

Technical Developments for the Community

Dean N. Williams

Fourth Annual ESGF & UV-CDAT F2F Conference

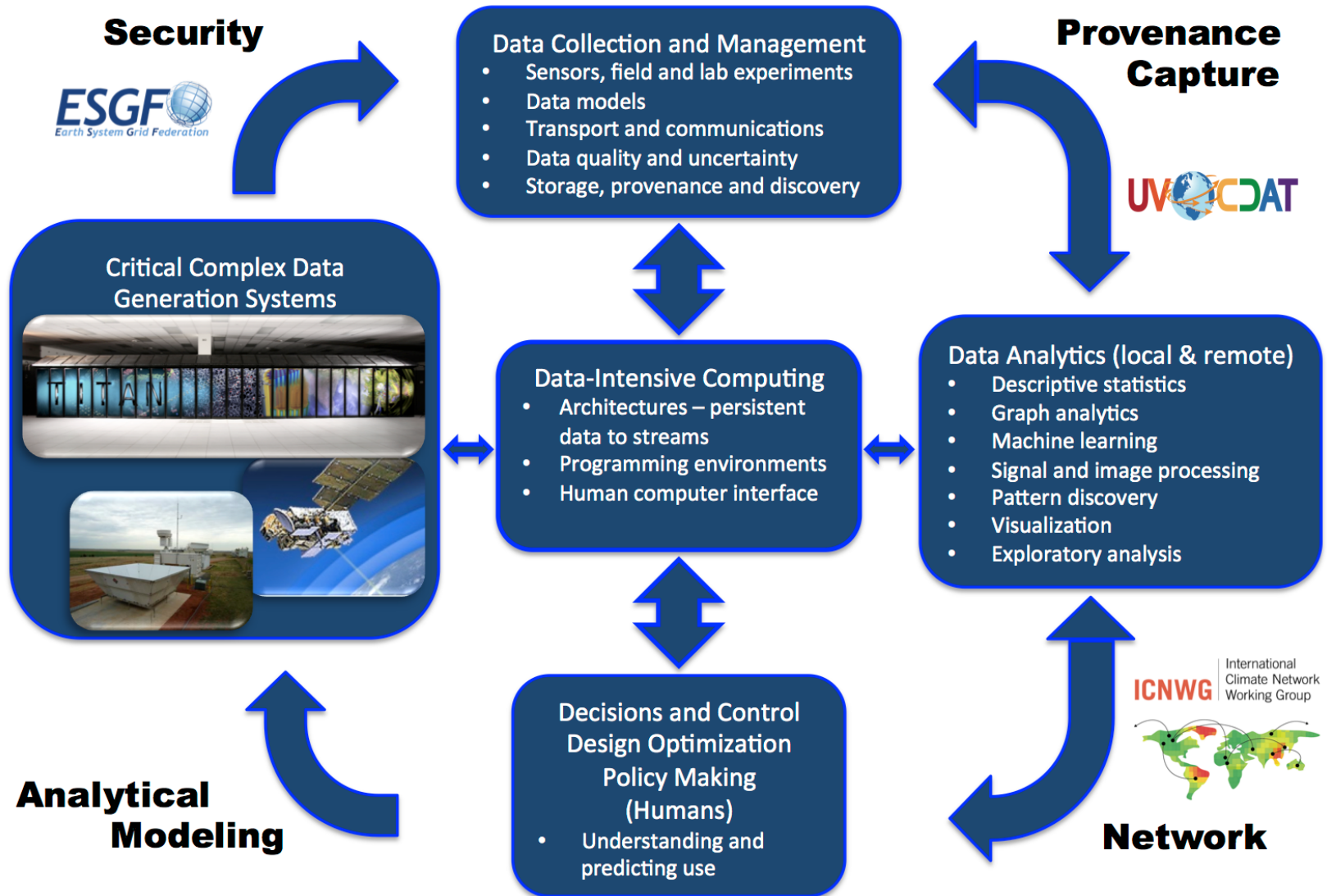
Livermore, California

December 9th – 11th, 2014

Grand Challenge of the Climate Data Ecosystem

Science Need and Challenge	Solution
Wide range of model runs, wide range of observations, and the wide range of analysis with many workflow types across multiple projects	Define use cases: 1) Data collection and Management (ESGF); 2) Efficient data extraction (Globus); 3) Data analysis and visualization (UV-CDAT, GrADS, NCL, Ferret), (e.g., regridder, satellite emulators); 4) Data-intensive computing; 5) High-performance computing (HPC), clusters; 6) Networks (ICNWG); 7) Decision and Control; and 8) Analytical modeling
Capture and record suites of model run experiments and compare results against other models and observations	Automated provenance and archiving; updating data (versioning) and replicating.
Quickly evaluate coupled model behavior by comparison with observations, and characterize uncertainties in model projections	Model metrics and diagnostics of the coupled system within one unified software system

Climate Community's Integrated Data Ecosystem and Workflow



Community Challenges for Enterprise System

Challenges	Description	Team
Heterogeneous Data Sets	The same infrastructure must also allow scientists to access and compare data sets from multiple sources, including from observational satellite, instrument sources, and reanalysis	ESGF
Installation	Software must adapted to multiple hardware platforms and operating systems located throughout the federation	Installation Working Team (ESGF-IWT)
Analysis, Diagnostics, and Visualizations	The generation of new and improved analysis, diagnostics, and visualization techniques for the better model development, intercomparison, and evaluation	Compute Working Team (ESGF-CWT)
Server-side and In Situ Computing	Server-side and in situ computation is necessary as the increase in data size and complexity of algorithms lead to data-intensive, compute-intensive challenges for diagnostics, UQ, analysis, model metrics, and visualization	Compute Working Team (ESGF-CWT), UV-CDAT
Provenance	Enables reproducibility, archiving and reuse of high-volume simulation data, provenance captures set up, execution and analysis details coupled with standard metadata creation, annotation, and forums for group discussions and sharing of any part of a workflow; Increases reproducibility, productivity and credibility of collaboration	Workflow and Provenance Working Team (ESGF-WPWT), UV-CDAT

Accelerating the Science

Features	Description	Team
Improve Usability	User support for infrastructure, including online training material, on-site training courses, and ongoing user support; Creates a flexible, extensible infrastructure for current and future efforts and related projects; Heightens productivity and user experience	Support Working Team (ESGF-SWT)
Data Sharing	Supports broad data sharing within community project teams and with scientific collaborations; Accelerates model development and result dissemination	CoG
User Interface	One stop shop to all needed capabilities, increases productivity, reproducibility; Specialized software when needed to enable web job submission, running, monitoring, and debugging capabilities on several HPC centers	CoG
Faster Analysis and Diagnostics (Parallelism)	Incorporate parallel capabilities into diagnostics framework for climatology generation and batch processing; add parallel support in UV-CDAT analysis and visualization frameworks where required	Compute Working Team (ESGF-CWT), UV-CDAT

Accelerating the Science

Deliverable	Description	Team
Faster Data Movement	Leverage the use of Globus transfer and sharing, integrated with ESGF; UV-CDAT at the LCF facilities and laboratories; work with ESnet to achieve good network performance throughput for rapid and secure data transport	Data Transfer Working Team (ESGF-DTWT) ; International Climate Network Working Group (ICNWG)
Improved Diagnostics	Incorporate standard diagnostics in UV-CDAT for all sub-model components, including more advanced diagnostics, UQ algorithms, ensembles, etc.	UV-CDAT
Exploratory and Analysis Visualization	Incorporate additional exploratory and analysis (i.e., EDEN) capabilities based on climate scientist and model developer requirements	UV-CDAT, Ferret, NCO, etc.
Improved Interface	Refine fundamental workflow issues with user interface in order to increase scientific productivity due to hard-to-use software	CoG
Tracking and Feedback	Display the current status and stats of user and federated sites	Dashboard Working Team (ESGF-DWT)

Atlassian: ESGF Confluence and Jira

ESGF

Earth System Grid Federation

Pages

Blog

SPACE SHORTCUTS

Meeting notes

JIRA reports

Product requirements

Retrospectives

Meeting notes (2)

File lists

Troubleshooting articles

PAGE TREE

Meeting notes

Recorded Meetings

Installation Working Team (esgf-iwt)

Compute Working Team (esgf-cwt)

Publication Working Team (esgf-pwt)

International Climate Network Working Group (ICNWG)

Identity Entitlement Access (esgf-idea)

Support Working Team (esgf-sw)

Replication and Versioning Working Team (esgf-rvwt)

User Interface Working Team (CoG)

Data Transfer Working Team (esgf-dwt)

Dashboard Working Team (esgf-dwt)

Node Manager Working Team (esgf-nmwt)

Quality Control Working Team (esgf-qwt)

Tracking and Feedback Notification Working Team (esgf-tfwt)

Workflow and Provenance Working Team (esgf-wpwt)

Metadata and Search Working Team (esgf-mswt)

Jira Setup

JIRA reports

Product requirements

Retrospectives

Meeting notes (2)

Requirements for a Biology node on ESGF

ESGF Members

File lists

GO-ESSP Discussion Topics

Troubleshooting articles

Pages

Earth System Grid Federation

Created by Dean N. Williams, last modified 5 minutes ago

The Earth System Grid Federation (ESGF) enterprise system is a collaboration that develops, deploys and maintains software infrastructure for the management, dissemination, and analysis of model output and observational data. ESGF's primary goal is to facilitate advancements in Earth System Science. It is an inter-agency and international effort led by the Department of Energy (DOE), and co-funded by National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), and international laboratories such as the Max Planck Institute for Meteorology (MPI-M), German Climate Computing Centre (DKRZ), the Australian National University (ANU) National Computational Infrastructure (NCI), Institut Pierre-Simon Laplace (IPSL), and the British Atmospheric Data Center (BADC). The ESGF mission is to:





- Support current CMIP activities, and prepare for future assessments
- Develop data and metadata facilities for inclusion of observations and reanalysis products for CMIP use
- Enhance and improve current climate research infrastructure capabilities through involvement of the software development community and through adherence to sound software principles
- Foster collaboration across agency and political boundaries
- Integrate and inter-operate with other software designed to meet the objectives of ESGF: e.g., software developed by European IS-ENES2, NASA, NOAA, ESIP, and the Australia's NCI
- Create software infrastructure and tools that facilitate scientific advancements

ESGF is a component architecture expressly designed to handle large-scale data management and computational resources for worldwide distribution. The team of computer scientists and climate scientists has developed an operational system for serving climate data from multiple locations and sources. Model simulations, satellite observations, and reanalysis products are all being served from the ESGF distributed data archive.

Sponsors


U.S. Sponsors:

DOE NASA NOAA NSF




European Sponsor

IS-ENES2






Australian Sponsor

NCI



Earth System Grid Federation




6th Coupled Model Intercomparison Project (CMIP6)

The 6th Coupled Model Intercomparison Project (CMIP6) is the largest international effort to coordinate and standardize climate model simulations. It is a multi-agency and international effort led by the Department of Energy (DOE), and co-funded by National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), and international laboratories such as the Max Planck Institute for Meteorology (MPI-M), German Climate Computing Centre (DKRZ), the Australian National University (ANU) National Computational Infrastructure (NCI), Institut Pierre-Simon Laplace (IPSL), and the British Atmospheric Data Center (BADC). The CMIP6 mission is to:

- Support current CMIP activities, and prepare for future assessments
- Develop data and metadata facilities for inclusion of observations and reanalysis products for CMIP use
- Enhance and improve current climate research infrastructure capabilities through involvement of the software development community and through adherence to sound software principles
- Foster collaboration across agency and political boundaries
- Integrate and inter-operate with other software designed to meet the objectives of ESGF: e.g., software developed by European IS-ENES2, NASA, NOAA, ESIP, and the Australia's NCI
- Create software infrastructure and tools that facilitate scientific advancements

ESGF is a component architecture expressly designed to handle large-scale data management and computational resources for worldwide distribution. The team of computer scientists and climate scientists has developed an operational system for serving climate data from multiple locations and sources. Model simulations, satellite observations, and reanalysis products are all being served from the ESGF distributed data archive.

Blog stream



Minor changes to publisher

Sasha Ames posted on Sep 03, 2014

I've made updates to the publisher software scripts with two minor changes. These changes are only committed thus far to the ames-dev branch of the esg-publisher git repository. esgscan_directory now will produce an error message when the directory format is malformed vs the esg.ini specification esgpubsh now has a --nodbwite option. this will go through the input file for publication validation steps shy of actually writing anything to postgresql...

About us

The Earth System Grid Federation for the climate community.

Today

Monday, Dec 8, 2014

Add Event

08 Today

Dec

ESGF - Compute Working Team (esgf-cwt)

ESGF Working Team

International Climate Network Working Group (ICNWG)

ESGF Working Team

09 Tuesday

Dec

GO-ESSP Tech Telcos

ESGF Working Team

11 Thursday

Dec

ESGF Installation Working Team (esgf-iwt)

ESGF Working Team

17 Wednesday

Dec

ESGF Publication Working Team (esgf-pwt)

ESGF Working Team

18 Thursday

Dec

ESGF Identity, Entitlement, Access (esgf-idea)

ESGF Working Team

22 Monday

Dec

ESGF - Compute Working Team (esgf-cwt)

ESGF Working Team

International Climate Network Working Group (ICNWG)

ESGF Working Team

23 Tuesday

Dec

GO-ESSP Tech Telcos

ESGF Working Team

25 Thursday

Dec

ESGF Installation Working Team (esgf-iwt)

ESGF Working Team

More

Recently updated

Earth System Grid Federation

5 minutes ago • updated by Dean N. Williams • view change

esgf node problems

Dec 04, 2014 • created by Sasha Ames

Troubleshooting articles

Dec 04, 2014 • created by Sasha Ames

Replication and Versioning Working Team Members

Dec 04, 2014 • updated by Stephan Kindermann • view change

2014-11-20 Meeting notes

