# Presentation and Poster Abstracts

## Day 1: Tuesday, 8 December 2015
## Science Drivers Project Requirements and Feedback

| Title and Presenter | Abstract |
|---|---|
| **WGCM Infrastructure Panel**<br><br>***Karl Taylor (DOE/LLNL, taylor13@llnl.gov)***<br>***V. Balaji (NOAA/GFDL, balaji@princeton.edu)*** | The Working Group on Coupled Modeling (WGCM) Infrastructure Panel (WIP) was formed in response to the WGCM's (2013) expressed need to provide scientific guidance and requirements for the global data infrastructure underpinning global climate science and modeling. This infrastructure includes ESGF software, and other tools such as: ES-DOC, CoG, CMOR, CF Conventions, and others. Chaired by V. Balaji (Princeton/GFDL) and K. Taylor (PCMDI), it outlined in 2014 a strategy to develop a series of "position papers" on global data infrastructure and its interaction with the scientific design of experiments; and to present them to the WGCM annual meeting for endorsement by the WGCM, the CMIP Panel, and the modeling groups. A series of position papers were unveiled at the WGCM-19 meeting (2015) in Dubrovnik. The 11 position papers currently in draft, and others in progress, will be available on the WIP website, https://goo.gl/eJxDvL. |
| **DOE Accelerated Climate Modeling for Energy**<br><br>**David Bader (DOE/LLNL, bader2@llnl.gov)** | The Accelerated Climate Modeling for Energy (ACME) project is sponsored by the U.S. Department of Energy's (DOE's) Office of Biological and Environmental Research (BER) to develop and apply a computationally advanced climate and Earth system model to investigate the challenges posed by the interactions of climate change and societal energy requirements.<br>The ACME model simulates the fully coupled climate system at high-resolution (15-25km) and will include coupling with energy systems, with focus on a near-term hindcast (1970-2015) for model validation and a near-term projection (2015-2050) most relevant to societal planning. The model further employs regional-refinement using advanced adaptive mesh methodologies in order to provide ultra-high-resolution to resolve critical physics and meteorological phenomena. The ACME model branched from the Community Earth System Model (CESM), and increasingly, its code will be designed to optimize performance on current and future DOE Leadership Class computers.<br>ACME's initial scientific goals address three areas of importance to both climate research and society:<br>1. **Water cycle:** How do the hydrological cycle and water resources interact with the climate system on local to global scales?<br>2. **Biogeochemistry:** How do biogeochemical cycles interact with global climate change?<br>3. **Cryosphere-ocean system:** How do rapid changes in cryosphere-ocean systems interact with the climate system?<br>ACME's computational focus looks to advance climate codes in the following four area:<br>**1. Computational Trends**<br>**2. Modern Software Engineering**<br>**3. Biogeochemistry**<br>**4. Cryosphere-Ocean System** |
| **Obs4MIPs**<br><br>**Peter Gleckler (DOE/LLNL, gleckler1@llnl.gov)** | |
| **IS-ENES**<br><br>**Sébastien Denvil (ENES/IPSL, sebastien.denvil@ipsl.jussieu.fr)** | |
| **CREATE-IP**<br><br>**Jerry Potter (NASA/GSFC, jpotter@ucdavis.edu)** | |

# Day 1: Tuesday, 8 December 2015
# Required Data Center and Interoperable Services

| Title and Presenter | Abstract |
|---|---|
| **ANU/NCI**<br><br>***Ben Evan (ANU/NCI, Ben.Evans@anu.edu.au)*** | |
| **DDC/DKRZ**<br><br>***Stephan Kindermann (ENES/DKRZ, kindermann@dkrz.de)*** | |
| **DOE LLNL/PCMDI**<br><br>***Dean N. Williams (DOE/LLNL, williams13@llnl.gov)*** | |
| **IS-ENES/IPSL**<br><br>***Sébastien Denvil (ENES/IPSL, sebastien.denvil@ipsl.jussieu.fr)*** | |
| **IS-ENES/CEDA**<br><br>***Phil Kershaw (ENES/CEDA, philip.kershaw@stfc.ac.uk)*** | |

# Day 2: Wednesday, 9 December 2015
# Advanced Computational Environments and Data Analytics

| Title and Presenter | Abstract |
|---|---|
| **Compute working Team Overview**<br><br>***Daniel Duffy (NASA/GSFC, daniel.q.duffy@nasa.gov)*** | |
| **WPS Overview and Demo**<br><br>**Charles Doutriaux (DOE/LLNL, doutriaux1@llnl.gov)** | |
| **Analytics as a Service Framework**<br><br>**Thomas Maxwell (NASA/GSFC, thomas.maxwell@nasa.gov)** | |
| **Ophedia**<br><br>**Sandro Fiore (ENES/CMCC, sandro.fiore@unisalento.it)** | The Ophidia project is a research effort on big data analytics facing scientific data analysis challenges in multiple domains (e.g. climate change). Ophidia provides declarative, server-side, and parallel data analysis, jointly with an internal storage model able to efficiently deal with multidimensional data and a hierarchical data organization to manage large data volumes ("*datacubes*"). The project relies on a strong background on high performance database management and OLAP systems to manage large scientific datasets.<br>The Ophidia analytics platform provides several *data operators* to manipulate *datacubes*, and *array-based primitives* to perform data analysis on large scientific data arrays (e.g. statistical analysis, FFT, DWT, subsetting, compression). Metadata management support (CRUD-like operators) is also provided. The server front-end exposes several interfaces to address interoperability requirements: WS-I$^{+}$, GSI/VOMS and OGC-WPS (through PyWPS). From a programmatic point of view a Python module (PyOphidia) makes straightforward the integration of Ophidia into Python-based environments and applications (e.g. IPython).The system offers a Command Line Interface (e.g. bash-like) for end-users, with a complete set of commands, as |

| | |
|---|---|
| | well as integrated help and manuals.<br>A key point of the talk will be the workflow capabilities offered by Ophidia. In this regard, the framework stack includes an internal workflow management system, which coordinates, orchestrates, and optimizes the execution of multiple scientific data analytics & visualization tasks (e.g. statistical analysis, metadata management, virtual file system tasks, maps generation, import/export of datasets in NetCDF format). Specific macros are also available to implement loops, or to parallelize them in case of data independence. Real-time workflow monitoring execution is also supported through a graphical user interface.<br>Some real workflows implemented at CMCC and related to different projects will be also presented: climate indicators in the FP7 EU CLIPC and EUBRAZILCC, fire danger prevention analysis in the INTERREG OFIDIA, and finally, large scale climate model intercomparison data analysis (e.g. precipitation trend analysis, climate change signal analysis, anomalies analysis) in the H2020 INDIGO-DataCloud. |
| **WPS Service and Back-end**<br><br>**Maarten Plieger (ENES/KNMI, maarten.plieger@knmi.nl)** | |

# Day 2: Wednesday, 9 December 2015
# ESGF Development for Data Centers and Interoperable Services

| Title and Presenter | Abstract |
|---|---|
| **CoG User Interface Working Team**<br><br>***Sylvia Murphy (NOAA/ERSL, sylvia.murphy@noaa.gov)*** | Throughout 2015, the ESGF User Interface Working Team (UIWT) has worked on upgrading and expanding the Earth System CoG Collaboration Environment to replace the old ESGF web front-end. Major new features include: a) integration of CoG into the ESGF software stack; b) federation of distributed CoGs; c) support for downloading data via Globus; and d) general improvements to the site, infrastructure upgrades, and security fixes. CoG is now ready to be deployed as the ESGF front-end at each node. For the next 6 months, the priority of the ESGF UIWT will be be supporting ESGF administrators and end-users, while at the same time collecting and prioritizing requirements for additional needed functionality. |
| **Compute Working Team**<br><br>***Charles Doutriaux (DOE/LLNL, doutriaux1@llnl.gov)*** | |
| **Dashboard Working Team**<br><br>***Paola Nassisi (ENES/CMCC, paola.nassisi@cmcc.it)***<br>***Sandro Fiore (ENES/CMCC, sandro.fiore@unisalento.it)***<br>***Giovanni Aloisio (ENES/CMCC, giovanni.aloisio@unisalento.it)*** | Monitoring the Earth System Grid Federation is a challenging topic. From an infrastructural standpoint the dashboard & desktop components provide the proper environment for capturing usage metrics, as well as system status information at local (node) and global (institution and/or federation) level. The Dashboard and the Desktop are strongly coupled and integrated into the ESGF stack and represent the back- and the front-end of the ESGF monitoring system.<br>The Dashboard acts as information provider, collecting and storing a high volume of heterogeneous metrics, covering machine performance, network topology, host/service mapping and registered users as well as download statistics. The Desktop is a web-based environment and provides an effective, transparent, robust and easy access to all the metrics and statistics provided by the Dashboard. It is written in Java and JavaScript programming languages and presents enhanced views with several gadgets (enriched with charts, tables and maps) for a simple and user-friendly visualization of aggregated and geo-localized information.<br>All the metrics collected by the ESGF monitoring infrastructure are stored in a system catalog that has been extended to support multiple information about the data usage statistics. More specifically, in addition to information like the number of downloads, downloaded datasets, users that have downloaded some data, the amount of data downloaded etc., new metrics are being provided. Some examples are statistics about data downloads grouped by model, variable or experiment, by country or over time, top ten list of the most downloaded datasets or clients distribution maps. To this end, specific data marts have been created to allow a fast access to this information.<br>Finally, to grant a programmatic access to the metrics managed by the Dashboard, a set of RESTful APIs has been defined (based on a JSON data interchange format) allowing the user to design and implement his/her own client applications. |
| **Data Transfer Working Team**<br><br>***Rachana Ananthakrishnan (DOE/ANL, ranantha@uchicago.edu)***<br>***Luckasz Lacinski (DOE/ANL, lukasz@uchicago.edu)*** | |
| **Support Working Team**<br><br>***Matthew Harris (DOE/LLNL, harris112@llnl.gov)*** | |
| **Identity Entitlement Access** | |

| | |
|---|---|
| **Team**<br><br>***Philip Kershaw (ENES/BADC, philip.kershaw@stfc.ac.uk)*** | |
| **Installation Working Team**<br><br>***Nicolas Carenton (ENES/IPSL, ncarenton@ipsl.jussieu.fr)***<br>***Prashanth Dwarakanath (ENES/Liu, pchengi@nsc.liu.se)*** | |
| **International Climate Network Working Group**<br><br>***Eli Dart (DOE/ESnet, dart@es.net)*** | |
| **Metadata and search Working Team**<br><br>***Luca Cinquini (NASA/JPL, Luca.Cinquini@jpl.nasa.gov)*** | For the ESGF Metadata and Search Working Team (ESGF-MSWT), the year 2015 was largely dominated by the general ESGF security incident, which prompted the whole federation to be brought offline. The ESGF-MSWT took advantage of this unfortunate situation to execute a much needed upgrade of the ESGF search services infrastructure, which would have been much more difficult as a backward-compatible upgrade. As a consequence, the upcoming ESGF 2.0 software stack will utilize Solr 5, deployed as a standalone engine embedded within Jetty, which includes many important new features such as atomic updates. The general master/slave/replica architecture hasn't changed, but the Solr slave shard will be exposed through the standard HTTP port 80 to avoid pesky firewall issues. Additional, support for publishing data to a new "local shard" has been introduced. From the User Interface perspective, many improvements have been added to the search pages, the administrator configuration utilities, and the data cart. In the next year, the ESGF-MSWT main focus will be to support the upcoming CMIP6 distributed data archive, and related observational data. Major areas of development will include metadata validation, partition of the global search space into virtual organizations, scalability and performance. |
| **Node Manager Working Team**<br><br>***Sasha Ames (DOE/LLNL, ames4@llnl.gov)***<br>***Prashanth Dwarakanath (ENES/Liu, pchengi@nsc.liu.se)*** | |
| **Persistent Identifier Services**<br><br>***Tobias Weigel (ENES/DKRZ, weigel@dkrz.de)***<br>***Stephan Kindermann (ENES/DKRZ, kindermann@dkrz.de)***<br>***Katharina Berger (berger@dkrz.de)*** | Persistent Identifier (PID) services for ESGF are concerned with the automated assignment and curation of persistent identifiers for CMIP6 data managed in ESGF at several levels of granularity. PIDs will be assigned to all CMIP6 files as well as several higher levels of aggregation, covering datasets, simulations and models. Identifier names are generated by CMOR and registered as part of the overall publishing workflow. An exemplary application based on the PID service is a smart user workspace tool that can pull additional information on given files from the federation, tell whether a new dataset version is available and ultimately provide access to it.<br>The presentation will give an overview on the service design as also described in the corresponding WIP paper and provide an update on the current development status. The service architecture is based on a distributed message queue to achieve high availability and throughput. The PID services interact with other ESGF components, including versioning, replication and citation services. |
| **Provenance Working Team**<br><br>***Bibi Raju (DOE/PNNL bibi.raju@pnnl.gov)*** | Provenance team aims to focus on the development of provenance solutions in support of reproducibility and performance investigations to accomplish the Accelerated Climate Modeling for Energy (ACME) computational goals. This includes development of a provenance format that can capture sufficient information to enable scientists to reproduce their previous calculations correctly as well as capture and link to performance information for specific workflows and model runs to enable in-depth performance analysis. The first step is to investigate methods for the capture, representation and storage, evaluation, access and use of provenance information. Over the last year, we have been developing a comprehensive workflow performance data model called Open Provenance Model-based Workflow Performance Provenance (OPM-WFPP). It enables the structured analysis of workflow performance characteristics and variability. It also links provenance information and performance metrics ontology.<br>The provenance capture ontology and system enables the capture of provenance from the high-level workflow through all relevant system levels in one integrated environment. A provenance production and collection framework is in place called Provenance Environment (ProvEn). It provides components supporting the production and collection of provenance information for distributed application environments. Semantic Web technologies and ontologies, including the Open Provenance Model – Workflow Performance Provenance (OPM-WFPP) ontology, are used by ProvEn for the representation, storage, and reporting of provenance. We are currently in the process of developing a provenance capture mechanism that can handle the high-velocity provenance information. |
| **Publication Working Team**<br><br>***Sasha Ames (DOE/LLNL,*** | |

| | |
|---|---|
| *ames4@llnl.gov)* | |
| **Quality Control Working Team**<br><br>***Martina Stockhause (ENES/DKRZ, stockhause@dkrz.de)***<br>***Guillaume Levavasseur (ENES/IPSL, glipsl@ipsl.jussieu.fr)***<br>***Katharina Berger (ENES/DKRZ, berger@dkrz.de)*** | The ESGF-QCWT aims to improve the quality of ESGF user services by integration of additional external documentations. The team coordinates the implementation of the errata service (IPSL) and the data citation service (DKRZ). We will present the team's progress over the last 12 months and give a roadmap for the next year with special emphasis on requirements, collaboration and risk aspects. |
| **Replication and Versioning Working Team**<br><br>***Stephan Kindermann (ENES/DKRZ, kindermann@dkrz.de)***<br>***Tobias Weigel (ENES/DKRZ, weigel@dkrz.de)*** | Ensuring ESGF CMIP6 data consistency across sites strongly depends on stable and agreed versioning and replication procedures. On one hand this requires common software components (versioning support as part of publication procudure and replication software like synda) - yet on the other hand operational agreements and the adherence to "versioning, replication and publication best practices" is necessary. The presentation will describe the currents status of the software as well as agreements aspects. As part of this also a short summary of the "replication and versioning" WIP paper is given. A roadmap for 2016 will be discussed highlighting open issues to be resoved. The collaboration aspects with the icnwg team and the publication team are summarized as well as future versioning and replication support aspects of the proposed persistent identifier ESGF |
| **Software Security Working Team**<br><br>***Prashanth Dwarakanath (ENES/LIU, pchengi@nsc.liu.se)*** | |
| **Tracking/Feedback Notification Working Team**<br><br>***Sasha Ames (DOE/LLNL, ames4@llnl.gov)***<br>***Sandro Fiore (ENES/CMCC, sandro.fiore@unisalento.it)*** | |
| **User Support Working Team**<br><br>***Torsten Rathmann (ENES/DKRZ, rathmann@dkrz.de)*** | |

# Day 3: Thursday, 10 December 2015
# Coordinated Efforts with Community Software Projects

| Title and Presenter | Abstract |
|---|---|
| **THREDDS Data Server (TDS)**<br><br>***John Caron (Independent, jcaron1129@gmail.com)*** | |
| **Science DMZ for ESGF Super Nodes**<br><br>***Eli Dart (DOE/ESnet, dart@es.net)*** | |
| **Named Data Networking (NDN)**<br><br>***Christos Papadopoulos (Colorado State, christos@colostate.edu)*** | |
| **Climate Model Output Rewriter (CMOR)**<br><br>***Denis Nadeau (DOE/LLNL, nadeau1@llnl.gov)*** | |
| **Synda** | |

| | |
|---|---|
| ***Sébastien Denvil (ENES/IPSL, sebastien.denvil@ipsl.jussieu.fr)*** | |
| **Globus**<br><br>***Rachana Ananthakrishnan (DOE/ANL, ranantha@uchicago.edu)*** | |
| **On-Demand Streaming Ensembles**<br><br>***Cameron Christensen (University of Utah, cam@sci.utah.edu)*** | |

# Day 3: Thursday, 10 December 2015
# Poster Session

| Title and Presenter | Abstract |
|---|---|
| **Climate4Impact Portal**<br><br>***Maarten Pileger (ENES/KNMI, maarten.plieger@knmi.nl)*** | The aim of climate4impact (C4I) is to enhance the use of Climate Research Data and to enhance the interaction with climate effect/impact communities. The portal is based on impact use cases from different European countries, and is evaluated by a user panel consisting of use case owners. It has been developed within the European projects IS-ENES and IS-ENES2 for more than 6 years, and its development currently continues within IS-ENES2. As the climate impact community is very broad, the focus is currently mainly on the scientific impact community. This work has resulted in the ENES portal interface for climate impact communities and can be visited at http://climate4impact.eu/.<br>C4I is connected to the Earth System Grid Federation (ESGF). A challenge was to describe the available model data and how it can be used. The portal warns users about possible pitfalls when using climate models. All impact use cases are described in the documentation section, using highlighted keywords pointing to detailed information in the glossary.<br>The main goal for C4I can be summarized by two objectives: The first, to work on a web interface, which generates a graphical user interface on WPS endpoints. These endpoints calculate climate indices and subset data using OpenClimateGIS/icclim on data stored in ESGF data nodes. Data is transmitted from ESGF nodes over secured OpenDAP and becomes available in a new, per user, secured OpenDAP server. The results are visualized using ADAGUC. Dedicated wizards for processing of climate indices are developed in close collaboration with users. The second, to expose C4I services to offer standardized services, which can be used by other portals, like the EU FP7 CLIPC portal. This has the advantage to add interoperability between several portals, as well as to enable the design of specific portals aimed at different impact communities, either thematic or national. |
| **The Climate Data Analytic Services (CDAS) Framework**<br><br>***Thomas Maxwell (NASA/GSFC, thomas.maxwell@nasa.gov)*** | Faced with unprecedented growth in the Big Data domain of climate science, NASA has developed the Climate Data Analytic Services (CDAS) framework. This framework enables scientists to execute trusted and tested analysis operations in a high performance environment close to the massive data stores at NASA. The data is accessed in standard (NetCDF, HDF, etc.) formats in a POSIX file system and processed using trusted climate data analysis tools (ESMF, CDAT, NCO, etc.). The framework is structured as a set of interacting modules allowing maximal flexibility in deployment choices.<br>CDAS services are accessed via a WPS API being developed in collaboration with the ESGF Compute Working Team to support server-side analytics for ESGF. The API can be executed using either direct web service calls, a python script or application, or a javascript-based web application. Client packages in python or javascript contain everything needed to make CDAS requests.<br>The CDAS architecture brings together the tools, data storage, and high-performance computing required for timely analysis of large-scale data sets, where the data resides, to ultimately produce societal benefits. It is is currently deployed at NASA in support of the Collaborative REAnalysis Technical Environment (CREATE) project, which centralizes numerous global reanalysis datasets onto a single advanced data analytics platform. This service permits decision makers to investigate climate changes around the globe, inspect model trends, compare multiple reanalysis datasets, and variability. |
| **ACME Dashboard**<br><br>***Matthew Harris (DOE/LLNL, harris112@llnl.gov)*** | Supporting the ACME community in model development, testing and usage requires the utilization of many complex and ever-changing components from model modules and script version to computer systems and diagnostics. In particular, in collaborative development efforts it is often difficult to keep track of the latest version of specific model and scripts, which set up parameters where used by collaborators or which runs still need to be completed. The ACME Dashboard is an integrated development environment that aims to support the required 'book keeping' and coordination effort by integrating secure resources access (storage, computing), component registers (data, models, diagnostics, workflows, etc.), provenance (usage information) and work execution (e.g., run workflow, use diagnostics) in one graphical environment. |
| **HPSS Connection to ESGF** | Accessing data stored on tape archives is difficult, time consuming, and prone to error. The ACME project plans to create 100s TB-PBs of data, all of which is not feasible to store on disk-based archives. To address |

| | |
|---|---|
| ***Sam Fries (DOE/LLNL, fries2@llnl.gov)***<br>***Alex Sim (DOE/ LBNL, asim@lbl.gov)*** | this, we are bridging HPSS and ESGF, allowing data sets stored on tape to be accessed through the same methods that climate scientists are already familiar with. LBNL's Berkeley Archival Storage Encapsulation (BASE) library provides a simple API for retrieving metadata as well as actual data from HPSS and other storage systems. We are creating a Python Web application that uses BASE to access and retrieve data, and allow that data to be published to ESGF. Our initial platform will test HPSS at NERSC with ESGF nodes at LLNL, with plans to deploy at other ACME sites such as OLCF and ALCF. |
| **Distributed Resource for the ESGF Advanced Management (DREAM)**<br><br>***Dean N. Williams (DOE/LLNL, williams13@llnl.gov)***<br>***Luca Cinquini (NASA/JPL, Luca.Cinquini@jpl.nasa.gov)*** | We envision that the Distributed Resource for the ESGF Advanced Management (DREAM) project will accelerate discovery by enabling climate researchers, among other types of researchers, to manage, analyze, and visualize data from earth-scale measurements and simulations. DREAM's success will be built on proven components that leverage existing services and resources. A key building block for DREAM will be the ESGF. Expanding on the existing ESGF, the project will ensure that the access, storage, movement, and analysis of the large quantities of data that are processed and produced by diverse science projects can be dynamically distributed with proper resource management.<br>Much of the DOE Office of Science data is currently generated by multiple stand-alone facilities. DREAM can collect data accumulated from these facilities and incorporate it into a fully integrated network accessible from anywhere in the world. The result is a completely new paradigm shift for data management, analysis, and visualization enabling researchers to:<br>1. Manage their calculations, data, tools, and research results;<br>2. Ensure that all data are sharable, reproducible and (re)usable—accompanied by appropriate metadata describing its provenance, syntax, and semantics at creation;<br>3. Advance application performance by selectively adapting APIs and services in response to scientific requirements and architectural complexities; and<br>4. Provide scalable interactive resource management—navigate data and metadata at multiple levels, provide architecture-aware data integration, analysis and visualization tools.<br>We will engage closely with DOE, NASA, and NOAA science groups working at leading edge compute facilities. These engagements—in domains such as biology, climate, and hydrology—will allow us to advance disciplinary science goals and inform our development of technologies that can accelerate discovery across DOE and other U.S. agencies more broadly. |
| **Climate Data Management System (CDMS)**<br><br>***Denis Nadeau (DOE/LLNL, nadeau1@llnl.gov)*** | |
| **Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT)**<br><br>***Aashish Chaudhary (Kitware, aashish.chaudhary@kitware.com)*** | |
| **CDATWeb**<br><br>***Matthew Harris (DOE/LLNL, harris112@llnl.gov)*** | |
| **NetCDF/HDF5**<br><br>***Ben Evans (NCI/ANU, Ben.Evans@anu.edu.au)*** | |
| **PROV**<br><br>***Ben Evans (NCI/ANU, Ben.Evans@anu.edu.au)*** | |
| **Climate Forecast (CF)**<br><br>***Karl Taylor (DOE/LLNL, taylor13@llnl.gov)*** | |
| **ES-DOC**<br><br>***Mark Greenslade (ENES/IPSL, momipsl@ipsl.jussieu.fr)*** | |
| **DRS**<br><br>***Ag Stephens (ENES/BADC, ag.stephens@stfc.ac.uk)***<br>***Guillaume Levavasseur (ENES/IPSL, glipsl@ipsl.jussieu.fr)*** | |

| | |
|---|---|
| **Data Citation Service**<br><br>***Martina Stockhause (ENES/DKRZ, stockhause@dkrz.de)***<br>***Katharina Berger (ENES/DKRZ, berger@dkrz.de)***<br><br>***(Question: Martina sent in two abstracts, is this the right one?)*** | The review of the CMIP6 data citation procedure resulted in the requirement of a citation possibility prior to the long-term archival of the data at the IPCC DDC (Data Distribution Centre) hosted at DKRZ.<br>A concept for a new citation module was developed and described in the WIP paper "CMIP6 Data Citation and Long-Term Archival" [1]. It consists of a repository, a GUI for data ingest and an API for data access [2]. This new component has to be integrated in the overall CMIP6 infrastructure. Several connections exist to the long-term archival, the ESGF development (esp. CoG portal, data versioning and data replication), and the other components providing additional information on the data (e.g., CIM documents) quality information and other annotations.<br>The poster gives a short summary of the citation concept for CMIP6 and the relations between the concept for data citation and data long-term archival to other CMIP6 infrastructure components. The focus will lie on the implementation of the citation concept and the technical integration of the citation module into the ESGF infrastructure, which is a part of the ESGF-QCWT efforts [3].<br>References:<br>[1] M. Stockhause, F. Toussaint, M. Lautenschlager (2015): CMIP6 Data Citation and LTA. Submitted as WIP white paper. http://www.earthsystemcog.org/projects/wip/resources<br>[2] Data Citation Service: http://cmip6cite.wdc-climate.de<br>[3] ESGF Quality Control Working Team (ESGF-QCWT). https://acme-climate.atlassian.net/wiki/display/ESGF/ESGF-QCWT+Charge |
| **PCMDI's Metrics Package**<br><br>***Paul Durack (DOE/LLNL, durack1@llnl.gov)***<br>***Peter Gleckler (DOE/LLNL, gleckler1@llnl.gov)*** | |
| **UV-CDAT Metrics (UVCMetrics)**<br><br>***Jeff Painter (DOE/LLNL, painter1@llnl.gov)***<br>***Brian Smith (DOE/ORNL, smithbe@ornl.gov)*** | |
| **Climate Variability and Diagnostics Package**<br><br>***Eric Nienhouse (NSF/NCAR, ejn@ucar.edu)*** | |
| **ESMValTool**<br><br>***Stephan Kindermann (ENES/DKRZ, kindermann@dkrz.de)*** | |
| **CMIP6 Errata as a New ESGF Service**<br><br>***Guillaume Levavasseur (ENES/IPSL, glipsl@ipsl.jussieu.fr)***<br>***Sébastien Denvil (ENES/IPSL, sebastien.denvil@ipsl.jussieu.fr)***<br><br>***(Question: is this a poster or the ESGF – QCWT)*** | Because of the experimental protocol inherent complexity of project like CMIP5 or CMIP6 it becomes important to record and to track reasons for datasets version changes. During CMIP5 it was impossible for scientists making use of data sets hosted by ESGF to know easily whether they were using a dataset having a known problem and whether this known problem were corrected by a newer version. Also very difficult was to have access to a description of this issue.<br>To move towards a better errata system is motivated by key requirements:<br>• Provide timely information about newly discovered issues. Because errors cannot entirely be eliminated, we should implement a centralized public interface to data providers, so that they can directly describe problems when they are discovered.<br>• Provide known issues information prior to download. The user has to be informed of known issues before downloading through the ESGF search interface.<br>• Enable users to interrogate a database to determine whether modifications and/or corrections have been applied to data they have downloaded. This service could rely on unique file identifiers so end users can discover whether files of interest to them 1) have been affected by known issues, 2) have been withdrawn, and/or 3) have been modified or corrected.<br>• Develop as part of the errata system a capability to notify end users of updates to files of interest to them.<br>The Quality Control Working Team aims to define and to establish a stable and coordinated procedure to collect and give access to errata information related to data sets hosted by ESGF. |
| **Enabling in-situ analytics in the Community Earth System Model via a Functional Partitioning Framework**<br><br>***Valentine Anantharaj (DOE/ORNL,*** | Efficient resource utilization is critical for improved end-to-end computing and workflow of scientific applications. Heterogeneous node architectures, such as the GPU-enabled Titan supercomputer at the Oak Ridge Leadership Computing Facility (OLCF), present us with further challenges. In many HPC applications on Titan, the accelerators are the primary compute engines while the CPUs orchestrate the offloading of work onto the accelerators, and moving the output back to the main memory. On the other hand, applications that do not exploit GPUs, the CPU usage is dominant while the GPUs idle.<br>We utilized Heterogeneous Functional Partitioning (HFP) runtime framework that can optimize usage of |

| | |
|---|---|
| ***anantharajvg@ornl.gov)*** | resources on a compute node to expedite an application's end-to-end workflow. This approach is different from existing techniques for in-situ analyses in that it provides a framework for on-the-fly analysis on-node by dynamically exploiting under-utilized resources therein.<br>We have implemented in the Community Earth System Model (CESM) a new concurrent diagnostic processing capability enabled by the HFP framework. Various single variate statistics, such as means and distributions, are computed in-situ by launching HFP tasks on the GPU via the node local HFP daemon. Since our current configuration of CESM does not use GPU resources heavily, we can move these tasks to GPU using the HFP framework. Each rank running the atmospheric model in CESM pushes the variables of of interest via HFP function calls to the HFP daemon. This node local daemon is responsible for receiving the data from main program and launching the designated analytics tasks on the GPU.<br>We have implemented these analytics tasks in C and use OpenACC directives to enable GPU acceleration. This methodology is also advantageous while executing GPU-enabled configurations of CESM when the CPUs will be idle during portions of the runtime. In our implementation results, we demonstrate that it is more efficient to use HFP framework to offload the tasks to GPUs instead of doing it in the main application. We observe increased resource utilization and overall productivity in this approach by using HFP framework for end-to-end workflow. |