## 2017 7th Annual ESGF F2F Conference Agenda

2017 Earth System Grid Federation (ESGF) Face-to-Face Conference (San Francisco, California, USA)

## **Registration:**

## **Conference venue:**

San Francisco Sheraton at Fisherman's Wharf, 2500 Mason Street, San Francisco, CA 94133, USA

**Remote participation:** Indicate on registration form that you wish to participate remotely and an email will be sent with information you need to join Webinar.

**Date:** Dec 4, 2017 to Dec 8, 2017—8:00 a.m. - 6:00 p.m. PST **Webinar Logistics:** 

- 2017 ESGF F2F Sheraton Fisherman's Wharf, San Francisco, California
- Tuesday, Wednesday, Thursday, Friday (December 5, 2017 through December 8, 2017)
- 7:30 a.m. | Pacific Standard Time (San Francisco, GMT-08:00) | 11 hours
- US TOLL: +1-415-655-0001
- Global call-in numbers: <a href="https://llnl.webex.com/llnl/globalcallin.php?serviceType=MC&ED=616508797&t">https://llnl.webex.com/llnl/globalcallin.php?serviceType=MC&ED=616508797&t</a> ollFree=0
- Meeting number (access code): 807 447 408
- Meeting password: esgf

Note. After logging on, please send your full name and affiliation to Angela (jefferson9@llnl.gov) for remote conference registration.

Time	Торіс			
Monday, December 4, 20	Monday, December 4, 2017			
2:00 p.m. – 4:00 p.m.	Pre-conference registration: Sheraton; Presidio Foyer			
5:00 p.m. – 6:00 p.m.	Social Activity: Meet and Greet (NO HOST)			
	Sheraton-Fisherman's Wharf – Restaurant/Bar			
Tuesday, December 5, 2017				
7:30 a.m. – 8:30 a.m.	Registration: Sheraton; Presidio Foyer			
8:00 a.m. – 8:30 a.m.	Coffee/tea reception and meet & greet: Sheraton, Presidio Foyer			

Time		Topic	
8:30 a.m. – 8:35 a.m.	Luce Cinquini	(NASA/IDI) & Angala Jaffarson (DOE/LINI)	
8:50 a.m. – 8:55 a.m.	— Welcome, Saf	(NASA/JPL) & Angela Jefferson (DOE/LLNL)  fety and Logistics	
8:35 a.m. – 8:40 a.m.	DOE opening comments — (Justin Hnilo, U.S. DOE's Office of Biological and Environmental		
		Program Manager for Data Management)	
	<ul> <li>Includ</li> </ul>	es welcome, safety, introduction, conference charge, and agenda overview	
8:40 a.m. – 9:00 a.m.		th System Grid Federation (ESGF) (Luca Cinquini—NASA/JPL)	
	How conf	erence attendees contribute to the conference's final report	
	• Framing o	of the 2017 7th Annual ESGF F2F Conference	
	Science D	rivers: Project Requirements and Feedback (A note from our sponsors)	
9:00 a.m. – 12:00 noon	Science Driver		
(3 hours)		sion Lead — Balaji	
	9:00 a.m. –	Karl Taylor and V. Balaji—Coupled Model Intercomparison Project,	
	9:30 a.m.	phase 6 (CMIP6) and the Working Group on Coupled Modeling	
		Infrastructure Panel (WIP)	
	9:35 a.m. – 10:05 a.m.	Peter Gleckler, Duane Waliser, Denis Nadeau, Robert Ferraro, Karl	
	10:03 a.m.	Taylor, Luca Cinquini, Paul Durack—An Update on Observations for Model Intercomparison Project (obs4MIPs) from an ESGF perspective:	
		progress plans and challenges	
	10:10 a.m. –	S. Denvil, M. Lautenshlager, S. Fiore, F. Guglielmo, S. Kinderman, M.	
	10:40 a.m.	Kolax, C. Pagé, W. Som de Cerff, the ENES Data Task Force —	
		Copernicus and H2020 Programme: Science Drive Overview	
	10:40 a.m. –	Break	
	10:55 a.m.		
	10:55 a.m. –	Jerry Potter, Laura Carriere, Judy Hertz—Collaborative REAnalysis	
	11:25 a.m.	Technical Environment Intercomparison Project (CREATE-IP)	
	11:30 a.m. –	Dean N. Williams, Dave Bader, Renata McCoy—Energy Exascale Earth	
	12:00 noon	System Model (E3SM) Workflow	
	Questions for	presenters to answer during their presentations	
	What are	the key things that are difficult to do today and are impeding scientific	
		or productivity and the sharing of data?	
		e key development effort that you see are needed for the future success of your	
	projects?	your timeline for data production and distribution from climate model and	
		ions, high-performance computer, network, and storage facilities needs and	
	investme		
	What is to	the estimated size of your distributed archive?	
		e your common developments, sharing of expertise, and accelerated	
	developr		
		e the administrative/sponsor requirements that arise from your project	
		y, metrics collection and reporting, persistent and digital object identifiers, data, user publication [i.e., long-tail publication], etc.)?	
		e your expected strategic roadmaps for the ESGF's short-term (1 to 3 years),	
		n (3 to 5 years), and long-term (5 to 10 years) development efforts?	
		e known use cases and workflows to help describe your ESGF future needs?	
	Homework ass		
		he conference adjourns, convert all known science drivers to use cases for evelopment.	
12:00 noon – 1:30 p.m.	ESUF (I	Lunch	
1:30 p.m. – 3:30 p.m.	Science Driver	Town Hall Discussion	
(2 hours)		sion Lead — Ben Evans	

Time		Торіс			
	Micha	Hall Panel: (Karl Taylor, V. Balaji, Peter Gleciel Lautenschlager, Jerry Potter, Renata McCogions to prepare for science driver presentat What is working, and what is not working? What are the key challenges to your programs What data services would address the identification what exists already today? What do we still need from ESGF? What are the key characteristics that these ser integrated, easy to customize, etc.)? What are the key impediments (on the data prodelivering these services? Which services should be developed with the measurable impact on science/programs?	ion and discussion concerning big data ed challenges? vices need to have to ovider/service provider	a challenge o be succe der side) i	es? ssful (i.e. n
3:30 p.m. – 3:45 p.m.		Break			
3:45 p.m. – 4:15 p.m.		Awards Ceremony			
4:15 p.m. – 6:00 p.m. (1 hour & 45 min)		r and Live Demonstration Session in Discussion Lead — Ben Evans			
	No.	Title	Name	Poster	Demo
	1	The Earth Data Analytics Services (EDAS)	Thomas Maxwell	Yes	Yes
		Framework	Dan Duffy		
	2	PAVICS: A platform for the Analysis and	D. Huard	Yes	Yes
		Visualization of Climate Science – toward inter-operable multidisciplinary workflows	T. Landry D. Byrns B. Gauvin-St- Denis		
	3	OGC Testbed-13 Earth Observation Clouds	T. Landry D. Byrns	Yes	No
	4	Using the ESGF CWT-API in the context of	Christian Pagé	Yes	Yes
		the EUDAT-EGI e-infrastructure and the ENES climate4impact platform	Xavier Pivan Asela Rajapakse Wim Som de Cerff		
			Maarten Plieger		
			Ernst de Vreede Alessandro		
			Spinuso		
			Lars Barring		
			Antonio Cofino Alessandro d'Anca		
			Sandro Fiore		
	5	Managing growth and complexity -	Phil Kershaw	Yes	No
		technologies to meet the challenges of	Jonathan Churchill		
		operating data, services and infrastructure at scale	Alan Iwi Bryan Lawrence		
		at scare	Neil Massey		
			Sam Pepler		
			Matt Pritchard Matt Pryor		
			Ag Stephens		
	6	Ophidia: an interoperable 'big data'	Sandro Fiore	Yes	Yes
		framework for climate change analytics	Charles Doutriaux		
		experiments	Cosimo Palazzo		
			Alessandro d'Anca		

Time		Topic			
			Zeshawn Shaheen Donatello Elia Jason Boutte Valentine Anantharaj Dean N. Williams Giovanni Aloisio		
	7	Federated data usage statistics in the Earth System Grid Federation	Alessandra Nuzzo Maria Mirto Paola Nassisi Katharina Berger Torsten Rathmann Luca Cinquini Sébastien Denvil Sandro Fiore Dean N. Williams Giovanni Aloisio	Yes	Yes
	8	WPS based processing services for the Copernicus Climate Change Service (C3S)	Stephan Kindermann Carsten Ehbrecht Ag Stephens Björn Brötz Wim Som de Cerff Maarten Plieger Sébastien Denvil	Yes	Yes
	9	Diagnostics Package for the E3SM Model	Chengzhu Zhang Zeshawn Shaheen Chris Golaz Jerry Potter	Yes	Yes
	10	ESGF Errata Service	Guillaume Levavasseur Atef Ben-nasser Mark A. Greenslade	No	Yes
	11	DREAM Data Services for Biological Data and Beyond	Sasha Ames Luca Cinquini Dean N. Williams	Yes	Yes
	12	Community Data Analysis Tools	Charles Doutriaux Denis Nadeau Dan Lipsa Dean N. Williams Aashish Chaudhary	Yes	Yes
	13	Visual Community Data Analysis Tools (vCDAT)	Matthew Harris Dan Lipsa James Crean Matthew Ma Charles Doutriaux Dean N. Williams Aashish Chaudhary	Yes	Yes
	14	Integrating ES-DOC with the ESG Publisher	Alan Iwi David Hassell Mark A. Greenslade Ag Stephens	Yes	Yes
	15	Compute Working Team End-User Application Programming Interface	Jason Boutte Charles Doutriaux	Yes	Yes
	16	A compliance-checking framework for	Ag Stephens	No	Yes

Time	Торіс				
		CMIP7	Antony Wilson Guillaume Levavasseur		
	17	Google Earth Engine and Project Jupyter	Tyler Erickson	No	Yes
	18	New Approach to Evaluate Large-scale Variability in CMIP models	Ji-Woo Lee, Kenneth R. Sperber, Peter J. Gleckler, Celine W. Bonfils, Karl E. Taylor, Charles Doutriaux	Yes	Yes
	•	What is working and what is not working? What are the key challenges to your application the ESGF infrastructure? How does your application/services integrate What do you still need from ESGF for softway What are the key impediments in delivering y installation, customization, etc.)? What are the key characteristics or functional the community within the ESGF infrastructur Which services or functions are your applicated what would be their measurable impact on softhe future)?	on concerning big da into ESGF? are integration? your application/servities that your application? e? ion's highest develop	ces in ESO	GF (i.e.
6:00 p.m.	Adjourn Day 1				
Wednesday, December (	6, 2017				
8:00 a.m. – 8:30 a.m.	Coffee	/tea reception and meet & greet: Sheraton; Pr	esidio Foyer		
8:30 a.m. – 9:45 a.m. (1 hour & 15 min)	8:30 8:45 8:45 9:30 9:45 Questi	a.m. – Group discussion a.m. – Conclusion recap a.m. – Conclusion recap a.m. – Conclusion recap ions for the ICNWG (i.e., network) plenary ICNWG network software and hardware integrities ICNWG network preparation services and too Automated replication network requirements projects) ICNWG network security requirements ICNWG dashboard integration into ESGF das	rt, Lukasz Lacinski, S Kindermann—Prese  discussion gration requirements ols (e.g., perfSONAR for ESGF (i.e., CMII	ntation on for Tier 1	ndermann data and Tier 2
		Resource discovery and allocation services Identify key gaps, identify benefitting commu	inities, and prioritize	ICNWG 1	future work

Time		Торіс
9:45 a.m. – 10:30 a.m.	Google Cloud	
	Presentation b	y Karan Bhatia (Google) and group discussion
	9:45 a.m. –	Karan Bhatia - presentation
	10:15 a.m.	
	10:15 a.m. –	Group discussion
10:30 a.m. – 10:45 a.m.	10:30 a.m.	Break
10:45 a.m. – 12:00 noon (1 hour & 15 minutes)		Data Analytics Working Team Plenary Discussion sion Lead and Presenters — Charles Doutriaux and Daniel Duffy
	10:45 a.m. – 11:00 a.m.	Charles Doutriaux and Daniel Duffy—Presentation on server-side computing progress
	11:00 a.m. – 11:15 a.m.	Cameron Christensen, Giorgio Scorzelli, Peer-Timo Bremer, Shusen Liu, Ji-Woo Lee, Brian Summa, Valerio Pascucci - Interactive Analysis and Visualization of Arbitrarily Large, Disparately Located Climate Data Ensembles Using a Progressive Runtime Server, On-Demand Data Conversion, and an Embedded Domain Specific Language Suitable for Incremental Computation
	11:15 a.m. – 12:00 noon	Group Discussion and Conclusion recap
	<ul> <li>Define a</li> <li>Data ana</li> <li>Performa</li> <li>Advance</li> <li>Provenar</li> <li>Automati</li> <li>Resource</li> <li>Identify I</li> </ul>	server-side computing scalable compute resource (clusters and HPCs) for ESGF data analysis lytical and visualization capabilities and services unce of model execution d networks as easy-to-use community resources (i.e., resource management) unce and workflow ion of steps for the computational work environment e management, installation and customer support key gaps, identify benefitting communities, and prioritize next steps services when multiple data sets are not co-located (future work)
12:00 noon – 1:30 p.m.	<b>T.</b> 111 <b>T</b> 111	Lunch
1:30 p.m. – 2:40 p.m. (1 hour & 10 minutes)		ement Access Working Team Plenary Discussion sion Lead — Philip Kershaw and Lukasz Lacinski
	1:30 p.m. – 1:45 p.m.	Philp Kershaw and Lukasz Lacinski—Presentation on authentication and authorization and IdEA progress
	1:45 p.m. – 2:30 p.m.	Group discussion
	2:30 p.m. – 2:40 p.m.	Conclusion recap
	<ul> <li>What too and how etc.)?</li> <li>What is r software</li> <li>What ser authoriza</li> <li>What lev</li> </ul>	authentication and authorization also have been identified for authentication and authorization (i.e., OAuth 2) well will they integrate with other projects (i.e., Copernicus, NASA DAACs, needed for authentication and authorization integration with the ESGF stack installation (i.e., address key needs)? vices must be made available today and in the future for authentication and attion? el of support would be expected from the science community? we want to assess the maturity and capability of authentication and

Time	Торіс		
	aurthoization (e.g., benchmarks or crowdsourcing)?		
	What are the future efforts to be expected from ESGF-IdEA?		
2:40 p.m. – 3:55 p.m. (1 hour & 15 minutes)	Status Update and Future Planning for ESGF User Interface, Search, and Dashboard Working Teams Plenary Discussion		
	Session Discussion Lead — Luca Cinquini, Guillaume Levavasseur, and Alessandra Nuzzo		
	2:40 p.m. – Luca Cinquini, Guillaume Levavasseur, and Alessandra Nuzzo—Status update and future planning for the ESGF UI, Search, and Dashboard Working Group		
	2:55 p.m. – Group discussion 3:30 p.m.		
	3:30 p.m. 3:45 p.m.  Break		
	3:45 p.m. – Conclusion recap 3:55 p.m.		
	<ul> <li>Questions</li> <li>Do you have any plan for engaging the user community to provide ongoing feedback for the user interface?</li> <li>How do you expect the search services to scale to support new data holdings in the next 5 years?</li> <li>Do you have any plans for federating the search services with other sites/agencies/institutions?</li> <li>How do you validate the metrics obtained from the dashboard, both for a single node,</li> </ul>		
	and across the whole federation?		
3:55 p.m. – 5:30 p.m. (1 hour & 15 minutes)	Installation and Software Security Working Team Plenary Discussion Session Discussion Leads — William Hill, Prashanth Dwarakanath, Luca Cinquini, and George Rumney		
	3:55 p.m. – William Hill—Presentation on Software Installation 4:05 p.m.		
	4:05 p.m. – Luca Cinquini—Presentation on Software Container (i.e., Docker) 4:15 p.m.		
	4:15 p.m. – George Rumney—Presentation of Software Security 4:25 p.m.		
	4:25 p.m. — Group discussion 5:20 p.m. — Conclusion recap 5:30 p.m.		
	<ul> <li>Questions</li> <li>How close are you to have an operation version of the Docker/Cloud ESGF?</li> <li>Which services or functionality are still missing from this architecture?</li> <li>How do you plan to address security risks with this architecture?</li> <li>Is there a plan for migrating an operational system from the current shell-based installer to Docker/Cloud?</li> </ul>		

Time	Торіс		
Thursday, December 7,	2017		
8:00 a.m. – 8:30 a.m.	Coffee/tea reception and meet & greet: Sheraton; Presidio Foyer		
8:30 a.m. – 9:45 a.m. (1 hour & 15 minutes)	Publication, Quality Control, Metadata, and Provenance Capture Working Team Plenary Discussion Session Discussion Leads — Sasha Ames and Heinz-Dieter Hollweg		
	8:30 a.m. – Sasha Ames—Presentation on Publication Progress 8:40 a.m.		
	8:40 a.m. – Heinz-Dieter Hollweg—Presentation on Quality Control Progress 8:50 a.m.		
	8:50 a.m. – Bibi Raju - Provenance Data harvest and Scientific Results Reproducibility 9:00 a.m.		
	9:00 a.m. – Martina Stockhause (or Stephan K.) – Errata Service 9:10 a.m.		
	9:10 a.m. – Group Discussion and Conclusion recap 9:45 a.m.		
9:45 a.m. – 10:45 a.m. (1 hour)	<ul> <li>Questions for publications, QC, metadata, and provenance capture plenary discussion</li> <li>Data integration and advanced metadata capabilities</li> <li>Data and metadata collection and sharing capabilities for possible provenance</li> <li>Data Quality and ancillary information</li> <li>Data preparation services and tools</li> <li>Authentication and security</li> <li>Local and remote publication services</li> <li>What are the key challenges that scientists encounter?</li> <li>What capabilities would address the identified challenges?</li> <li>What do we still need?</li> <li>What are the impediments for ESGF node providers and software developers to provide these missing capabilities?</li> <li>Which requirements need to be addressed with the highest priority and what would be their measurable impact on science?</li> <li>Machine Learning Plenary Discussion</li> <li>Session Discussion Lead — Sookyung Kim, Sandro Fiore</li> </ul>		
	9:45 a.m. – Sookyung Kim—Presentation on Community Machine Learning 9:55 a.m.  9:55 a.m. – S. Denvil, M. Lautenschlager, S. Fiore, F. Guglielmo, M. Juckes, S. Kindermann, M. Kolax, C. Pagé, W. Som de Cerff. The ENES Data Task Force — Copernicus and H2020 Program: Machine Learning and Big Data Needs and Overview  10:05 a.m. – Tom Landry - Imagery, text and geospatial Machine Learning applications in Montreal's booming ML landscape		
	10:15 a.m. Group Discussion 10:45 a.m.  Questions for the machine learning plenary discussion  • What problems machine learning and deep learning methodologies can solve in climate domain?  • What can it not solve?		

Time	Торіс	
10:45 a.m. – 11:00 a.m.	<ul> <li>What is the recent metrics in deep learning which can applied to climate data?</li> <li>What exist already in climate community using artificial intelligence?</li> <li>What is the highest priority problem using machine learning in climate community?</li> <li>What are the key challenges to ESGF implementing machine learning algorithms?</li> <li>How can we solve these challenges with respect to programs?</li> <li>How can we solve data labeling and scalability issue?</li> </ul> Break	
11:00 a.m. – 12:00 noon	Diagnostics Plenary Discussion	
(1 hour)	Session Discussion Lead — Zeshawn Shaheen, Tom Landry, others  11:00 a.m. — Zeshawn Shaheen—Presentation on the Community Diagnostics Package  11:10 a.m. — S. Denvil, M. Lautenschlager, S. Fiore, F. Guglielmo, M. Juckes, S. Kindermann, M. Kolax, C. Pagé, W. Som de Cerff. The ENES Data Task Force — Copernicus and H2020 Program: Diagnostics Needs and Overview  11:20 a.m. — Tom Landry—Presentation on Canada Diagnostics  11:30 a.m. — Group discussion  11:50 a.m. — Conclusion recap  Questions for the diagnostics plenary discussion  • What are the key diagnostics challenges that scientists encounter?  • What diagnostics exists already today?  • What diagnostics are still need?  • What are the diagnostics impediments for resource providers (i.e., hardware) and software developers to provide these missing capabilities?  • Which diagnostics requirements need to be addressed with the highest priority and	
	what would be their measurable impact on science?	
12:00 noon – 1:30 p.m.	Lunch	
1:30 p.m. – 3:00 p.m. (1 hour & 30 minutes)	CMIP6 Data Node Operations Team (CDNOT) Plenary Discussion  Session Discussion Lead — Sébastien Denvil  1:30 p.m. — Sébastien Denvil—What is CDNOT and what is it we want to achieve  1:50 p.m. — Katharina Berger (DKRZ), Stephan Kindermann (DKRZ) — Feedback from an insider.	
	2:10 p.m. – Serguei Nikonov (Princeton University), V.Balaji (Princeton University), Aparna Radhakrishnan (Engility Corporation), Hans Vahlenkamp (Princeton University) — CMIP6 and ESGF reciprocation: Case study using GFDL preparation to CMIP6  2:10 p.m. – Kim Serradell (Barcelona Supercomputing Center) — First steps with ESGF.	
	2:20 p.m. – Group Discussion and Conclusion recap 3:00 p.m. — Group Discussion and Conclusion recap  Questions for the CDNOT plenary discussion  What are the ESGF services and tools that are needed for CDNOT to be successful  Should CDNOT's mode of operation be made more widely accessible to other project	ects

Time	Торіс	
	<ul><li>and the community?</li><li>What is the distinction between CDNOT and ESGF?</li></ul>	
3:00 p.m. – 4:10 p.m. (1 hour & 10 minutes)	Node Manager and Tracking / Feedback Notification Plenary Discussion Session Discussion Lead — Sasha Ames and Tobias Weigel	
	3:00 p.m. – Sasha Ames—Presentation on the Node Manager 3:10 p.m.	
	3:10 p.m. – Tobias Weigel— Presentation on PID Services and Tracking/Feedback 3:20 p.m.	
	3:20 p.m. – Group discussion and Conclusion recap 4:10 p.m.	
	<ul> <li>Questions for the node manager and notification plenary discussion</li> <li>What are the key challenges for the node manager and notification?</li> <li>What services would address the identified challenges?</li> <li>What exists already today? What do we still need?</li> <li>What are the key characteristics that these services need to have to be successful (i.e. integrated, easy to customize etc.)?</li> <li>What are the key impediments (on the data provider / service provider side) in delivering these services?</li> <li>Which services should be developed with the highest priority and what would be their measurable impact on science?</li> </ul>	
4:10 p.m 4:30 p.m.	Break	
4:30 p.m. – 5:30 p.m. (1 hour)	User Support and Documentation Plenary Discussion  Session Discussion Lead — Matthew Harris  4:30 p.m. — Matthew Harris —Presentation on the User Support for the Community  4:40 p.m. — Group discussion  5:20 p.m. — Conclusion recap  5:20 p.m. — Conclusion recap  Questions  • What level of support and documentation are needed for ESGF services, tools and the community?  • What do support and documentation do data provider and users want to see from ESGF?	
	<ul> <li>What type of support and documentation is there for ESGF (i.e., FAQs, Jupyter Notebook, online tutorials, presentations)?</li> </ul>	
	<ul> <li>Where are the support tools and documentation located?</li> <li>What can we expect in the future in terms for user support and documentation?</li> </ul>	
5:30 p.m.	Adjourn Day 3	
Fuiday Dagambar 9 201		
Friday, December 8, 201		
8:00 a.m. – 8:30 a.m.	Coffee/tea reception and meet & greet: Sheraton; Presidio Foyer	
8:30 a.m. – 10:00 a.m.	<ul> <li>ESGF Executive Committee Breakout Meeting: Sheraton; Lombard Room</li> <li>Discuss of the construction of the annual conference report</li> <li>Discuss meeting location and time of the next ESGF F2F conference</li> </ul>	

Time	Торіс
	Discuss strategic and implementation documents
	Working Teams Meeting
	All working teams discuss conference findings for their area of annual reporting
10:00 a.m. – 10:15 a.m.	Break
10:15 a.m. – 12:00 noon	ESGF Development Teams Report Back on Conference Findings: Sheraton; Presidio
	Ballroom
	Session Discussion Lead — Tom Landry
	ESGF Team Leads findings on conference feedback
	Prioritize the feedback
	Open discussion
12:00 noon	Adjourn Day 4
12:00 noon – 1:30 p.m.	Lunch
1:30 p.m. – 5:30 p.m.	General Code Sprint (optional): Sheraton; Lombard Room
·	Working Teams and Leads
5:30 p.m.	Conference Adjourn Day 4
	Concludes the 7 <sup>th</sup> Annual ESGF F2F Conference