Technical Developments for the Community

Dean N. Williams

Fourth Annual ESGF & UV-CDAT F2F Conference

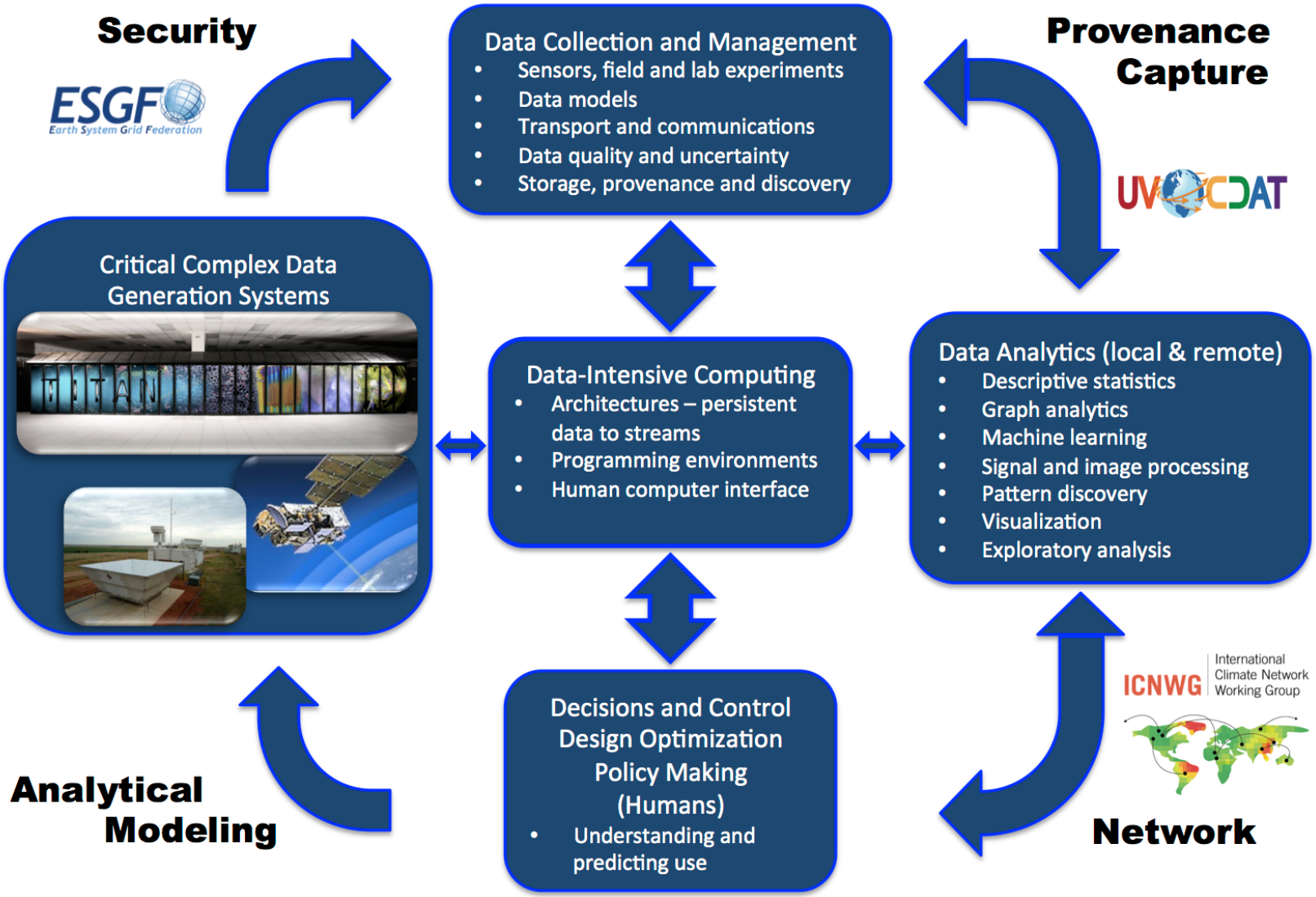
Livermore, California

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

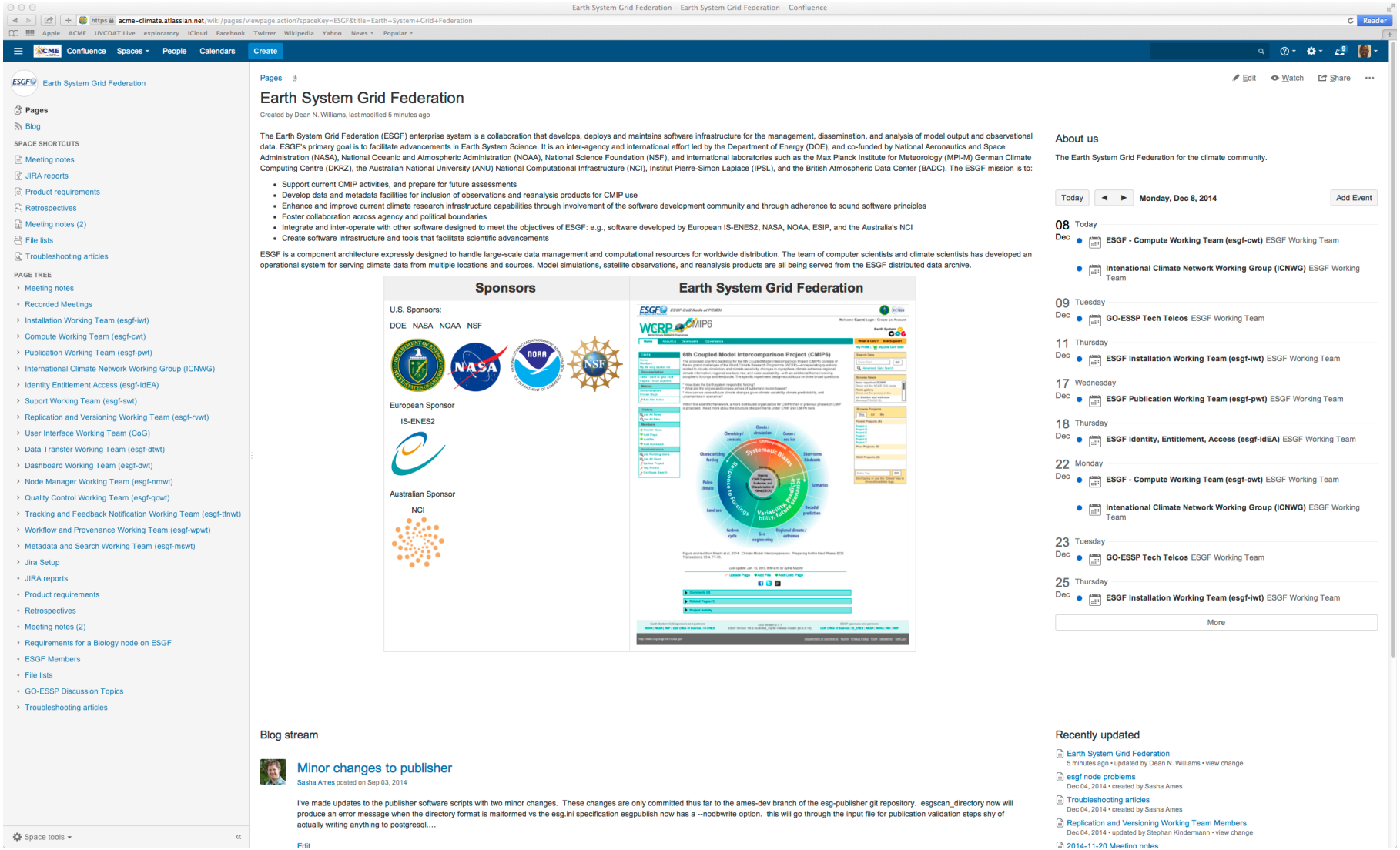
Accelerating the Science

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



Atlassian: UV-CDAT Confluence and Jira