor intensive tasks, providing intelligent support for complex tasks, and reducing duplication of effort through collaboration support. Thanks to full provenance capture throughout the end-to-end model development, diagnostic, and production cycle, it will enable unprecedented levels of reproducibility. The test bed environment will enable group and project collaboration with the data sharing and computing infrastructure. It will provide a testing-to-production environment for simulation and evaluation (with model metrics, diagnosis, and intercomparison). The integration with ESGF infrastructure will enable distributed archiving of high-volume simulation data, and the integration with UV-CDAT will provide thin as well as thick client interface to the data with sophisticated visualization and automated model diagnostic framework and rapid, interactive user analysis of in situ data. To achieve these goals, we will leverage and integrate existing software tools and develop necessary new software capabilities.

Thought the duration of the project, we will develop the overarching web front-end (ACME-FE, **Figures 24 and 25**) based on Python and Django. We will integrate it with Akuna, which will provide workflow execution, and metadata capture; the workflow includes all the processes needed to run and diagnose the model, create custom analyses, and publish the data to ESGF. Akuna will be integrated with ProvEn, which will be used to capture full provenance of a workflow.

We will leverage Globus Nexus for account and group creation, and authentication for Globus Online use as well as Akuna and ACME-FE. Globus Online technology will be used for file transfer needs. The ACME-FE will be integrated with ESGF, where it will enable ESGF project search, data search, and simple and automated data publication to ESGF. It will also be integrated with UV-CDAT Live and web informatics and exploratory analysis web interfaces to meet visualization, diagnostics, model metrics, UQ, and user analysis needs.



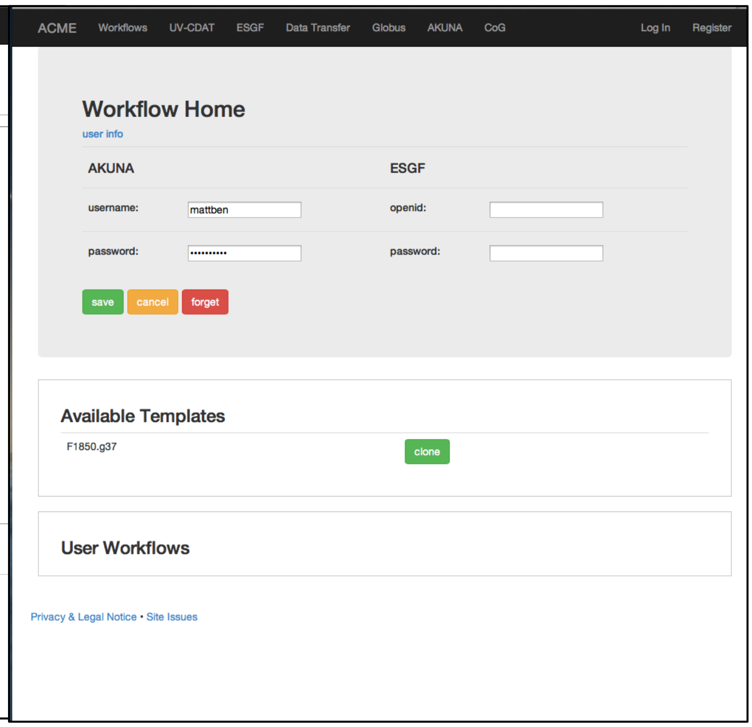


Figure 24. *(left) ACME-FE prototype UI showing model run cases.* **Figure 25.** *(right) ACME-FE prototype UI showing workflow login front page.*

Table 24. *ACME test bed expected milestones for the coming year.*

Expected Time of Achievement	Milestones
January 2014	Complete the design of the overall UI framework.
February 2014	Explore collaborative open source frameworks to help with the development of the ACME-FE.
March 2014	Develop a specialized collaborative test bed portal with a checklist for approvers of new model setups before job launching. Demonstrate the prototype UI at the ACME review March meeting.
April, May 2014	Support metadata creation, annotation, and forums for discussions and sharing of any part of a workflow and capturing of this metadata in ProvEn.
June 2014	Beta test the ACME test bed and support computational infrastructure

Expected Time of Achievement	Milestones
	for several HPC centers.
July 2014	Release production test bed for ACME users.
August 2014	Refine test bed to add additional features needed for v.0 fully coupled system.
September 2014	Document and beta test additional features.
October, November 2014	Deliver the next release version of the test bed in preparation of the v.1 fully coupled system.

6.3.4 OpenClimateGIS

OpenClimateGIS (OCGIS) [OpenGIS, 2013] is a NOAA-funded open source Python software library designed for geospatial manipulation, subsetting, computation, and translation of climate data sets stored in local NetCDF-CF files or files served remotely via OPeNDAP. The software is optimized to work with high-dimensional climate data sets allowing a number of common GIS operations (e.g. intersects, clip, spatial averaging, coordinate system transformations) to be performed on a file storage format notoriously difficult to manipulate in modern GIS software. OCGIS shares a number of structural similarities to UV-CDAT, most importantly the common Python programming language and NumPy-based array storage. These similarities will simplify the integration of OCGIS's geospatial functionality with the UV-CDAT library. Incorporating GIS into UV-CDAT will provide it with a set of unique analysis capabilities for localized climate analysis.

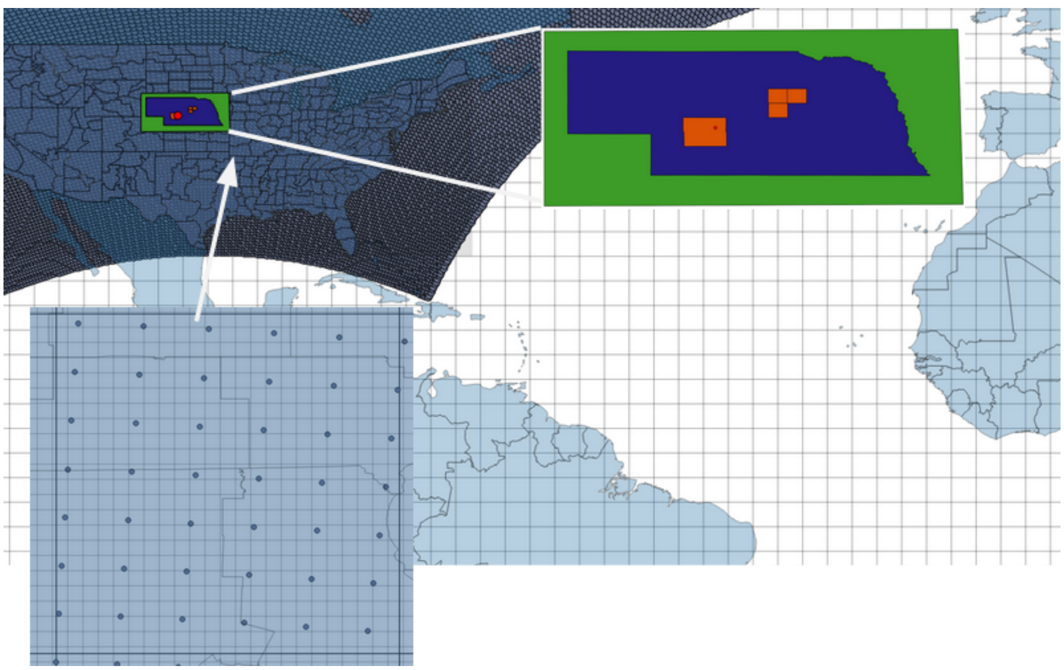


Figure 26. *This graphic displays ESRI shapefile output from OpenClimateGIS (OCGIS) using three different climate products as input. The three climate products (NARCCAP, CMIP5 Decadal Simulation, Maurer Observational) have differing coordinate systems that were consolidated by OCGIS. The inset map displays different approaches to spatial subsetting available inside OCGIS: point, bounding box, and collections of arbitrary geometric boundaries.*

There are two major development stages for OCGIS integration into UV-CDAT. These stages are described in more detail below.

1. Introducing ESRI shapefile as a data set objects inside UV-CDAT.

2. Enabling the spatial subsetting operations (i.e. intersects, clip), coordinate transformations (i.e. remapping), and spatial averaging as options inside UV-CDAT.

The first stage will bring support for the vector-based, ESRI Shapefile storage format as a UV-CDAT data set object. This new data set object will follow the UV-CDAT interface to the extent possible. One obstacle will be the handling of multi-geometries or geometry collections (e.g. a multipolygon like the Hawaii state boundary) but that may be addressed by introducing a method for splitting those geometry types into single geometries.

The second stage will produce a series of UV-CDAT GIS operations to use the new data set objects for spatial subsetting and averaging of grid-based data set objects in UV-CDAT. It is expected that ESMP will be used for the conservative regridding/spatial averaging at this stage. There will also be the ability to perform coordinate transformations (i.e. remapping) on target data sets. Implementation at this phase is not as clear in terms of how the operations will be exposed in the Python code and UV-CDAT user interface. However, we are targeting an implementation as easy as possible at the user level:

```
>>> f=cdms2.open(shapefile)
>>> data_s = f("somevar",latitude=(lat1,lat2),longitude=(lon1,lon2))
>>> f=cdms2.open(model_output_file)
>>> data_m = f("somevar",latitude=(lat1,lat2),longitude=(lon1,lon2))
>>> data_s2 = data_s.regrid(data_m.getGrid())
```

As an illustrative example—following completion of the second development stage—a UV-CDAT user will be able to load in a U.S. state boundaries shapefile, select a state, and spatially subset other loaded data sets by that state's boundary.

The milestone dates below are considered tentative and may change following OCGIS team prioritization or changes to the software integration strategy.

Table 25. *OpenClimateGIS expected achievable milestones for the coming year.*

Expected Time of Achievement	Milestones
March 2014	Bring OpenClimateGIS into the UV-CDAT build system.
April 2014	Prototype the data interface to the ESRI Shapefile format in UV-CDAT. Identify modifications to OCGIS data interface required to simplify access to the ESRI Shapefile. Begin demonstration to the UV-CDAT developers of the prototype data interface to ESRI Shapefile format.
May 2014	Finalize the ESRI Shapefile data interface in UV-CDAT.
June 2014	Develop an implementation plan for bringing OCGIS GIS operations into UV-CDAT.
July 2014	Implement the plan to bring OCGIS GIS operations into UV-CDAT.
August 2014	Target for initial implementation of GIS operations in UV-CDAT. Alpha release UV-CDAT with GIS operations.
September 2014	Begin demonstrations and feedback from UV-CDAT users (UV-CDAT beta release).
November 2015	Perform initial merge of OCGIS GIS functionality into the UV-CDAT trunk following the Stage 2 development plan above.

7 Planned Development and Integration

The ESGF and UV-CDAT teams have been designing, developing, implementing, fielding, supporting, maintaining, and facilitating access and analysis and visualization to simulation and observational data sets for well over a decade. Our clients have hailed our projects as a "game changer." The American Meteorological Society (AMS) in 2010 writes, "The CMIP3 multi-model dataset archive...led to a new era in climate system analysis and understanding," and in 2013 the U.S. Federal Laboratory Consortium awarded the ESGF project for "Outstanding Partnership." As the climate community processes and needs have changed over the years, our flexible software architecture has adapted to meet their changing requirements. This section summarizes planned development and integration to the ESGF and UV-CDAT software needed to meet future requirements for continued scientific productivity and discovery.

7.1 Software Development Capabilities

Since the mid-1990s, our software development teams have continuously delivered effective and intuitive automated solutions to problems of all sizes and complexities on time to the climate community, especially for the CMIP large-scale data efforts. Our expert developers thrive on learning and implementing new technologies and techniques to bring the most user friendly and intelligent software directly to our community of users.

Our established coding standards, development processes and best practices help ensure the integrity and usability of our many community climate change products, such as ESGF, UV-CDAT, ES-DOC, CoG, CF, CVs, Globus, and many others. In addition, our software development lifecycle encourages customer involvement throughout. Our Agile development process allows us to adjust our schedules and priorities as necessary to quickly provide new solutions and meet our customers' ever-changing needs. For the most part, our goal for this workshop was to integrate the many software components of the climate community together and map out when we think they will be available for use by the user community.

With that, our software design teams provided detailed designs in their presentations including high-level architecture diagrams, in-depth data and class definitions, and user interface mock-ups at the face-to-face meeting. A rigorous testing process will help to ensure a quality product, while the requirements and design documentation guarantees that we meet the climate customers' needs now and into the future.

7.2 Integrated Data Environment Capabilities

As our tools continue to evolve and enter new areas of software develop for the climate change community, we are constantly seeking new ways to integrate with a long list of tools that already exist and doing their jobs well in the community. In addition, we are also trying to upgrade components of these tools and parallelize them for greater efficiency and speed. The AMWG, LMWG, and OMWG diagnostics packages are examples of upgrades that we've made to better the user experience in diagnosing model subcomponents. As always we make it a point to leverage the input and output capabilities, through APIs of any existing tools to provide the most effective solutions.

As shown throughout the report, the teams have focused their efforts to expand into new areas of data mining, parallelism, workflows, HPC, cluster computing, and analysis over an integrated distributed/federated network. As discussed in previous reports, we intend to fully explore the possibility of providing a configurable and scalable cloud-based environment, so that resources and services can be instantiated on demand for specific short-lifetime projects or to meet such requirements as elastic allocation of computing processes. From the meeting, some of the integrated data activities include:

- User and developer support: Provide new documentation, tutorials, and website updates, maintenance, support for platforms, monitoring of system tools usage, etc. Engage with the community for requirements, adoption, and feedback.

- Community outreach: Engage with the climate-science community through workshops, governance, literature, etc.
- CoG: New ESGF front-end that will help users keep track of the many multiple projects that are supported by ESGF. This will also help users keep track of facet customization on a per-project basis; upload files of any kind to a project workspace; allow user profiles; and allow better integration with other Python/Django software such as UV-CDAT, RCMES, and OCG.
- Diagnostic packages: Develop AMWG, LMWG, OMWG, PCWG, and MPAS diagnostic packages along with others to be determined.
- Big data/parallel data processing architecture: Provide three parallel processes that have been identified for running batch and interaction parallel processes along with client-server parallel interaction.
- Web and information visualization: Advance UV-CDAT analysis and visualization capabilities to a web application.
- Software process: Continue to build upon and improve UV-CDAT's build, test, and documentation processes.
- Refactor code to make it easier to add functionality such as Climate Data Operators and NetCDF Operators: Simplify the extendable and pluggable feature to support the addition of new codes.
- Interface enhancements: Partner with the climate science community to improve the user experience.
- Improve scripting: Improve workflow scripts and user-controlled parameters.
- Improve visualization: Develop new GIS, 1D, 2D, and 3D visualization and interface for the climate community.
- Improve notification: The notification software will be upgraded to allow the user to select the type of notification response (i.e., e-mail or RSS Feed) from their profile. When the ESGF node manager changes data, information indicating the change will be automatically pushed to each individual user that request to be notified. The data change will also be push to the CoG's projects errata site for logging the list of errors and corrections.

Our software expertise extends beyond our tools indicated in this report. With a vast knowledge of and experience in tools that help to federate community resources, we will continue to find the best tools to meet our customers' needs, and we help manage, maintain and ensure the integration of data throughout.

7.3 Products

7.3.1 CoG

CoG is a web application that integrates services for data and metadata management, with tools for scientific collaboration, and formal elements of projects governance. For example, CoG provides a powerful user interface to search and display results from a back-end search engine conforming to the ESGF API; it allows users to store and inspect search records in a persistent Data Cart, analogous to the shopping cart metaphor that is standard in commercial web sites; it allows scientists to easily request a project workspace that they can share with colleagues to document and reference their scientific findings; and it allows PIs to formalize the governance of their projects by documenting their structure, their bodies, their processes, repositories, and other aspects of the project. Additionally, as part of its ESGF adoption by ESGF, CoG is being evolved to support a multi-site distributed architecture, where scientific projects can be instantiated at any site, and shared seamlessly across the federation.

Features:

- Support of multiple projects within same web portal

- Connections among projects (parent, children, peers), including other MIPs
- Facet customization on a per-project basis
- Customizable navigation bar for each project
- Templated content to describe projects
- Customizable left menu for each project
- Embedded wiki capabilities—users and administrators can edit/share the site content (this also means that the CMIP5 documentation could be co-located with the CMIP5 data)
- Capability to upload files of any kind (documents, presentation, images) to a project workspace
- Allow users to comment on any page
- Better support for delivering and viewing images after a search
- Integration with social media (Facebook, Google+, Twitter)
- User's profile (interests, photos, projects, etc.)
- Ability to send news to current and associated projects
- Ability to add resources to a project as shared bookmarks
- Ability to tag projects with custom tags, and to search for projects by tag

7.3.2 ESGF Dashboard

The ESGF Dashboard provides a top-level view of the entire distributed ESGF infrastructure. The Dashboard collects historical information about the status of the ESGF federation and delivers high-level, real-time metrics. Reports and graphs to be filtered or sorted in virtually any manner so that users can obtain data when and how they need it.

From an architectural viewpoint, the ESGF Dashboard is composed of the following three parts: the information provider, the dashboard catalog, and the user interface. The *information provider* represents the system's back-end and is responsible for retrieving and storing the peer-to-peer metrics. The *dashboard catalog* is a temporal database used by the information provider to persistently store the metrics. Finally, the *user interface* is a web application that exploits the MVC design pattern. It relies on a strong adoption and implementation of Web 2.0 concepts such as mash-up, Google maps, and permalinks. It provides several views at four different (hierarchical) granularity levels: federation, peer-group, host, and service. These views provide charts and summaries about the service availability, the round trip time between pairs of nodes, the host status, and other information. The views, which can be customized (at the presentation level) by the end user, are delivered in different ways and formats. From an implementation point of view the Dashboard GUI (ESGF Desktop) is written in JS. The ESGF Desktop is an extensible, flexible and pervasive environment including multiple applications such as a web terminal, monitoring tools, download statistics applications, and community-based gadgets.

One of the next big integrations involving the ESGF Dashboard will include the network information from the major sites: LLNL, BADC, DKRZ, ANU/NCI, and KNMI.

Features:

- Across-the-organization look at planning and execution
- Filterable, sortable, printable reports
- Network topology (peer-groups composition)
- Node type (host/services mapping)

- Registered users (extracting the information from the ESGF-security user table)
- Downloaded data (both at the node and federation level)
- System metrics (e.g., round-trip time, service availability, CPU, memory, disk, processes, etc.)

7.3.3 UV-CDAT Live

The UV-CDAT "smart client," named UV-CDAT Live, is an AJAX-based, cross-browser approach that will give users the benefits of UV-CDAT's rich "thick client" look and feel without the need of having to download and install onerous prerequisite software and the UV-CDAT software stack. This is truly a "rich Internet application" that is integrated with ACME tools, such as model diagnostics and metrics. It can be run as a stand-alone rich web application or within the ACME end-to-end test bed for complete model input and output control and diagnostics.

Over time, the UV-CDAT "smart client" approach overtime will replace the old UV-CDAT desktop application. We are expecting the web version of UV-CDAT to have local and remote access to data and provide the same interactive responsiveness to hot keys as the "thick client." In support of increased user productivity, the live version of the UV-CDAT will also include workflow and province capture for reproducibility and resource sharing.

Features:

- Delivers the UV-CDAT application and tools over a secure web Hypertext Transfer Protocol (HTTPS) connection
- No need to install UV-CDAT software (except for browser) or provide automated installation and updates
- Automatically updates UV-CDAT without user action or concerns
- Has the look and feel of UV-CDAT desktop applications and interactions

7.3.4 ACME End-to-End UI

The advanced model development, testing, and execution infrastructure has been designed to strongly accelerate the model development and testing cycle for the new DOE Earth system model, known as ACME, by automating labor-intensive tasks, providing intelligent support for complex tasks, and reducing duplication of effort through collaboration support. The test-bed environment will provide the group of collaborating DOE scientists with the data and computing infrastructure needed for rapid development and assessment of new scientific modules and provide a testing-to-production environment for simulation and evaluation (i.e., model metrics, diagnosis, and intercomparison). Deployment and integration of existing software tools, as well as the development of necessary new software capabilities to accomplish this, will be driven by the scientific requirements to develop and use the overall coupled earth system model and the individual component models (atmosphere, land, ocean, sea ice, and land ice) within it.

To achieve the goals of the ACME project, the team will build upon the efforts of the Climate Science for a Sustainable Energy Future, ESGF, UV-CDAT, GO, and the Advanced Simulation Capability for Environmental Management-developed Akuna framework.

Features:

- Creates a flexible, extensible infrastructure for future ACME efforts and related DOE projects
- Automates laborious, repetitive simulation data tasks to improve productivity
- Supports broad data sharing within ACME project teams and with scientific collaborations
- Accelerates model development through visualization, diagnostics, and testing
- Enables reproducibility and archiving of high-volume simulation data

- Specialized collaborative portal with a checklist for approvers of new model setups before job launching
- Supports metadata creation, annotation, and forums for group discussions and sharing of any part of a workflow and provenance metadata capture
- Supports computational infrastructure for several HPC centers
- Specialized software when needed to enable web job submission, running, monitoring, and debugging capabilities on several HPC centers
- Job launching interface for submission of hundreds of production runs and enabling specialized monitoring of multiple ensemble, automated runs
- Rule based controls for model set-up
- User web help center with online tutorials, FAQs, and a help desk

8 Glossary

Acronym	Meaning and Website
ACME	Accelerated Climate Modeling for Energy: DOE's effort to build an Earth system modeling capability tailored to meet the climate change research strategic objectives
AGU	American Geophysical Union (http://sites.agu.org/)
AIMS	Analytics and Informatics Management Systems project responsible for all LLNL's climate software, including ESGF, UV-CDAT, and the DOE ACME Test Bed (http://aims.llnl.gov/)
Akuna	Advanced Simulation Capability for Environmental Management (ASCEM) developed AKUNA framework (http://akuna.labworks.org/)
AMS	American Meteorological Society (https://www.ametsoc.org/)
AMWG	Atmosphere Model Working Group
ana4MIPs	A pilot activity to make reanalysis products more accessible for climate model intercomparisons (http://www.wcrp-climate.org/documents/ezine/WCRPnews_17082012.pdf)
ANL	Argonne National Laboratory, sponsored by the DOE (http://www.anl.gov/)
ANU/NCI	Australian National University/National Computational Infrastructure (http://www.nci.org.au)
API	Application Programming Interface (http://en.wikipedia.org/wiki/Application_programming_interface)
AR5	Fifth IPCC Assessment Report, published in 2013 (http://www.ipcc.ch/report/ar5/#.UwVGOCTm6Gg)
ARM	Atmospheric Radiation Measurement is a U.S. Department of Energy scientific user facility, providing data from strategically located in situ and remote sensing observatories around the world (http://www.arm.gov/)

Acronym	Meaning and Website
BADC	British Atmospheric Data Centre (http://badc.nerc.ac.uk/)
BER	Office of Biological and Environmental Research under the DOE Office of Science (http://science.energy.gov/ber/)
CA	Certificate Authority is an entity that issues digital certificates (http://en.wikipedia.org/wiki/Certificate_authority)
CF	Climate and Forecast metadata convention, for processing and sharing NetCDF data files (http://cf-pcmdi.llnl.gov/)
CFSR	NOAA's Climate Forecast System Reanalysis (http://cfs.ncep.noaa.gov/cfsr/)
Client-Server	Relationship between two computer programs, where the client program makes a service request, which the server program fulfills (http://en.wikipedia.org/wiki/Client-server)
CMCC	Euro-Mediterranean Center on Climate Change (http://www.cmcc.it/)
CMIP5	Coupled Model Intercomparison Project 5, sponsored by WCRP/WGCM, and related multi-model database planned for the IPCC AR5 (http://cmip-pcmdi.llnl.gov)
CoG	A web environment that enables users to create project workspaces, connect projects into networks, share and consolidate information within those networks, and seamlessly link to tools for data archival, reformatting and search, data visualization, and metadata collection and display (https://earthsystemcog.org/)
CORDEX	Coordinated Regional Climate Downscaling Experiment, providing global coordination of Regional Climate Downscaling for improved regional climate change adaptation and impact assessment (http://wcrp-cordex.ipsl.jussieu.fr/)
CVs	Controlled Vocabularies.
Data Node	Internet location providing data access or processing (http://en.wikipedia.org/wiki/Node-to-node_data_transfer)
DKRZ	German Climate Computing Centre (http://www.dkrz.de/)
DOE	Department of Energy, the U.S. government entity chiefly responsible for implementing energy policy (http://www.doe.gov/)
DV3D	3D Climate data visualization using python and VTK (http://portal.nccs.nasa.gov/DV3D/)
EDEN	Exploratory Data analysis ENvironment is a visual analytics tool for exploring multivariate data sets. (http://cda.ornl.gov/projects/eden/)
EOS	NASA's Earth Observing System (http://eospso.gsfc.nasa.gov/)
ES-DOC	Earth System Documentation, an international effort to develop metadata services for a set of climate and related projects (http://earthsystemcog.org/projects/es-doc-models/)
ESGF	Earth System Grid Federation, led by LLNL, a worldwide federation of climate and computer scientists deploying a distributed multi-petabyte archive for climate science (http://esgf.org)

Acronym	Meaning and Website
ESMF	Earth System Modeling Framework is software for building and coupling weather, climate, and related models (http://www.earthsystemmodeling.org/)
ESnet	DOE Energy Science Network (https://www.es.net/)
ESRL	NOAA Earth System Research Laboratory (http://www.esrl.noaa.gov/)
Exascale	Computer processing capabilities of order 10^{18} operations per second (http://en.wikipedia.org/wiki/Bit)
EYR-Global	Enlighten Your Research-Global
F2F	Face-to-Face: being in the presence of another (http://www.thefreedictionary.com/face-to-face)
FAQs	Frequently Asked Questions
FUNET	Finnish university and research network (http://www.csc.fi/hallinto/funet)
Gbps	Gigabit per second, 10^9 bits of information (http://en.wikipedia.org/wiki/Data_rate_units)
GB	Gigabyte, 10^9 bytes of information (http://en.wikipedia.org/wiki/Gigabyte)
GeoMIP	Geo-engineering Model Intercomparison Project (http://climate.envsci.rutgers.edu/GeoMIP/)
GO	Globus Online is an open-source software toolkit used for building grids (https://www.globus.org/)
GridFTP	Is an extension of the standard File Transfer Protocol for high-speed, reliable, and secure data transfer (http://toolkit.globus.org/toolkit/docs/latest-stable/gridftp/)
HPC	High-Performance Computing (http://en.wikipedia.org/wiki/Supercomputer)
ICNWG	International Climate Network Working Group, formed under the Earth System Grid Federation (ESGF), is to help set up and optimize network infrastructure for their climate data sites located around the world (http://icnwg.llnl.gov/)
INTERNET2	U.S. research and education network (http://www.internet2.edu/)
IPCC	Intergovernmental Panel on Climate Change, a scientific body of the United Nations, periodically issues assessment reports on climate change (http://www.ipcc.ch/)
IPSL	Institut Pierre-Simon Laplace (http://www.ipsl.fr/)
IS-ENES2	Infrastructure for the European Network of Earth System Modelling 2 (https://verc.enes.org/ISENES2/)
JANET	U.K. research and education network (https://www.ja.net/)
JPL	NASA Jet Propulsion Laboratory (http://www.jpl.nasa.gov/)
KNMI	Royal Netherlands Meteorological Institute (http://www.knmi.nl/index_en.html)

Acronym	Meaning and Website
LANL	Los Alamos National Laboratory, sponsored by the DOE (http://www.lanl.gov/)
LBNL	Lawrence Berkeley National Laboratory, sponsored by the DOE (http://www.lbl.gov/)
LLNL	Lawrence Livermore National Laboratory, sponsored by the DOE (https://www.llnl.gov/)
LMWG	Land Model Working Group
LUCID	Local Urban Climate Model and its Application to the Intelligent Design of Cites in the U.K. (http://www.homepages.ucl.ac.uk/~ucftiha/index.html)
MERRA	NASA's Modern Era Retrospective-analysis for Research and Applications (http://gmao.gsfc.nasa.gov/merra/)
Metadata	Data properties, such as their origins, spatio-temporal extent, and format (http://en.wikipedia.org/wiki/Metadata)
MPAS	Model for Prediction Across Scales
NARCCAP	North American Regional Climate Change Assessment Program (http://www.narccap.ucar.edu/)
NASA	National Aeronautics and Space Administration (http://www.nasa.gov/)
NCAR	National Center for Atmospheric Research, sponsored by the National Science Foundation (http://www.ncar.ucar.edu/)
NCI	National Computational Infrastructure at the Australian National University (http://www.nci.org.au/)
NCCS	NASA Center for Climate Simulation (http://www.nccs.nasa.gov/)
NetCDF	A machine-independent, self-describing, binary data format (http://www.unidata.ucar.edu/software/netcdf/)
NOAA	National Oceanic Atmospheric Administration, an agency of the U.S. Commerce Department (http://www.noaa.gov/)
NSF	National Science Foundation, an agency of the U.S. (http://www.nsf.gov/)
NYU	New York University (http://www.nyu.edu/)
obs4MIPs	Observation for Model Intercomparison Project, a pilot activity to make observational products more accessible for climate model intercomparisons (http://obs4mips.llnl.gov:8080/wiki/)
OCGIS	OpenClimateGIS is a Python package designed for geospatial manipulation, subsetting, computation, and translation of climate datasets stored in local NetCDF files or files served through THREDDS data servers (https://earthsystemcog.org/projects/openclimategis/)
OMWG	Ocean Model Working Group

Acronym	Meaning and Website
OPeNDAP	Open-source Project for Network Data Access Protocol is a data transport architecture and protocol widely used by Earth scientists (http://www.opendap.org/)
OpenID	Allows users to use an existing account to sign in to multiple websites, without needing to create new passwords (http://openid.net/)
ORNL	Oak Ridge National Laboratory, sponsored by the DOE (http://www.ornl.gov/)
OSG	Open Science Grid
ParaView	An open-source, multi-platform data analysis and visualization application. (http://www.paraview.org/)
PB	Petabyte, 10^{15} bytes of information (http://en.wikipedia.org/wiki/Petabyte)
PCMDI	Program for Climate Model Diagnosis and Intercomparison, located at LLNL (http://www-pcmdi.llnl.gov/)
PKI	Public Key Infrastructure is a set of hardware, software, people, policies, and procedures needed to create, manage, distribute, use, store, and revoke digital certificates (http://en.wikipedia.org/wiki/Public-key_infrastructure)
SURFnet	Netherlands network (http://www.surf.nl/)
TAMIP	Transpose Atmospheric Model Intercomparison project (http://www.metoffice.gov.uk/hadobs/tamip/)
TB	Terabyte, 10^{12} (a trillion) storage bytes (http://en.wikipedia.org/wiki/Terabyte)
THREDDS	Thematic Real-time Environmental Distributed Data Services (https://www.unidata.ucar.edu/software/thredds/current/tds/)
UV-CDAT	Ultrascale Visualization Climate Data Analysis Tools, provides access to large-scale data analysis and visualization tools for the climate modeling and observational communities (http://uv-cdat.org)
VCS	Visualization and Control System (http://uv-cdat.org)
VisTrails	An open-source system that supports data exploration and visualization (http://www.vistrails.org/)
VO	Virtual Organization is one whose members are geographically apart, usually working by computer e-mail and groupware while appearing to others to be a single, unified organization with a real physical location (http://en.wikipedia.org/wiki/Virtual_organization)
VTK	Visualization ToolKit, an open-source, freely available software system for 3D computer graphics, image processing, and visualization (http://www.vtk.org/)
WCRP	World Climate Research Programme, which aims to facilitate analysis and prediction of Earth system variability and change for use in an increasing range of practical applications of direct relevance, benefit, and value to society (http://www.wcrp-climate.org/)
Web portal	A point of access to information on the World Wide Web

Acronym	Meaning and Website
	(http://en.wikipedia.org/wiki/Web_portal)
Wget	The non-interactive network downloader is describe as a computer program that retrieves content from web servers and is part of the GNU Project (https://www.gnu.org/software/wget/)
Workflow	A sequence of operations, performed by person(s), organization(s), or mechanism(s) (http://en.wikipedia.org/wiki/Workflow)

9 References

1. [Cinquini, 2013] Luca Cinquini, Daniel Crichton, Chris Mattmann, Gavin M. Bell, Bob Drach, Dean Williams, John Harney, Galen Shipman, Feiyi Wang, Philip Kershaw, Stephen Pascoe, Rachana Ananthakrishnan, Neill Miller, Estanislao Gonzalez, Sebastian Denvil, Mark Morgan, Sandro Fiore, Zed Pobre, Roland Schweitzer, "The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geospatial Data", IEEE Future Generation Computer Systems, http://dx.doi.org/10.1016/j.future.2013.07.002, 17 September. 2013.

2. [OpenGIS, 2013] OpenClimate GIS Website Overview, 2013. https://www.earthsystemcog.org/projects/openclimategis/

3. [Steed, 2013] Chad A. Steed, Daniel M. Ricciuto, Galen Shipman, Brian Smith, Peter E. Thornton, Dali Wang, Xiaoying Shi, Dean N. Williams, Big data visual analytics for exploratory earth system simulation analysis, Computers & Geosciences, Volume 61, December 2013, Pages 71-82, ISSN 0098-3004, http://dx.doi.org/10.1016/j.cageo.2013.07.025.

4. [Williams, 2013] Dean N. Williams, Timo Bremer, Charles Doutriaux, John Patchett, Sean Williams, Galen Shipman, Ross Miller, David R. Pugmire, Brian Smith, Chad Steed, E. Wes Bethel, Hank Childs, Harinarayan Krishnan, Prabhat, Michael Wehner, Claudio T. Silva, Emanuele Santos, David Koop, Tommy Ellqvist, Jorge Poco, Berk Geveci, Aashish Chaudhary, Andy Bauer, Alexander Pletzer, David Kindig, Gerald L. Potter, Thomas P. Maxwell, "Ultrascale Visualization of Climate Data", IEEE *Computer Magazine*, September 2013, vol. 46 no 9, pp. 68-76, DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/MC.2013.119.

10 Participants and Contributors to the 2013 Report and Meeting

Figure 27. *Group photo of the 2013 international meeting attendees.*

10.1 Attendees

Last Name	First Name	Affiliation
Aloisio	Giovanni	Euro Mediterranean Center on Climate Change (CMCC)
Ananthakrishnan	Rachana	Argonne National Laboratory
Ayachit	Utkarsh	Kitware, Inc.
Bader	David	DOE/LLNL
Bang	Jae Young	University of Southern California
Bell	Gavin M.	DOE/LLNL
Berger	Katharina	Deutsches Klimarechenzentrum GmbH
Blau	Eric	Argonne National Laboratory
Bremer	Peer-Timo	DOE/LLNL
Budai	Jeff	NOAA/NCDC (STG, Inc.)
Burnett	Benjamin	NYU-Poly
Canada	Curits	DOE/LANL
Carenton-Madiec	Nicolas	Institut Pierre Simon Laplace (IPSL)
Chaudhary	Aashish	Kitware, Inc.
Cholia	Shreyas	DOE/LBNL/NERSC
Chunpir	Hashim	DKRZ
Cinquini	Luca	NASA/JPL and NOAA/ESRL
Cofino	Antonio S.	University of Cantabria
Crichton	Daniel	NASA/JPL
Dart	Eli	DOE/ESnet
DeLuca	Cecelia	NOAA/ESRL
Denvil	Sébastien	CNRS/IPSL
Doutriaux	Charles	DOE/LLNL
Dwarakanath	Prashanth	National Supercomputer Centre
Ehbrecht	Carsten	DKRZ
Evans	Ben	NCI/Australian National University
Fillmore	David	Tech-X Corporation
Fiore	Sandro	Euro Mediterranean Center on Climate Change (CMCC)

Last Name	First Name	Affiliation
Geernaert	Gary	DOE BER Headquarters
Gleckler	Peter	DOE/LLNL
Greenslade	Mark	Institut Pierre Simon Laplace (IPSL)
Hankin	Steve	NOAA / PMEL
Hardy	Carla	DOE/LLNL
Harney	John	DOE/ORNL
Harris	Matthew	DOE/LLNL
Hester	Mary	DOE/ESnet
Hnilo	Jay	DOE BER Headquarters
Juckes	Martin	STFC
Kershaw	Philip	NCAS/British Atmospheric Data Centre
Kindermann	Stephan	German Climate Computing Center, DKRZ
Koch	Dorothy	DOE BER Headquarters
Koop	David	NYU-Poly
Krassovski	Misha	CDIAC, Oak Ridge National Laboratory
Lee	Huikyo	NASA/JPL
Lee	Tsengdar	NASA Headquarters
Leung	Elo	DOE/LLNL
Manross	Kevin	NSF/NCAR/CISL
Mattmann	Chris	NASA/JPL
Maxwell	Thomas	NASA/Goddard
McCoy	Renata	DOE/LLNL
Mendes	Wanderley	INPE
Nienhouse	Eric	NSF/NCAR
Ogochi	Koji	Japan Agency for Marine-Earth Science and Technology (JAMSTEC)
Painter	Jeffrey	DOE/LLNL
Pascoe	Stephen	Science and Technology Facilities Council, UK
Peterschmitt	Jean-Yves	LSCE/IPSL
Pletzer	Alexander	Tech-X Corp.
Potter	Gerald	NASA/GSFC
Qizhong	Wu	Beijing Normal University
Radhakrishna	Aparna	NOAA/GFDL
Schuster	Douglas	NSF/NCAR
Sim	Alex	Lawrence Berkeley National Laboratory
Silva	Claudio	NYU
Smith	Brian	DOE/ORNL
Stephan	Eric	Pacific Northwest National Laboratory
Stevens	Laura	CICS-NC/NCDC
Stockhause	Martina	DKRZ
Taylor	Karl	DOE/LLNL/PCMDI
Williams	Dean	DOE/LLNL
Yoo	Wucherl	DOE/LBNL
Zhang	Bao	Beijing Normal University

10.2 Meeting and Report Organizer

Dean N. Williams, LLNL

10.3 Joint International Agency Meeting Committee

Dean N. Williams, DOE/LLNL, ESGF, UV-CDAT

Ian Foster, DOE/ANL, Globus, ESGF

Chris Mattmann, Dan Crichton and Luca Cinquini, NASA/JPL, ESGF, CMAC

Laura Carriere, NASA/GSFC/NCSS, CDS, ESGF

Cecelia DeLuca, NOAA/ESRL, CoG

Stephen Pascoe, IS-ENES2/BADC, ESGF

Sébastien Denvil, IS-ENES2/IPSL, ESGF

Ben Evans, ANU/NCI, ESGF

Martin Juckes, IS-ENES2/BADC, ESGF

Michael Lautenschlager, IS-ENES2/DKRZ, ESGF

Cecelia DeLuca, NOAA/ESRL, CoG

Eric Guilyardi, IS-ENES2/IPSL, ES-DOC

Giovanni Aloisio, IS-ENES2/CMCC, ESGF

10.4 Program Managers

Dr. Gerald Geernaert, DOE BER HQ

Dr. Justin Hnilo, DOE BER HQ

Dr. Dorothy Koch, DOE BER HQ

Dr. Tsendgar Lee, NASA HQ

Acknowledgments

The meeting organizers wish to thank the national and international funding agencies for providing travel funding for attendees to join the meeting in person, LLNL for hosting the annual event, and presenters for their contributions to the meeting and to this document. The organizers also wish to acknowledge Angela Jefferson for her help in processing endless paperwork, finding the meeting locations, and arranging many other important meeting logistics.

The development and operation of ESGF and UV-CDAT continue to be supported by the efforts of principal investigators, software engineers, data managers, and system administrators from many agencies and institutions worldwide. Primary contributors to these open-source software products include: ANL, ANU, BADC, CMCC, DKRZ, ESRL, GFDL, GSFC, IPSL, JPL, Kitware, NCAR, NYU, ORNL, LANL, LBNL, LLNL (leading institution), Tech-X, and the University of Utah. There are many other institutions that have contributed to the efforts of ESGF and UV-CDAT. Apologies, if you are not among the listed. The DOE, NASA, NOAA, IS-ENES2, and the Australian Department of Education provide major funding.