2015 Earth System Grid Federation Annual Report

June 19, 2015

Authors: ESGF Executive Committee and Working Team Leads

Abstract

The Earth System Grid Federation (ESGF) is a multi-agency, international collaboration whose purpose is to develop the software infrastructure needed to facilitate and empower the study of climate change on a global scale. ESGF's architecture employs a system of geographically distributed peer nodes that are independently administered yet united by common federation protocols and application programming interfaces. The cornerstones of its interoperability are the peer-to-peer messaging, which is continuously exchanged among all nodes in the federation; a shared architecture for search and discovery; and a security infrastructure based on industry standards. ESGF integrates popular application engines available from the open-source community with custom components (for data publishing, searching, user interface, security, and messaging) that were developed collaboratively by the team. The full ESGF infrastructure has now been adopted by multiple Earth science projects and allows access to petabytes of geophysical data, including the Coupled Model Intercomparison Project (CMIP)—output used by the Intergovernmental Panel on Climate Change assessment reports. ESGF is a successful example of integration of disparate open-source technologies into a cohesive functional system that serves the needs of the global climate science community.

Brief project summary

As global, broad-based model projections have become critical to understanding climate change, effectively managing the vast volumes of resulting data has become a major challenge for the climate and computational scientists who support the projections. To manage the massively distributed data volumes, the Earth System Grid Federation (ESGF) connects diverse federated archives for knowledge discovery. These distributed data sets have aided many researchers in producing significant articles and reports, such as the Intergovernmental Panel on Climate Change (IPCC) Third Assessment Report (AR3) and Fifth Assessment Report (AR5). Today, ESGF houses the tens of petabytes of data generated by such projects as the Department of Energy's Accelerated Climate Modeling for Energy (ACME) and the international CMIP and securely makes these data available to other climate scientists and non-researchers. Moreover, the access ESGF provides has translated into an impressive volume of new research[1]. In the next three to five years, it is estimated that the ESGF distributed archive will grow to hundreds of petabytes of data storage.

Objectives

The internationally distributed peer-to-peer ESGF "data cloud" archive represents the culmination of an effort that began in the late 1990s to (1) develop efficient, community-based tools to obtain relevant meteorological and other observational data, and (2) to develop custom computational models and export analysis tools for climate-change simulations, such as those used in IPCC reports. Established in 1995 by the World Climate Research Programme's (WCRP) Working Group on Coupled Modeling to support CMIP, ESGF provides a community-based infrastructure for climate model diagnosis, validation, intercomparison, documentation, and data access. Through ESGF, scientists are able to analyze general

---

[1] Between 2005 and 2015, based on the CMIP3 and CMIP5 ESGF archives, over 1,500 peer-reviewed journal articles were published.

circulation models in a systematic fashion, a process that facilitates model improvement. Virtually the entire international climate modeling community has participated in CMIP since its inception.

ESGF portals are gateways to scientific data collections hosted at sites around the globe; they allow the user to register[2] and potentially access all data and services within ESGF. Currently, more than 40 portals are in use, including one at Lawrence Livermore National Laboratory (LLNL) (http://pcmdi9.llnl.gov). ESGF allows international climate research teams to work in highly distributed research environments, using unique scientific instruments, petascale-class computers, and extreme amounts of data. A key to ESGF's success is its ability to effectively produce, validate, and analyze research results collaboratively. For example, new results generated by one team member are immediately accessible to the rest of the team, and they can annotate, comment on, and otherwise interact with those results.

Desired outcomes and deliverables

The growing international interest in ESGF development efforts has attracted many others who want to make their data more widely available and easy to use. For example, the World Climate Research Programme (WCRP), which provides governance for CMIP, has now endorsed the ESGF software foundation to be used for ~70 other model intercomparison projects (MIPs), such as obs4MIPs, TAMIP, CFMIP, and GeoMIP. At present, more than 40 projects are supported by ESGF, including data from 25 worldwide climate research centers spanning 21 countries. Virtually all climate researchers now have used ESGF directly or indirectly.

The world is full of large-scale data management and retrieval enterprise systems. In the U.S., for climate science alone, systems include the Global Change Master Directory, the Network-Object Mobile-Agent Dynamic System (NOMADS), the Atmospheric Radiation Measurement Archive, the Regional Climate Model Evaluation System, and the National Aeronautics and Space Administration (NASA) Distributed Active Archive Centers, to name a few. However, ESGF is the sole distributed data system and the only one to offer interoperability among disparate data sets (i.e., simulations, observations, and reanalysis data) for assessment study.

An important design criterion for ESGF was that it be open source and take advantage of open-source tools. The system leverages web servers and a few key protocols to provide secure communication. ESGF is not only the cornerstone platform for climate research and information sharing and collaboration, but is also designed to be a growing research platform for further investigation into new computing paradigms, including smart, autonomic, self-managing, and self-repairing computing infrastructures. ESGF is being examined for additional collaboration opportunities in science applications outside the climate community (such as energy, materials, nuclear energy, biology, chemistry, and fusion). This by no means indicates that we will abandon the climate community; we think that these efforts will strengthen the entire scientific community, with advances that will be beneficial for all.

Finally, besides our usual trademark advances in federation, analysis, and visualization, our efforts will look to expand in the areas of data mining, workflows, provenance and metadata, high-performance computing (HPC), data movement, and data ontology. These are all challenging areas related to the management, manipulation, storage, access, analysis, and visualization of large-scale scientific data.

Management plan

The management of ESGF has changed from a principal investigator (PI)–driven decision-making process to a committee-based decision-making process involving a Steering Committee and an Executive

---

[2] Users have a shared identity and authentication to all sites, thus, they register only once to gain access.

Committee[3]. Members of these committees are normally tied to one or more projects in the community with a specific mission or goal. To better coordinate international efforts, the Executive Committee has written a strategic planning document that describes ESGF development over the next ten years. This living document was be presented to the Steering Committee and posted on ESGF's website for community review.

The Steering Committee provides a forum for ESGF funding agencies to communicate, engage with, and coordinate their support for ESGF, and to help fashion a common vision for its evolution. The Executive Committee, which accepts guidance from and reports to the Steering Committee, provides general guidance and makes high-level decisions in directing and coordinating the course of ESGF, ensuring that they are consistent with multiple sponsor needs. In this process, individual PIs can make decisions, but their actions must be relayed to the Steering Committee. This governance model drives innovation and quality of services and helps to balance the conflict that exists between new development and day-to-day operations. The Executive Committee is scheduled to hold regular meetings every other month at minimum, and the Steering Committee will meet at least twice per year.

Additionally, the ESGF management and development teams have adopted the following tools to improve the coordination among developers and enhance the long-term stewardship and documentation of the software stack:

- For project task organization, the national and international team of ESGF developers and managers use the Atlassian on-demand Confluence and JIRA Agile software tools.
- Git is used for version control of the software stack repository.
- Github is used for online bug/enhancement tracking and reporting.

Each code-release version is tagged in Git and released to the community. The latest official release of the ESGF software stack is version 1.8. The development teams (17 sub-teams in all, see **Table 1**) comprise of 5 to 20 members that commit between 10% to 50% of their time. Two leads are assigned to each of the sub-teams, and the leads report progress to the ESGF Executive Committee over a designated time period.

Recent annual progress reports and presentations from the ESGF Executive Committee and sub-team leads can be found on the ESGF website (http://esgf.llnl.gov) along with the conference report for the 2015 ESGF and Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT) Face-to-Face (F2F) Conference (http://aims-group.github.io/pdf/2014-ESGF_UV-CDAT_Conference_Report.pdf).

Data plan

As discussed at the 2015 ESGF and UV-CDAT F2F conference, governance and use cases are real issues that determine how requirements affect operations and software development as they relate to projects and data. Therefore, encouraged by many supporting funding agencies, a significant fraction of projects utilizing ESGF to disseminate and analyze data were represented at the conference to voice concerns regarding current and future community data use cases. Discussions focused on maintaining essential operations and developing new and improved software to handle ever-increasing data variety, complexity, velocity, and volume. Data use cases for computing and data science activities that are critical to the community meeting its scientific mission (both as individual projects and as a federation of projects) are summarized in a series of presentations (found at http://esgf.llnl.gov/facetoface.html) and in the 2015 ESGF and UV-CDAT F2F conference report (http://aims-group.github.io/pdf/2014-ESGF_UV-

---

[3] The new ESGF governance and governance structure can be found at http://esgf.llnl.gov/governance.html.

CDAT_Conference_Report.pdf). Section 4 of the conference report describes three primary project data use cases and plans—for CMIP6, the Coordinated Regional Climate Downscaling Experiment (CORDEX), and ACME.

Web presence

A web presence is intrinsic to ESGF. It exists on three levels: (1) the ESGF user interface, which enables scientists to host, manage, and share scientific projects seamlessly from any location; (2) software development websites, such as the Confluence/JIRA Agile development tool and the Github code management and software repository; and (3) the official ESGF site, where anyone can learn about ESGF, uncover the latest software events, and download the ESGF software stack (http://esgf.llnl.gov).

In September 2015, the ESGF user interface (or portal front end) will be officially switched over to the CoG content management system. This new web interface will allow individual projects to be more identifiable and to customize their interface as needed. At the same time, it will also recognize the many sponsors that support the distributed archive and the development of its underlying software. To do so, a single project logo or label identifies each project at the space near the top of each webpage. As the user switches to other sites and projects in the federation to retrieve data or utilize resources, the project logo changes.

Sites not using the official ESGF front end may choose whether to display a project logo for their user interface in a fashion similar to CoG. However, since ESGF is the means by which the data are distributed to the community, these sites must have the icon/logo or words "powered by ESGF" displayed at all times to indicate ESGF's presence.

ESGF network activities can be explored at the International Climate Network Working Group (ICNWG) website: http://incwg.llnl.gov.

Project progress toward objectives

Progress on the analyzed requirements from each ESGF sub-project is under way. Many of the ESGF, UV-CDAT, and collaborating sub-teams are already engaged in high-priority innovations to sustain the climate community's data and computational activities. Although many projects utilize the ESGF data ecosystem, due to time constraints only the following few projects were able to provide feedback and requirements: CMIP6, WGCM Infrastructure Panel (WIP), NASA's satellite and weather program, CREATE-IP, obs4MIPs, CORDEX, climate4impact, NCPP, NMME, HIWPP, ACME, and GeoMIP. With this limitation and not in any particular order of priority, progress has been made on the following ESGF sub-tasks:

Table 1: *Project task progress.*

Category	Topic	Progress Statement
Management	Governance	A Steering Committee and an Executive Committee for ESGF governance has been put in place (http://esgf.llnl.gov/governance.html). The Executive Committee completed the ten-year strategic roadmap document, presented it to the Steering Committee, and is now working on an implementation document to achieve the goals of the strategic roadmap.
	Operations *Leads: ESGF XC*	For sustained success, operations are key for community confidence in the ESGF infrastructure. A support process has been put into place to help with operations. See "Community Outreach" section in the table.

Sub-task Progress	Compute Working Team *Leads:* *Dan Duffy (NASA)* *Charles Doutriaux (DOE)*	The Compute Working Team (CWT) is engaged in developing the capability to enable data-proximal analytics within ESGF. To guide the discussion, the team created several potential analytical use cases for the data stored in ESGF, using anomalies as the prototypical example of the type of analytics that scientists would want to perform. Next, the team focused on tailoring the interface to the analytic capabilities and reviewed several different potential programming interfaces, before deciding on the Web Processing Service (WPS) as the standard for defining the analytical services, inputs, and outputs. In addition to being rather simple to deploy, the geospatial community makes heavy use of WPS-enabled services. The team wrote the initial specification document and posted it on Confluence for team review. Starting with the pyWPS, the team has developed an initial implementation of WPS, which can be found in the GitHub repository. The next steps for the CWT will be continued development of the underlying analytic capabilities and exposure of anomalies and simple functions, such as average, maximum, and minimum, through the WPS. This will require WPS extension. The CWT will strive to have at least two WPS beta test sites with, at a minimum, the same set of test data. Finally, on the server side, the CWT is leaning toward a modular system, allowing for easy switching before many types of implementations of the same functions/algorithms. A rigorous test suite will ensure that all methods lead to the same results. A modular and flexible system is important because many technologies may lead to different levels of performance based on different problems.
	Publication Working Team *Leads:* *Sasha Ames (DOE)* *Rachana Ananthakrishnan (DOE)*	A publication service has been deployed using the Globus-based web front end in CADES for beta testing. This front end has been successfully used to publish project data sets for AMIP. This use case has provided the team an opportunity to better understand user needs and how to rapidly respond and integrate requested functionality. The dynamic nature of the ACME project has created the need for a publisher-configuration modification and access service, which is under development. The publication command line software now supports the addition of facet values for previously published data sets and a heterogeneous format handler, which will first be used with ACME. Additionally, the team is designing a more flexible configuration file format, motivated by data needs for ACME, based on XML to allow for differing facet value configurations under a single project heading.
	International Climate Network Working Group *Leads:* *Eli Dart (DOE)* *Mary Hester (DOE)*	During the past six months, ICNWG has made significant progress towards improving data transfer performance between ANU, BADC, DKRZ, and LLNL. Network performance has surpassed the goal of 500MB/sec in multiple test cases, but data transfer performance is lower (and in some cases, significantly lower). Additional diagnostics and system tuning work is under way to identify the causes of lower data transfer performance and assess the effort required to implement performance improvements.
	CoG User Interface Working Team *Leads:* *Cecelia DeLuca (NOAA)* *Luca Cinquini (NOAA)*	The CoG team has nearly completed the work necessary to make CoG the next-generation ESGF user interface. CoG has evolved from a single website application to a system of federated sites. Much effort was devoted to making CoG part of the ESGF standard installation, running the Django application behind an Apache *HTTPD* server, and supporting seamless OpenID authentication for all sites. Additionally,

		search capabilities have been significantly upgraded, including support for multiple facet options, bread crumbing, RESTful search URLs, and retrieval of the user's last search constraints and results. A major area of development consisted of integrating Globus data transfer with the CoG search interface and data cart, which will greatly improve the speed, reliability, and convenience with which users download data from the ESGF servers to their desktop. Finally, several improvements in style, appearance, and functionality were applied throughout the site.
	Identity Entitlement Access *Leads:* *Philip Kershaw (IS-ENES)* *Rachana Ananthakrishnan (DOE)*	The last six months have seen the conclusion of the first set of enhancement tasks for the ESGF access control system. This work has been focused around improvement of the user experience for data download. The Wget script data download system is simplified so that users no longer need X.509 certificate-based authentication. The user interface for OpenID login has also been improved. These changes were integrated into the 1.8 release of ESGF in March 2015. Efforts are now focused on new functionality to support user delegation and the eventual replacement of OpenID 2.0, which is now being deprecated across the IT community. The new work is based on the popular OAuth 2.0 framework and the OpenID Connect specification. An implementation of OAuth 2.0 is available and has been deployed at the Centre for Environmental Data Archival in the UK. The new system will be tested in pilots to integrate access control with Globus Online, CoG, and the ENES Climate Impacts Portal. Once complete, it will be integrated into a new release.
	Replication Working Team *Leads:* *Stephan Kindermann (IS-ENES)* *Tobias Weigel (IS-ENES)*	A new version of the Synda replication tool was developed that includes support for GridFTP: (https://forge.ipsl.jussieu.fr/prodiguer/wiki/docs/synda). The new tool version is made available on Github: (https://github.com/Prodiguer/synda).
	Versioning Working Team *Leads:* *Stephan Kindermann (IS-ENES)* *Tobias Weigel (IS-ENES)*	Various alternatives to retain the versioning history of ESGF published data sets were discussed, as were different versioning requirements at data centers. Conflicting requirements were detected, requiring a more formal collection of aspects and solution approaches in a "versioning requirements and roadmap" document, which will lead to a recommendation document for the WIP. Development will start in July 2015 of a prototype service that includes persistent identifier versioning information in the metadata associated with CMIP6 data.
	Data Transfer Working Team *Leads:* *Lukasz Lacinski (DOE)* *Rachana Ananthakrishnan (DOE)*	This team is working towards updating the Globus transfer option for data access in ESGF. The effort there is twofold: update the installation and configuration of the servers on ESGF data node needed for transfer and enhance the web download option on the ESGF portal. In the last few months, updates have been made to the installer to move to a binary installation rather than source installation and to an updated version of the server. This work is being targeted for the 1.9 release. The team has also supported use of the Globus download option and is working on an improved design for it. The team also worked with the NASA Jet Propulsion Laboratory team, especially in the design of the Globus transfer option in the CoG front end.

	Dashboard Working Team *Lead:* *Sandro Fiore (IS-ENES)*	Since the focus of the dashboard is to provide statistics related to user metrics (besides the small set of system metrics already available), in recent months the activities were focused on the implementation of back-end mechanisms to store, collect, and analyze specific information about data usage statistics and to develop and test federation-level aggregation mechanisms. Moreover, existing sensors have been refined and fixed where needed. In the coming weeks, the team will be working closely with the module ESGF Node Manager developers to integrate fine-grain statistics (e.g., ESGF by model, by variable, by data sets), by site and at the ESGF level. This activity will be carried out in close synergy with the EU IS-ENES2 project.
	Node Manager Working Team *Leads:* *Sasha Ames (DOE)* *Prashanth Dwarakanath (IS-ENES)*	The working team has presented and discussed requirements for replacing the node manager. Notable features of the new node manager will include Django service deployment running through Apache and modified module communication. This new node manager will be more robust and fault-tolerant and synchronizes attribute management services across the federation. It will also prevent loss of services when an attribute service provider becomes unreachable or goes down temporarily. Thus key sites to continue to communicate as peers using a revised protocol, but additional sites will communicate solely with an assigned primary site, with other sites serving as backups. The team has also implemented a prototype server to test communication and key functionality. They have performed continuous design review and design updates in tandem with the prototype implementation to determine the specifications to follow as they transition toward production-quality software.
	Quality Control Working Team *Leads:* *Martina Stockhause (IS-ENES)* *Katharina Berger (IS-ENES)*	The concept for the integration of external information into the ESGF platform is based on the examples Errata module (IPSL) and Citation module (DKRZ). WGCM Infrastructure Panel papers (in preperation) will formalize specifications to improve the ESGF software especially in the CMIP6 context. The crucial requirement of an effective versioning system and support was investigated in cooperation with ESGF replication and versioning teams. A coding sprint will soon start the implementation of these new features into the ESGF data publication process.
	Tracking and Feedback Notification Working Team *Lead:* *Sasha Ames (DOE)*	Plans are underway to retool the AccessLoggingFilter and perhaps replace it with a module that would work in the upcoming Apache-based web services to run on an ESGF node. The team is considering requirements for a notification service that would, at the top-level handle feedback to node administrators regarding status/connectivity and user notification of updates to previously accessed data sets.
	Metadata and Search Working Team *Lead:* *Luca Cinquini (NASA)*	Work on the ESGF search module has focused on providing operational upgrades of the current search infrastructure, while starting to design and prototype the architecture for scaling to support CMIP6 model output and larger observational data sets. Since the beginning of the year, the search module has undergone three releases (3.9.0, 3.9.1, 3.9.5), which included bug fixes, improvements to publishing and validation, and speed up of wildcard queries. At the same time, studies were started to upgrade the ESGF search engine to Solr4 and Solr-Cloud. This upgrade will allow better support for geospatial searches, more scalability and fault tolerance, and overall better performance. In the next six months, a roadmap will be drafted to execute this switch

		and prepare the ESGF infrastructure for the next generation of models and observations.
	Provenance Capture Working Team *Leads:* *Kerstin Kleese-Van Dam (DOE)* *Bibi Raju (DOE)*	The Provenance Working Team has been focusing on developing provenance solutions for reproducibility and performance investigations. The design of an initial Provenance Schema is complete, and the team is aiming to release a Provenance repository implementation in the next couple of weeks. The provenance library (to create and send provenance messages from the modeling code and other software tools) is currently being tested and the receiving ProvEn API is nearly completed. A version of the Velo software for ACME is set up on the ACME test system and will be used for submitting and monitoring ACME jobs. The team is working on integrating the Pegasus workflow manager and Velo for submitting and monitoring jobs and to collect provenance information.
New Potential Subtask	Software Security Working Team	With the new security issues, we may want to think about creating a separate working team that is solely dedicated to the integrity of the software stack.
Community Outreach	User Support Working Team *Leads:* *Torsten Rathmann (IS-ENES)* *Matthew Harris (DOE)*	After testing and using Askbot (a generic stack overflow), it was decided to abandon it in support of the mailing list. The user support mailing list (esgf-user@lists.llnl.gov) is going strong and a good place for users to ask questions once they have reviewed online documentation. In addition to the support mailing list, the support working team stood up a Frequently Asked Questions (FAQ) website (http://esgf.github.io/esgf-swt/). This new support site has comprehensively revised all of the ESGF FAQ pages from past sites and the ESGF wiki. With more than 50 questions and answers, the site covers a broad spectrum of user requests. The new FAQ support site is hosted by Github and has a new user-friendly interface design. With this new FAQ site, links to single questions are also possible.
	Documentation Working Team *Leads:* *Matthew Harris (DOE)* *Sam Fries (DOE)*	On the documentation front, the ESGF is defined on its website at http://esgf.llnl.gov. The website offers additional documents, such as sponsors, acknowledgements, governance, committees, publications, tutorials, supported projects, wikis, and much more. In the last year, the hosted website experienced major upgrades to its installation and tutorial documentation pages. Further, the "issue tracker" used to develop and plan changes to ESGF have migrated to a new host on Github for each of the working teams; it may be found under the "Wikis" menu: http://esgf.llnl.gov/wiki.html.
	Installation Working Team *Leads:* *Nicolas Carenton (IS-ENES)* *Prashanth Dwarakanath (IS-ENES)*	The installation working team has cut one major release and one hotfix release in 2015. Version 1.8 of the stack was out on February 27 and included the very first steps of the RPMization process that we decided to follow. Git and curl were the two candidates and are now installed from RPMs and no longer compiled from source. The other major improvement to the installer that was made before 1.8 that included the subsystem installer design review. Indeed, the revised function calls subsystem installers in whatever languages they were written in instead of sourcing them. This change required moving global variables from esg-node to a dedicated file called esg-init and helped to get rid of the bash stack size issue the team faced during installations from scratch. This new mechanism has already been implemented for Live Access Server (LAS) and CoG sub installers. Developers can now write their subcomponent installer in the

<table>
<tr>
<td></td>
<td>language of their choice. The next upcoming release is currently being prepared, and the installation working team has been focusing on packaging an Apache front end for ESGF as well as Globus as an RPM. This release will include CoG installation and operation using port 8000 and will also make the Solr service accessible from the standard http port (80) in addition to the Solr port 8983. This is a significant step toward overcoming firewall-related issues that are currently preventing sites from accessing all of the Solr services across the federation, leading to an inconsistent view of published data sets. The pilot deployment of Apache front end and the new http port connectivity to Solr has already been done on the Swedish data node (http://esg-dn1.nsc.liu.se) as a lead up to the next release.</td>
</tr>
</table>

For more details on the explanation of the sub-tasks above and their progress, please visit ESGF's Atlassian Confluence and JIRA site; the 2015 4th Annual Earth System Grid Federation and Ultrascale Visualization Climate Data Analysis Tools Conference Report, Lawrence Livermore National Laboratory, Livermore, CA, LLNL-TR-666753; the ESGF Github developer's site; and the ESGF official website.

If you cannot reach any these sites or if you need additional information, please contact Dean N. Williams at williams13@llnl.gov.

Appendix 1: List of Publications and Presentations

Group Citations:

John. L. Schnase, Tsengdar J. Lee, Chris A. Mattmann, Christopher S. Lynnes, Luca Cinquini, Paul M. Ramirez, Andrew F. Hart, Dean Williams, Duane Waliser, Pamela Rinsland, W. Phillip Webster, Daniel Q. Duffy, Mark A. McInerney, Glenn S. Tamkin, Gerald L. Potter, and Laura Carriere, "Big Data Challenges in Climate Science," *IEEE Geoscience and Remote Sensing Magazine* (2015) **[in process]**.

Peter J. Gleckler, Charles Doutriaux, Paul J. Durack, Karl E. Taylor, Yuying Zhang, Dean N. Williams, Erik Mason, and Jérôme Servonnat, "A Package to Make Performance Summaries from Climate Model Intercomparisons more Accessible," *Eos, Transactions American Geophysical Union* (2015) [**in process**].

Dean N. Williams and Rose Hansen, "A Global Repository for Planet-Sized Experiments and Observations," *Bulletin of the American Meteorological Society* (2015) [**under second peer-review**].

C. Palazzo, A. Mariello, S. Fiore, A. D'Anca, D. Elia, D. N. Williams, and G. Aloisio, "A Workflow-Enabled Big Data Analytics Software Stack for eScience," *The Second International Symposium on Big Data Principles, Architectures & Applications* (BDAA 2015), HPCS 2015, Amsterdam, The Netherlands, (2015) [**accepted for publication**].

Hashim Iqbal Chunpir, Dean N. Williams, and Thomas Ludwig, "Evolution of e-Research: From Infrastructure Development to Service Orientation," *HCI International Conference* (2015) [**accepted for publication**].

Chris A. Mattmann, Christopher S. Lynnes, Luca Cinquini, Paul M. Ramirez, Andrew F. Hart, Dean Williams, Duane Waliser1, and Pamela Rinsland, "Next Generation Cyber-infrastructure to Support Comparison of Satellite Observations with Climate Models," *Conference on Big Data from Space* (BIDS '15), Research, Technology and Innovation (RT&I), Frascati, Italy (2015) [**accepted for publication**].

Dean N. Williams, "Visualization and Analysis Tools for Ultrascale Climate Data," *Eos, Transactions American Geophysical Union* 95 (42), 377–378 (2014), DOI: 10.1002/2014EO420002.

D.N. Williams, Giri Palanisamy, Galen Shipman, Thomas Boden, and Jimmy Voyles, "Department of Energy Strategic Roadmap for Earth System Science Data Integration," *Big Data, 2014 IEEE International Conference Proceedings*, Washington D.C., USA, pp. 772–777 (2014) DOI:10.1109/BigData.2014.7004304.

Matthew Harris, "Webengine," *Proceedings of the World Congress on Engineering and Computer Science 2014*, Vol. I, WCESC 2014, San Francisco, USA, pp. 131–135 (2014) http://www.iaeng.org/publication/WCECS2014/.

Sandro Fiore, Alessandro D'Anca, Donatello Elia, Cosimo Palazzo, Dean N. Williams, Ian T. Foster, and Giovanni Aloisio, "Ophidia: A full software stack for scientific data analytics," *IEEE High Performance Computing & Simulation* (HPCS), 2014 International Conference, pp. 343–350 (2014) 10.1109/HPCSim.2014.6903706.

L. Cinquini, D.J. Crichton, C. Mattmann, J. Harney, G.M. Shipman, F. Wang, R. Ananthakrishnan, N. Miller, S. Denvil, M. Morgan, Z. Pobre, G.M. Bell, C.M. Doutriaux, R.S. Drach, D.N. Williams, P. Kershaw, S. Pascoe, E. Gonzalez, S. Fiore, and R. Schweitzer, "The Earth System Grid Federation: An open infrastructure for access to distributed geospatial data," *Future Generation Computer System* 36, 400–417 (2014), DOI: 10.1016/j.future.2013.07.002.

 S. Fiore, A. D'Anca, D. Elia, C. Palazzo, I. Foster, D. Williams, and G. Aloisio, "Ophidia: A Full Software Stack for Scientific Data Analytics," *Proc. of the 2014 International Conference on High Performance Computing & Simulation* (HPCS 2014), Bologna, Italy, pp. 343–350 (2014) ISBN: 978-1-4799-5311-0.

Reports:

D.N. Williams et al., *4th Annual Earth System Grid Federation and Ultrascale Visualization Climate Data Analysis Tools Conference Report*, Lawrence Livermore National Laboratory, Livermore, CA (2015) LLNL-TR-666753, http://aims-group.github.io/pdf/2014-ESGF_UV-CDAT_Conference_Report.pdf.

D.N. Williams et al., *3th Annual Earth System Grid Federation and Ultrascale Visualization Climate Data Analysis Tools Conference Report*, Lawrence Livermore National Laboratory, Livermore, CA (2014) LLNL-TR-650500, http://uvcdat.llnl.gov/media/pdf/ESGF_UV-CDAT_Meeting_Report_December2013.pdf.

Presentations:

The latest presentations for the 2014 ESGF & UV-CDAT Conference can be found on the following website: http://esgf.llnl.gov/facetoface.html.

Posters:

Distributed Computation Resources for Earth System Grid Federation (ESGF), Daniel Duffy, Charles Doutriaux, and Dean Williams, poster at the 2014 American Meteorological Society Conference, Phoenix, January 2015.

Distributed Computation Resources for Earth System Grid Federation (ESGF), Daniel Duffy, Charles Doutriaux, and Dean Williams, poster at the 2014 American Geophysical Union Conference, San Francisco, December 2014.

Awards:

2015 Federal Laboratory Consortium (National) Award for Advancing Federal Research and Technology for the Ultrascale Visualization for Climate Data Analysis Tools (UV-CDAT)

2014 Federal Laboratory Consortium (Far West Regional) Award for Advancing Federal Research and Technology for the Ultrascale Visualization for Climate Data Analysis Tools (UV-CDAT)

2013 Federal Laboratory Consortium (Far West Regional) Award for Advancing Federal Research and Technology for the Earth System Grid Federation (ESGF)

Appendix 2: Science highlights transmitted during the past year

Data used from the ESGF archives have helped to generate ~800 peer-reviewed CMIP5 articles: (http://cmip.llnl.gov/cmip5/publications/).