



National Center for Atmospheric Research

Climate and Global Dynamics Division • Climate Change Research Section

1850 Table Mesa Drive • Boulder, CO 80305

Tel: 303-497-1321 • Fax: 303-497-1715 • Email: wmw@ucar.edu

Web: <http://www.cgd.ucar.edu/CCR/warren/>

NCAR

20 March 2013

R&D 100 Awards
100 Enterprise Drive
Suite 600, Box 912
Rockaway, NJ 07866-0912

Dear R&D 100 Awards:

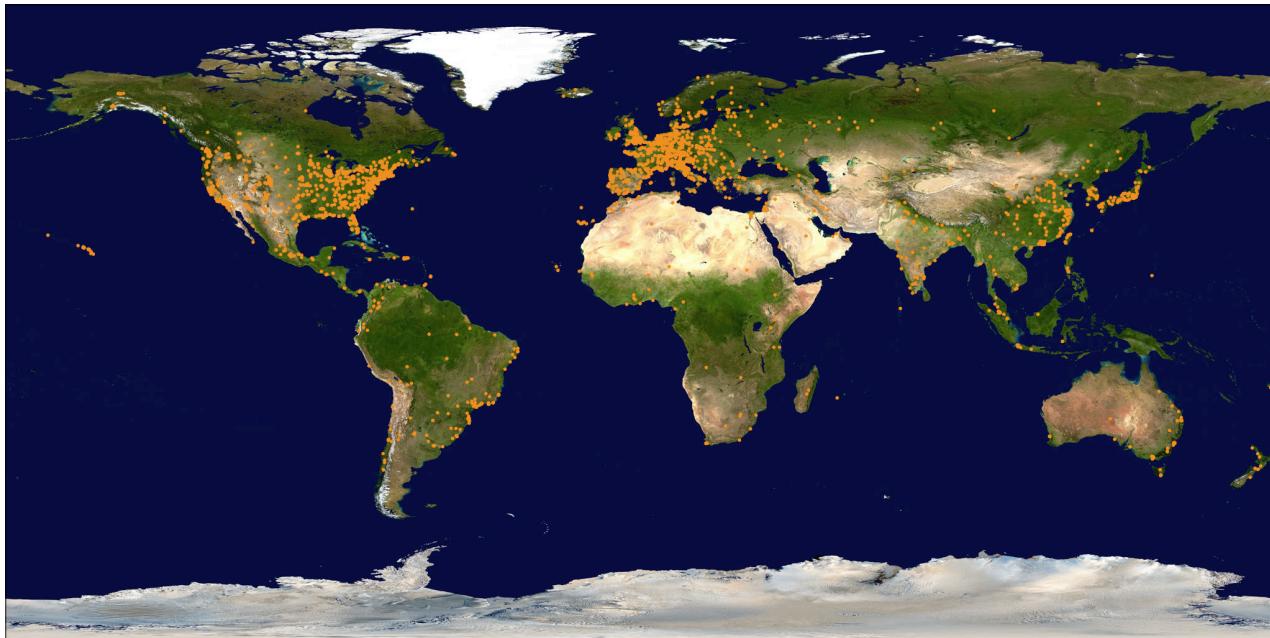
We are writing to support the selection of the Earth System Grid Federation (ESGF) as an R&D 100 awardee for 2013.

The impact of the ESGF on the advancement of climate science research and our understanding of climate change cannot be overstated. Our best estimates of future climate originate from the model projections performed by the world's leading modeling centers. Because no single model is demonstrably superior to all others, a result of both the extreme complexity of the climate system and scientific uncertainties in model formulations, the ability to examine the results of many models has proved to be crucial in improving of scientific understanding of future climate change.

Model simulations from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) will underpin a major part of the upcoming Fifth Scientific Assessment by the Intergovernmental Panel on Climate Change. Through ESGF, thousands of researchers can access the CMIP5 database, which now exceeds 50 Petabytes (and will ultimately reach 100 Petabytes) distributed over the modeling center archives on 4 continents. The ability of thousands of researchers to easily access, search, discover and obtain the data from a large, multi-model archive of climate simulations is truly transformational. The ESGF system provides secure access to a scientific gold mine of information.

In the future, the ESGF system will enable major advances in our understanding of the past and future climate change. It will also allow the scientific community to provide the more detailed and more reliable forecasts of climate change that governments and society require for developing successful adaption and mitigation strategies.

To highlight the success of the ESGF, we show a world map of the current distribution of users.



Climate scientists anywhere in the world can access the ESGF research centers and climate data products. Each dot represents a registered site, which total over 25,000.

The National Center for Atmospheric Research is operated by the University Corporation for Atmospheric Research under sponsorship of the National Science Foundation.

An Equal Opportunity/Affirmative Action Employer

Because of these impacts, we firmly believe that the ESGF developers are deserving of an R&D 100 award.

Sincerely,



Warren Washington,
Member, National Academy of Engineering,
Former Chair, National Science Board
Senior Scientist, National Center for Atmospheric Science



Benjamin Santer,
Member, National Academy of Science
Distinguished Member of the Technical Staff
Lawrence Livermore National Laboratory



Ghassem Asrar,
Executive Director,
World Climate Research Programme



V. Ramaswamy
Director, NOAA Geophysical Fluid Dynamics Laboratory



Robert Dickinson,
Member, National Academies of Science and Engineering
Professor of Geological Sciences, University of Texas



Antonio J. Busalacchi
Professor and Director, U. of MD Earth System Science Interdisciplinary Center
Chairman of the Joint Scientific Committee for the World Climate Research Programme
Chairman of the National Research Council Board on Atmospheric Sciences and Climate



NATIONAL SCIENCE FOUNDATION

4201 Wilson Boulevard
Arlington, VA 22230

DIVISION OF ATMOSPHERIC AND GEOSPACE SCIENCES

To Whosever it may concern

March 22, 2013

I was a program director at DOE Office of Science, Office of Biological and Environmental Research from 2004-2010. During that period, with the Office of Advanced Scientific Computing and Research, our office co-funded a project called the Earth System Grid Center for Enabling Technologies (ESG CET). The lead PI for this project was Dean Williams of Program for Climate Model Diagnosis and Intercomparison (PCMDI) located at Lawrence Livermore National Laboratory.

The goal of this project was to build a “**science gateway**” to climate resources that provides data, information, models, analysis, and visualization tools and computational capabilities for management and analysis. ESG-CET was a multi-institutional effort across various DOE national laboratories, National Center for Atmospheric Research, NOAA/Pacific Marine Environmental Laboratory), and University of Southern California, working to maximize the accessibility of climate simulation data by the international research community.

ESG was responsible for the data management of hundreds of terabytes of climate data ranging from high-resolution modeling, grand challenge-scale computations on leadership computing systems, regional climate modeling, coupled climate/carbon cycle modeling, land/biosphere modeling, atmospheric chemistry modeling, detection and attribution of climate change, and model intercomparison projects.^{1,2}

The ESG-CET subsequently led to Earth System Grid Federation (ESGF³), a project being led by Dean Williams. To date, ESGF has made available over 60 large-scale CMIP5 simulation model runs (more than 1.8 petabytes of data) from 25 worldwide climate research centers spanning 21 countries. ESGF also provides access to 18 highly visible national and international climate data products, with more on the way. Climate scientists anywhere in the world can access the ESGF research centers and climate data products. There are over 25,000 registered sites, with currently more than 25 portals are in use, including one at PCMDI, and over 2.5 Petabytes downloaded by the community.

At NSF, in FY 11, we made ~16 small grants to university researchers for analysis of the Coupled Model Intercomparison Project (CMIP) simulations. They all accessed the data at PCMDI through the use of ESGF. In March 2012, a workshop

on *Coupled Model Intercomparison Project Phase 5 (CMIP5) Model Analysis* was held where PIs presented their results. The website for the Workshop is <http://cmip5.wcrp-climate.org/workshop/sponsors.shtml>. A list of publications resulting from the CMIP project may be found at <http://cmip.llnl.gov/cmip5/publications/allpublications>

Anjuli Bamzai, Ph.D.

Program Director Climate and Large-Scale Dynamics
NSF Division of Atmospheric and Geospace Sciences
Directorate of Geosciences
National Science Foundation

References:

1. [Data Management and Analysis for the Earth System Grid](#)
Overview of climate data management challenges and server-side analysis in ESG. D. N. Williams, R. Ananthakrishnan, D. E. Bernholdt, S. Bharathi, D. Brown, M. Chen, A. L. Chervenak, L. Cinquini, R. Drach, I. T. Foster, P. Fox, S. Hankin, V. E. Henson, P. Jones, D. E. Middleton, J. Schwidder, R. Schweitzer, R. Schuler, A Shoshani, F. Siebenlist, A. Sim, W. G. Strand, N. Wilhelmi, M. Su, “Data Management and Analysis for the Earth System Grid”, in the Journal of Physics: Conference Series, SciDAC ’08 conference proceedings, volume 125. ([Citation:] D N Williams et al 2008 J. Phys.: Conf. Ser. 125 012072)
2. [The Earth System Grid: Enabling Access to Multi-Model Climate Simulation Data](#). D N Williams, R Ananthakrishnan, D E Bernholdt, S Bharathi, D Brown, M Chen, A L Chervenak, L Cinquini, R Drach, I T Foster, P Fox, D Fraser, J Garcia, S Hankin, P Jones, D E Middleton, J Schwidder, R Schweitzer, R Schuler, A Shoshani, F Siebenlist, A Sim, W G Strand, M Su, N. Wilhelmi. Bulletin of the American Meteorological Society, February 2009.
3. [The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geospatial Data](#), IEEE special issue, *Future Generation Computing Systems*: Luca Cinquini, Daniel Crichton, Chris Mattmann, Gavin M. Bell, Bob Drach, Dean Williams, John Harney, Galen Shipman, Feiyi Wang, Philip Kershaw, Stephen Pascoe, Rachana Ananthakrishnan, Neill Miller, Estanislao Gonzalez, Sebastian Denvil, Mark Morgan, Sandro Fiore, Zed Pobre, Roland Schweitzer; accepted, due out in 2013.



18 March 2013

Dr. Donald E. Anderson, Jr.
NOAA Climate Program Office
Program Manager, NOAA Environmental Software Infrastructure and Interoperability (NESII) Group, Modeling, Analysis Predictions and Projections (MAPP) Program, and the National Climate Predictions and Projections (NCPP) Platform

To whom it may concern:

I am writing to support the Earth System Grid Federation (ESGF) for an R&D award. As a program manager at the National Oceanic and Atmospheric Administration, I can attest to the need for the open sharing of climate data and metadata, for research, international assessments, and studies of local impacts. Many of the groups that I manage rely on ESGF for these functions.

In particular, ESGF was the vehicle for data collection and distribution for the Fifth Coupled Model Intercomparison Project (CMIP5), a scientific foundation for the next Intergovernmental Panel on Climate Change (IPCC) assessment (AR5). With contributions from many national and international partners, ESGF delivered a system that has made it possible for the worldwide community to access this data. New pilot projects with NASA, such as obs4MIPs, promise to bridge the gap between the observational and modeling communities and greatly improve the model evaluation process.

ESGF is of particular value to my programs through its integration in the Earth System CoG collaboration environment. Through CoG, ESGF provides data support for community-based activities such as the National Climate Predictions and Projections (NCPP) Platform, which provides information about the local and regional climate, an atmospheric Dynamical Core Model Intercomparison Project (DCMIP), and a Downscaling Intercomparison Project, hosted by NCPP, to be held this summer.

NOAA programs look forward to working with the ESGF effort as it evolves and hope that the team will be recognized for their contribution to the climate community.

Thank you,

A handwritten signature in black ink that reads "Donald E. Anderson Jr."

Donald E. Anderson, Jr.



March 13, 2013

Jason Paragas, PhD
Deputy Program Director for Bio-Security

Subject: *Support Letter for ESGF*

To Whom It May Concern:

The complexity of the life sciences is being unraveled through innovation brought by high throughput technologies, systems biology and more recently biocomputational approaches. As success in these areas grows, vast and disparate data silos are evolving that were derived from scarce and rare resources. In fact, much of the research effort may be housed in stand-alone computers, email correspondence, and paper notebooks. The extraordinary atomization of our collective research efforts in the life sciences dilutes the potential treasure. Effectively the majority of the work product from the life sciences is lost. The American Academy for the Advancement of Science reported that the Federal investment in research and development is approximately 147 billion in FY10¹. The life sciences investment is approximately a third of that number. A slim fraction of that investment is transformed into an inconvenient form that collectively advances the field. We, as a life sciences community, are in essence returning copper back to the ground. The greatest opportunity in the life sciences today is not in the frontiers of science but to transform our research effort into a computable resource.

This is not a lofty idea. In fact it has been accomplished, but in the climate sciences field. The efforts of the Earth Sciences Grid Federation have transformed the community from a series of disparate efforts into a dynamic field that expends efforts in scientific frontiers and not in data impedance mismatches. Yet preserving the ability for individual teams to openly compete ideas.

Porting the principles of the Earth Sciences Grid Federation to the life sciences community will disruptively transform and accelerate our efforts. Furthermore, a life sciences application of the Earth Sciences Grid Federation will enable the

¹ AAAS Report XXXIV Research and Development FY 2010

²http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_me



vision from the recent Whitehouse directive to develop plans to increase access to the Federally funded research². Extending the principles of Earth Sciences Grid Federation into the life sciences will accelerate innovation, contain costs, and preserve the data derived from sacred research resources.

Sincerely,


Jason Paragas, PhD

²http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf



Lawrence Livermore National Laboratory

Biosciences and Biotechnology Division

March 8, 2013

To the R&D 100 Awards Committee:

As the Division Leader for the Biosciences & Biotechnology Division in the Physical and Life Sciences Directorate at Lawrence Livermore National Laboratory (LLNL), I oversee the development of new technologies that will help to improve and accelerate biomedicine and core biotechnologies.

An area of particular interests is the building of an infrastructure for the accelerated therapeutic community. The goal here would be to extend and/or develop new technologies to improve accessing of biological data for the purpose of accelerating drug design and development. To achieve this goal, an underlying computational infrastructure is needed to facilitate the exchange of biological data to allow the integration of data from many existing experimental efforts in genome sequence, structural genomics, biochemical kinetics, and pharmacology. These data would be used in massive-scale systems biology representations of biochemical pathways (endogenous and xenobiotic metabolism, and regulatory mechanisms) within the human body, detailed as interconnected compartmentalized organs.

This infrastructure will allow an integrated system to be developed, including: 1) system-level bioinformatics data (e.g. genomics, transcriptomics, proteomics); 2) protein interaction data (e.g. kinetic characteristics of enzymes) based on atomistic protein structure function and cheminformatics); and 3) adverse outcome data (side effects) from clinical trials and elsewhere. The technical motivation for the infrastructure is based on the success of the Earth System Grid Federation (ESGF) and the status that it has achieved over the past decade in the climate domain—demonstrating the capability to help a scientific community self-organize to build an information and knowledge infrastructure that has revolutionized how climate modeling is done.

Looking forward, the biology community is looking to applying the successful ESGF infrastructure to meet our needs and the possible integration of cross-discipline study between the biology and climate.

Sincerely,

Kenneth W. Turteltaub
Division Leader
Biosciences and Biotechnology Division



National Aeronautics and
Space Administration
Goddard Space Flight Center
Greenbelt, MD 20771



March 7, 2013

Reply to Attn of:

606

R&D 100 Awards
100 Enterprise Drive
Suite 600, Box 912
Rockaway, NJ 07866-0912

Re: Support Letter for ESGF to Receive R&D 100 Awards

To Whom It May Concern:

It is with great pleasure and enthusiasm that we write this letter of recommendation and support for the Earth System Grid Federation (ESGF) to receive an R&D 100 Award. As members of the international observational and re-analysis data prepared working groups, we can attest to the critical importance of validating observational and re-analysis data against climate model simulations. Today, links between observational data experts and model developers/analysts are weak, and as a result, valuable data resources in understanding predictive climate change are underutilized and poorly understood.

As a result of this underutilization, two pilot projects have emerged to help bridge the gaps between the communities in order to better the model evaluation process: observation for Model Intercomparison Projects (obs4MIPs); and re-analysis for Model Intercomaprison Projects (ana4MIPs). These pilot projects aim to apply their observational and re-forecast data sets to the Coupled Model Intercomparison Project (CMIP) process for simulation data verification and validation. It is through the ESGF infrastructure that the co-mingling of data products can exist. At NASA, ESGF is one of the primary conduits in which we distribute NASA climate model output, observations, and re-analysis data to the climate change community. Conversely, it is the primary conduit in which we ascertain and access climate simulation and other related data products.

For the first time ever, ESGF is making it possible for disparate data and resources to be joined together under one knowledge discovery system, across different U.S. federal agencies (DOE, NOAA, NASA, NSF) and international borders. Without ESGF, it would be difficult to generate reports such as the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5), due out at the end of 2013.

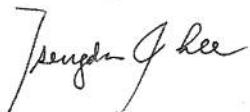
We have been pleased to be part of this ongoing ESGF effort and hope that the team will

be recognized for their outstanding contribution to climate change science and other scientific domains.

Sincerely,



Dr. W. Phillip Webster
Chief, Computational and Information Sciences and Technology Office, NASA Goddard



Digitally signed by TSENGDAR LEE
DN: c=US, o=U.S. Government,
ou=NASA, ou=People, cn=TSENGDAR
LEE, 0.9.2342.19200300.100.1.1=tjlee1
Date: 2013.03.07 12:44:05 -05'00'

Dr. Tsengdar Lee
Program Manager, High End Computing, NASA Headquarters



California Energy Systems for the 21st Century

Dr. Noah Goldstein, Director, Livermore Energy Systems Informatics Capability
Lawrence Livermore National Laboratory

Letter of Support for ESGF

March 1, 2013

To whom it may concern:

I relish this opportunity to recommend the Earth System Grid Federation (ESGF) for an R&D 100 award for its innovation and contributions. At LLNL's Livermore Energy Systems Informatics Capability (LESIC), we are tasked with identifying and implementing systems for integrating and managing large volumes of Energy Systems data. Among the challenges we face is how to structure data across local and regional repositories that can have varying degrees of sensitivity, yet as a whole will contribute greatly to our mission-based projects. In ESGF, we see a solution that will enable us to maximize data integration for simulation in a labor- and computationally- efficient manner.

In the upcoming months, we hope to begin a large integrated Advanced Computing project, called the California Energy Systems for the Twenty-First Century, or CES-21. This project will enable the California Electric and Gas utilities to leverage LLNL's simulation and modeling expertise to solve some of the most challenging problems of our Energy System modularization, including renewable power integration, streaming data, and massively expanding volumes of sensitive data. We have the responsibility of managing those data associated with that effort, as well as developing new techniques for seamless simulation and data analysis. In CES-21, we are looking to innovative approaches like ESGF to provide some leadership in creating solutions to our problems, especially in the domains of data integration and management, pre- and post-processing, as well as data analysis and visualization.

Here at LLNL and LESIC, we hope that ESGF will enable CES-21 and similar projects to help with creating a more efficient grid, a safer energy system, with enthusiastic support of our stakeholders and public partners. We are very pleased that ESGF exists as an easy-to-use model of data distribution, integration, and processing for some of society's most important problems. At LESIC, we look forward to applying ESGF's innovations in the Energy System domain.

Thank you,

Noah Goldstein
Goldstein8@llnl.gov



UNITED STATES DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
GEOPHYSICAL FLUID DYNAMICS LABORATORY
Princeton University
Forrestal Campus, US Route 1
201 Forrestal Road
Princeton, New Jersey 08540

22 February 2013

Support Letter for ESGF

Ronald J Stouffer
Physical Scientist GFDL/NOAA
Head of the Climate Model Intercomparison Project (CMIP) Panel

To whom it may concern:

The CMIP activity is an organized effort by the world's climate modeling community to freely provide climate information to other scientists and anybody who is able to download and process the data. These data sets are being used for analysis and increased scientific understanding. They provides the scientific basis for many of the international climate reports, including the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment report due out later this year. The scientific findings in the IPCC reports provide input to climate policy negotiations between and among countries. The CMIP database is therefore of extremely high value to society both inside the U.S. and the world.

The total amount of data under CMIP Panel oversight is about 4 Pb and is one of the larger databases in the world. This data does not reside on any one server but is distributed around the world across 10's of data servers. ESGF provides the infrastructure which allows the data to be useful to those who try to obtain data from the CMIP database. This software is extremely complex as it involves allowing various servers to access information on each other across the internet and the security issues associated with that process.

There is an urgent need for more funding support for the ESGF activity. Without the support, it is possible that serving the CMIP database as outlined above will no longer be possible which will greatly hinder the advancement of climate science.

Therefore I strongly support the ESGF proposal.

