Earth System Grid Federation (ESGF)

Computation External Review Committee

February 11−13, 2014

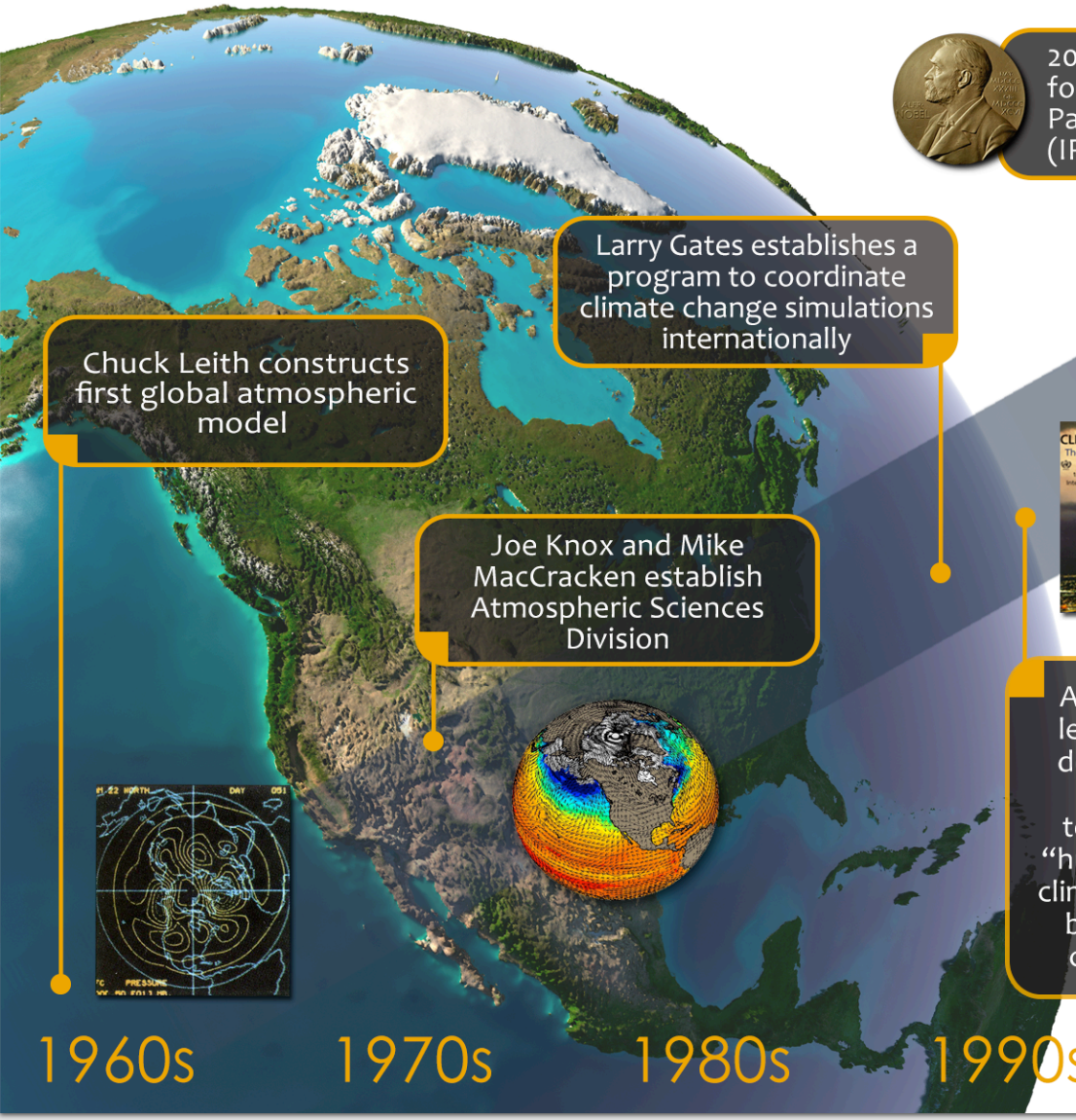
Dean N. Williams, Principal Investigator

Lawrence Livermore National Laboratory

LLNL-PRES-648666

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

LLNL was at the forefront of climate science from the beginning and remains a leader in the field



Chuck Leith constructs first global atmospheric model

Larry Gates establishes a program to coordinate climate change simulations internationally

Joe Knox and Mike MacCracken establish Atmospheric Sciences Division

1960s 1970s 1980s 1990s 2000s 2010s

2007 Nobel Peace Prize for Intergovernmental Panel of Climate Change (IPCC) work

Award-winning research, led by Ben Santer, on the detection and attribution of climate change that took a bold stand on the "human influence on global climate" and was supported by the Lab in the face of considerable push-back

- E.O. Lawrence Award
- MacArthur Fellowship
- National Academy of Sciences

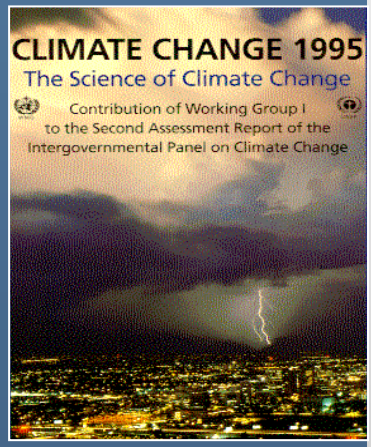
Steve Klein leads group focusing on cloud processes, a key source of uncertainty in climate projections

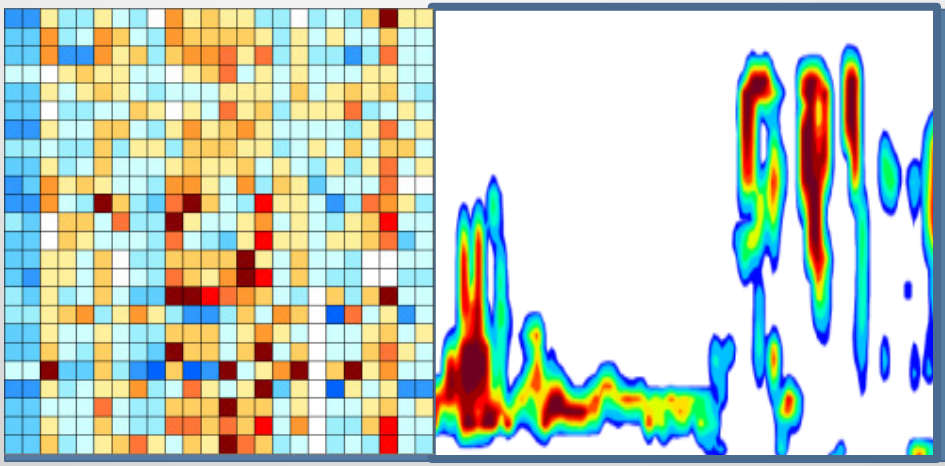
Tom Guilderson receives E.O. Lawrence Award in 2011 for his work on the global carbon cycle

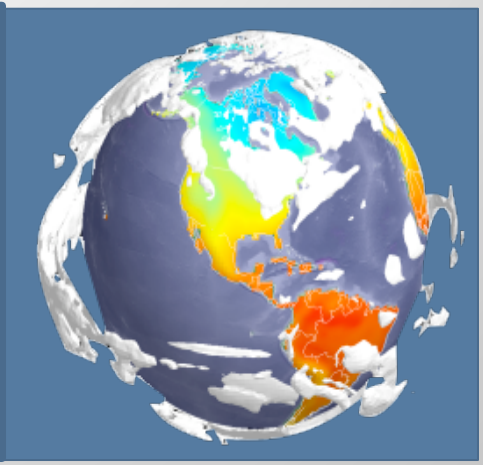
The 5th IPCC assessment report was published in November of 2013

World leader in climate change detection and attribution

DOE/SC Program on Climate Model, Diagnosis and Intercomparison at LLNL has played a key role in the international climate community







"The balance of evidence suggests a discernible human influence on global climate." Ben Santer

- DOE Program at LLNL since 1989
- Goal: Quantify fidelity of model simulations and uncertainty in projections
- Research:
  - Understand fidelity of climate model simulations
  - Diagnose and fingerprint human signatures in climate change
  - Advance climate data management, analytic and informatics tools
- Publications: 186 peer review publications from DOE funding over the last seven years with 21 in Science, Nature and PNAS

ESGF[1] is a coordinated multiagency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate

1. 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, "The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geospatial Data", *IEEE Future Generation Computer Systems*, http://dx.doi.org/10.1016/j.future.2013.07.002, 17 September 2013.

The climate community relies on ESGF data archives

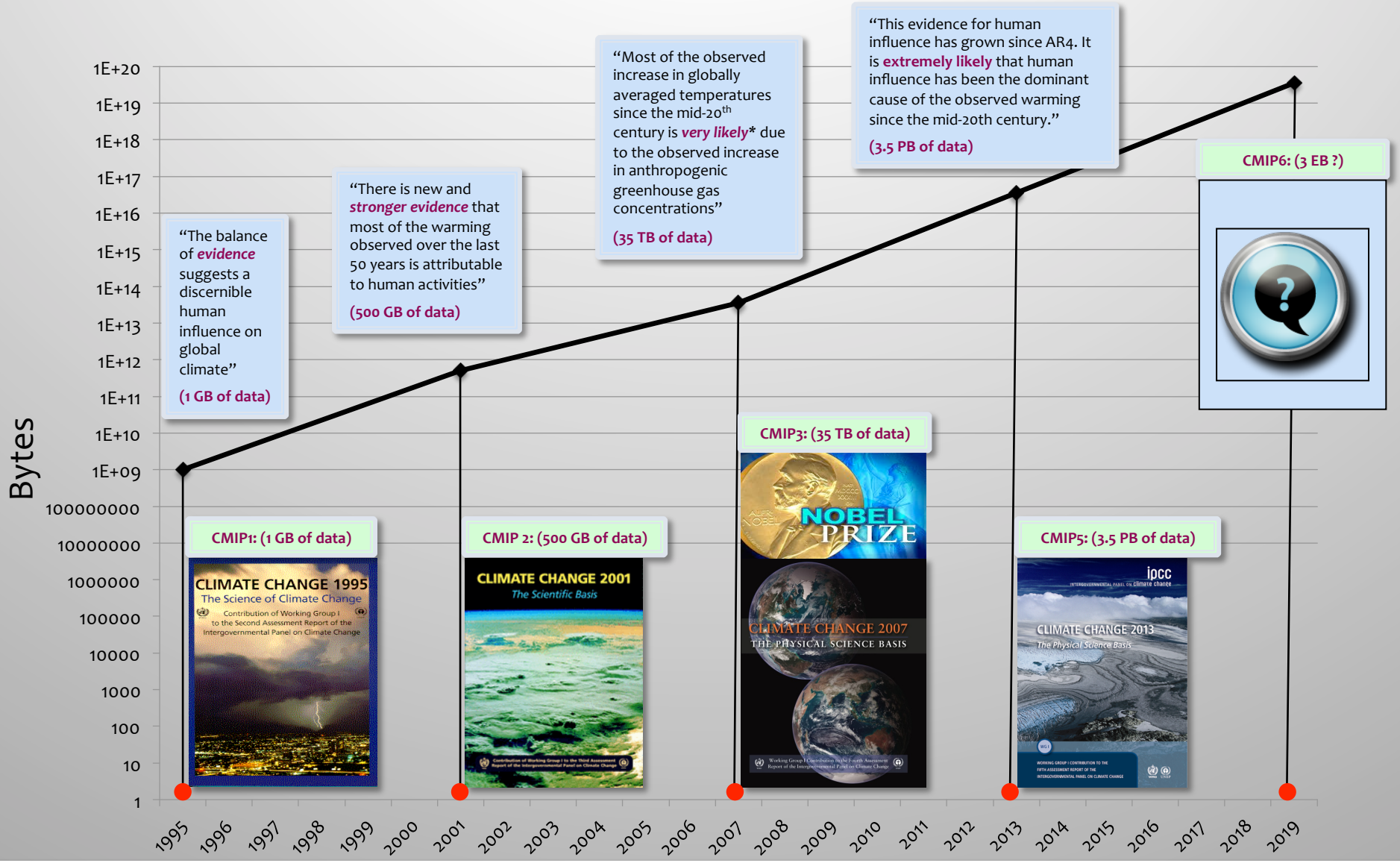
Institution	Data Set	P	Type	Description	Use	Status
ANL	CAM-SE High-Res Gridded	L	Model	Gridded	Atmosphere	Published
BADC	TAMIP	M	Model	Gridded	Atmosphere	Collection
BNL	ARM CSAPR Rainfall	L	Observational	Gridded	Atmosphere	Collecting
BNL	NOAA NEXRAD MOSAIC Rainfall	L	Observational	Gridded	Atmosphere	Collecting
LANL	POP	H	Model	Gridded	Ocean	
LLNL	**CMIP5, CMIP3**	H	Model	Gridded	Atmosphere Land Ocean	Collecting
LLNL	NASA MERRA, CERES	M	Reanalysis	Satellite	Atmosphere	Published
LLNL/PNNL	NASA TRMM,	H	Observational	Satellite	Atmosphere	Collecting (TRMM)
LLNL/PNNL	NASA AIRS, MLS, TES	H	Observational	Satellite	Atmosphere	Collecting
LLNL/PNNL	CAM5 Sensitivity Runs	H	Model	Gridded	Atmosphere	Awaiting availability
MPI-M	CMIP5	H	Model	Gridded	Atmosphere Land Ocean	Collecting
MPI-M	EUCLIPSE	H	Model	Gridded	Atmosphere	Will collect
MPI-M	LUCID	M	Model	Gridded	Atmosphere Land Ocean	Will collect
MPI-M	CORDEX	M	Model	Gridded	Atmosphere Land Ocean (regional)	Will collect
NCAR	NARCCAP	H	Model	Regional Gridded	Atmosphere	Collecting
NCAR	PCM	M	Model	Gridded	Atmosphere Land Ocean	Published
NCAR	CCSM	H	Model	Gridded	Atmosphere Land Ocean	Published
NCAR	CADIS	H	Model	Gridded	Atmosphere Land Ocean	Published
ORNL	CLM Single Point	H	Model	Single Point	Land	Published
ORNL	CLM Gridded	H	Model	Gridded	Land	Published

Institution	Data Set	P	Type	Description	Use	Status
ORNL/LLNL/ PNNL	ARMBE (a.k.a. CMBE)	M	Observational	Single Point Gridded	Atmosphere	Published 6 variables at ORNL
ORNL	C-LAMP	H	Model	Single Point Gridded	Land	Published
ORNL/PNNL	AmeriFlux	H	Observational	Gap-filled surface weather forcing data (U.S. sites)	Land	Published
ORNL	Fluxnet Canada	M	Observational	Gap-filled surface weather forcing data (Canadian sites)	Land	Awaiting Authorization
PNNL	USGS Basin Boundaries	L	Geospatial Reference	Vector	Land	Collected
PNNL	STATSGO, LAI, SAI (MODIS)	L	Reanalysis	Single Point Gridded	Land	Collected
PNNL	NLDAS2	L	Observational Reanalysis	Single Point	Land	Collected
PNNL	DEM (MOPEX)	M	Geospatial Reference	Gridded	Land	Collected
PNNL	CLM4 Single Point, sample of parameters	H	Model	Single Point	Land	Collected
PNNL/ORNL	USGS Stream flow for MOPEX basins	M	Observational	Single Point Time Series	Land	Collected
PNNL	MTSAT	M	Observational	Satellite	Atmosphere	Collecting
PNNL/LLNL	NASA CALIPSO, CloudSat	L	Observational	Satellite	Atmosphere	Collecting
SNL	Hydrobase3	H	Observational climatologies	Gridded	Ocean	Awaiting availability
SNL	World Ocean Atlas 2009	H	Observational climatologies	Gridded	Ocean	Will collect
SNL	LES runs	M	Model	Single Point Gridded	Ocean	Will collect
SNL	Florida Current Project	H	Observational	Single Point	Ocean	Will collect
SNL	RADARSAT-1	H	Observational	Gridded	Sea ice	Will collect
SNL	ESA Cryosat Siral	M	Observational	Single Point Gridded	Sea ice Ice sheet	Determining availability
SNL	NSIDC, Hadley Ctr Passive microwave	H	Observational	Gridded	Sea ice Ice sheet	Will collect
SNL	ICESat	H	Observational	Gridded	Sea ice	Will collect
SNL	IceBridge	H	Observational	Gridded	Sea ice	Will collect

International climate research charge: Coupled Model Intercomparison Projects (CMIPs)

- The Intergovernmental Panel on Climate Change (IPCC) prepares an assessment report about climate science approximately every 6 years
- This climate modeling research requires enormous scientific and computational resources
  - ✓ Involves over two dozen modeling research groups worldwide; and
  - ✓ Encourages cooperative community international research activities
- The World Climate Research Program (WCRP) serves as the primary coordinating body for this research activity
  - ✓ The WCRP's Working Group on Coupled Modeling (WGCM) relies on Lawrence Livermore National Laboratory (LLNL) to support these activities by coordinating and maintaining the distributed petabyte data archive

LLNL has been a major contributor to every Intergovernmental Panel on Climate Change (IPCC): CMIP data history



Data archive challenges for CMIP5 is 100× compared with CMIP3

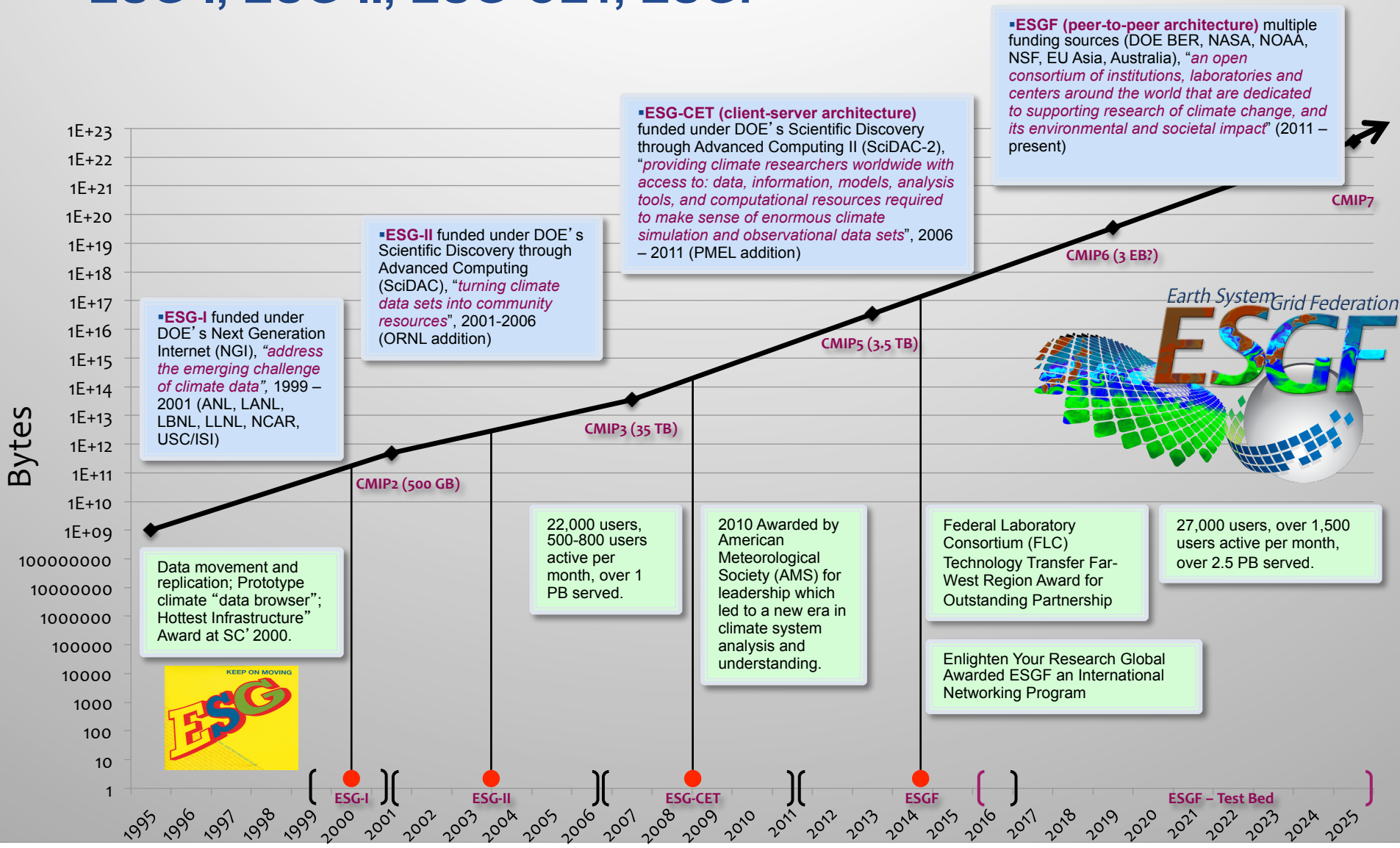
CMIP3 Modeling Centers		volume (GB)
BCCR	Norway	862
CCCma	Canada	2,071
CNRM	France	999
CSIRO	Australia	2,088
GFDL	USA	3,843
GISS	USA	1,097
IAP	China	2,868
INGV	Italy	1,472
INMCM3	Russia	368
IPSL	France	998
MIROC3	Japan	3,975
MIUB	Germany/Korea	477
MPI	Germany	2,700
MRI	Japan	1,025
CCSM	USA	9,173
UKMO	UK	973
Totals		**34,989 (TB)**

Archive size: 35 TB

CMIP5 Modeling Centers		volume (TB)
BCC	China	51
CCCma	Canada	51
CMCC	Europe (Italy)	158
CNRM	France	71
CSIRO	Australia	81
EC-EARTH	Europe (Netherland)	97
GCESS	China	24
INM	Russia	30
IPSL	France	121
LASG	China	100
MIROC	Japan	350
MOHC	UK	195
MPI	Germany	166
MRI	Japan	269
NASA	USA	375
CESM	USA	1,200
NCC	Norway	32
NCEP	USA	26
NIMR/KMA	Korea	14
NOAA GFDL	USA	158
Totals		**3,108 (PB)**

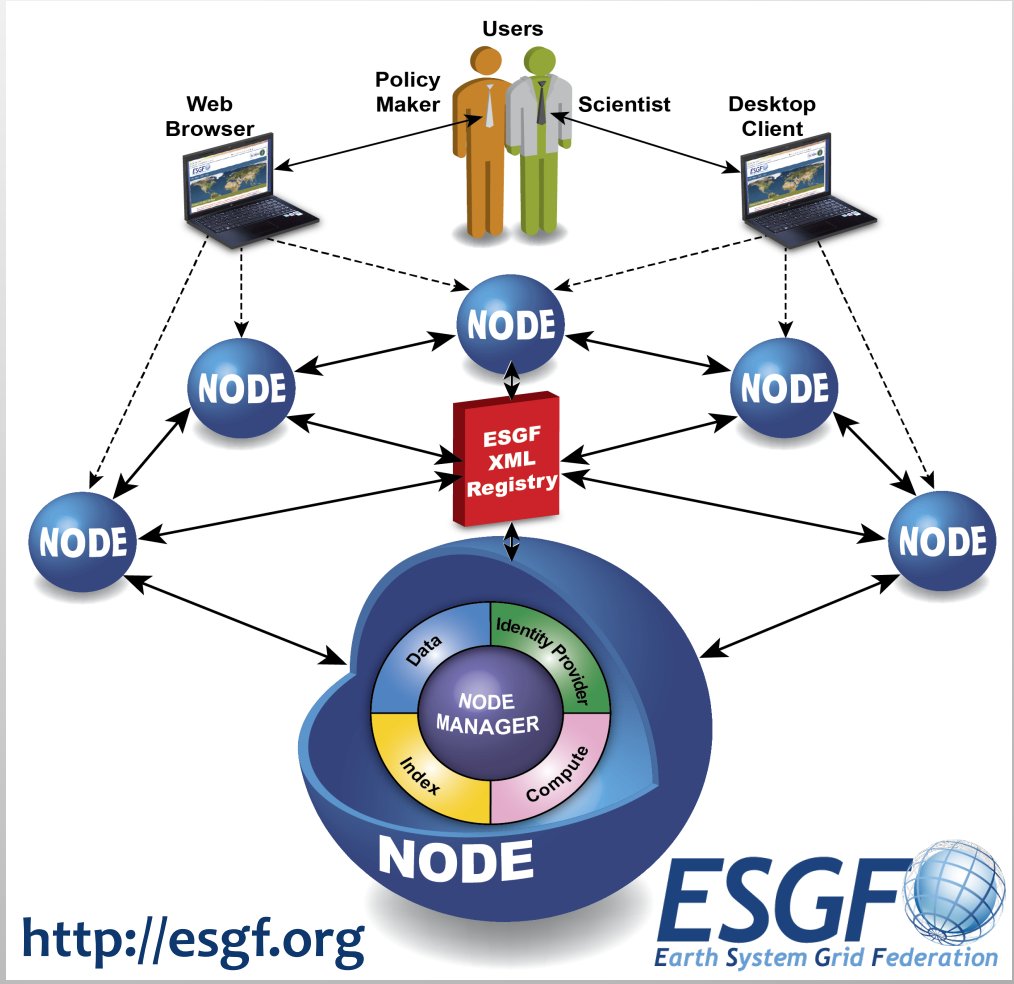
Archive size: 3.1 PB

History and community of the Earth System Grid (ESG): ESG-I, ESG-II, ESG-CET, ESGF



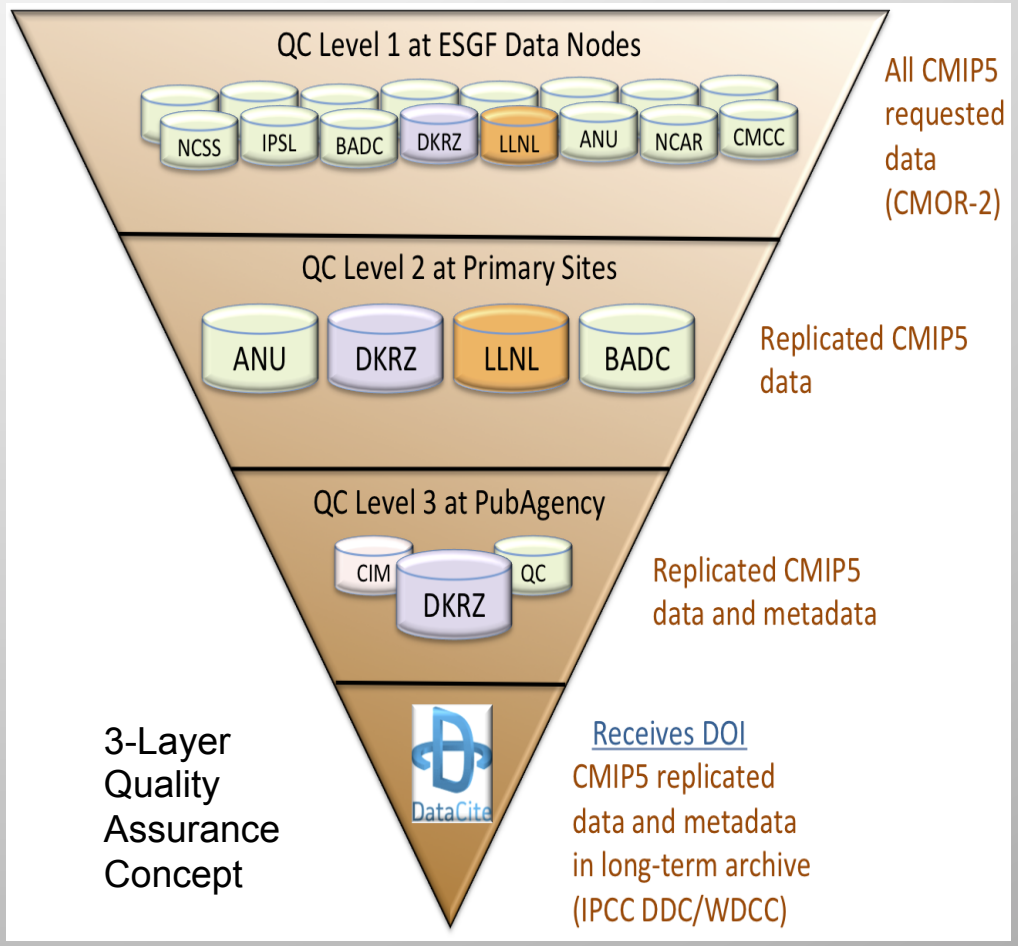
ESGF distributed data archival and retrieval system

- Federated peer-to-peer architecture
- Support discipline-specific portals
- Support browser-based and direct client access
- Single sign-on
- Automated script and GUI-based publication tools
- Full support for data aggregations
  - ✓ A collection of files, usually ordered by simulation time, that can be treated as a single file for purposes of data access, computation, and visualization
- User notification service
  - ✓ Users can choose to be notified when a data set has been modified



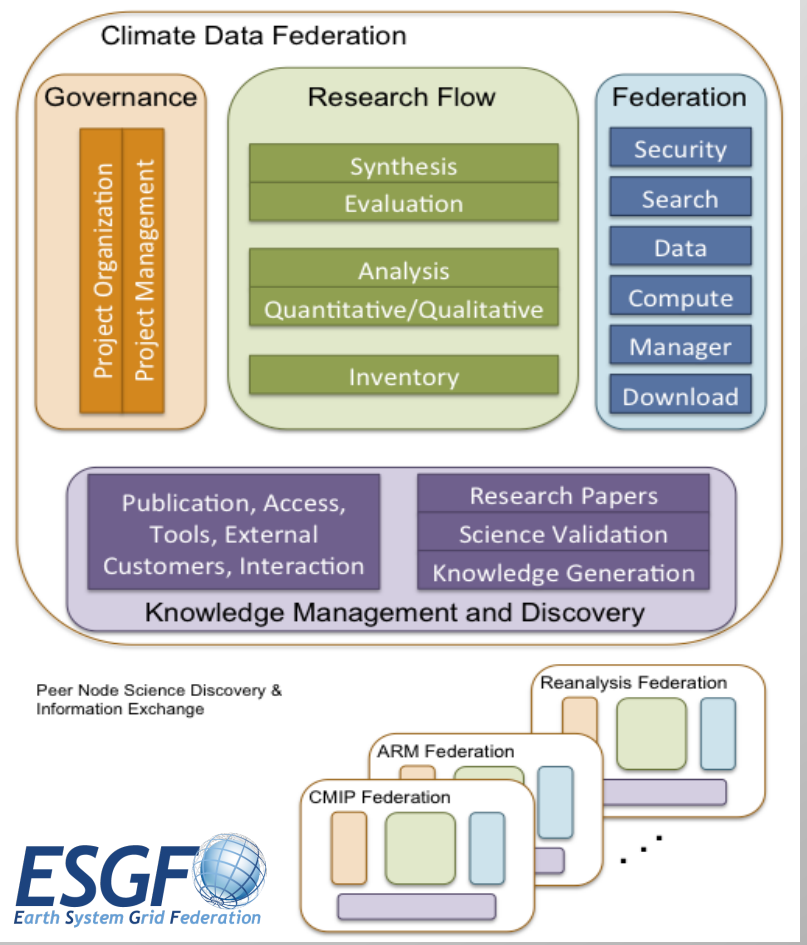
ESGF provides data quality control processing

- Publishing data to an ESGF portal performs QC Level 1 (QCL1) check
  - ✓ QCL1 data are visible to users and are identified as QCL1 on the UI
- DKRZ (MPI) quality control code is run on data to perform QC Level 2 (QCL2) check
  - ✓ QCL2 data are visible to users and are identified as QCL2 on the UI
- Visual inspections are performed for inconsistencies and metadata correctness at QC Level 3 (QCL3) check
  - ✓ QCL3 data are visible to users and are identified as QCL3 on the UI
  - ✓ Digital Object Identifiers (DOIs) are given to data sets that pass the QCL3 check



ESGF software system integration data federation services (i.e., data services)

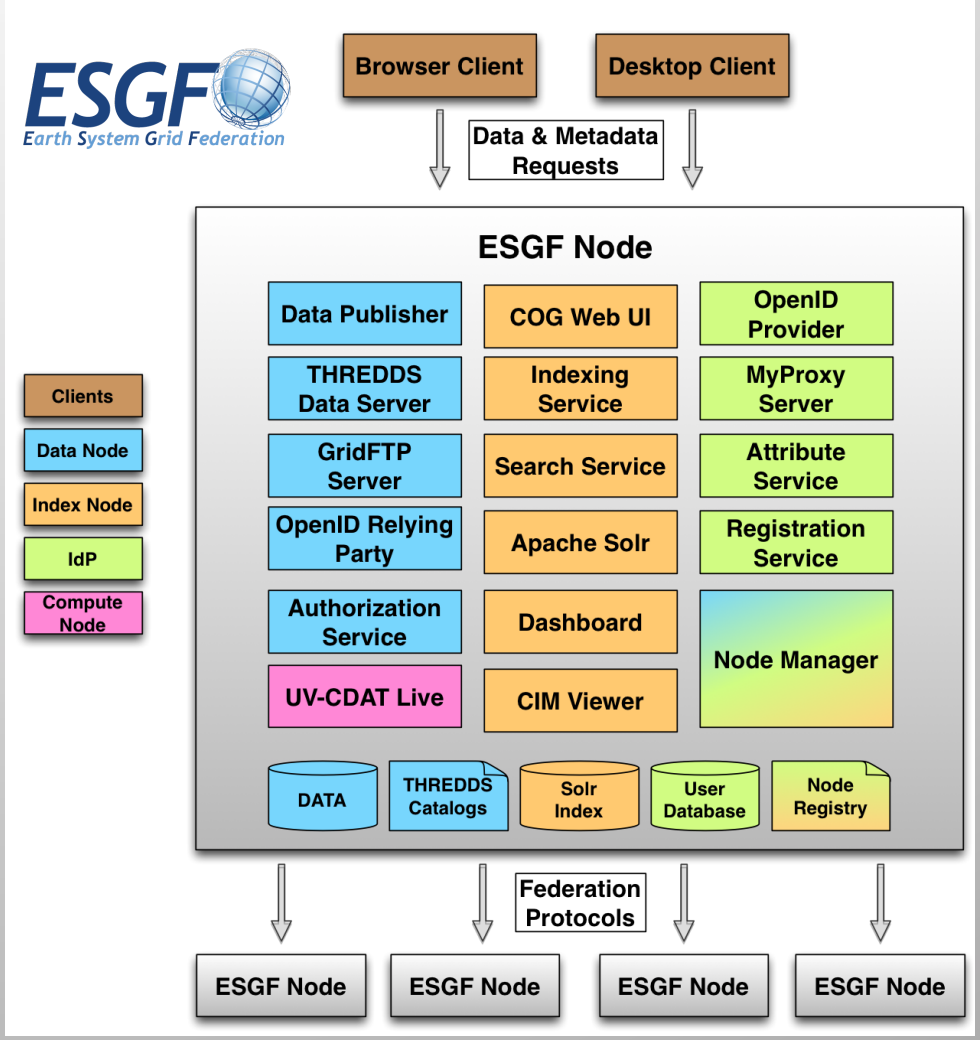
- NetCDF Climate and Forecast (CF) Metadata Convention
  - ✓ LibCF, Mosaic
- Climate Model Output Rewriter 2 (CMOR-2)
- Regridders: GRIDSPEC, SCRIP, and ESMF
- Publishing
- Search and Discovery
- Replication and Transport
  - ✓ GridFTP, OPeNDAP, DML, Globus Online, ftp, BeSTMan (HPSS)
  - ✓ Networks
- Data Reference Syntax (DRS)
- Common Information Model (CIM)
- Quality Control
  - ✓ QC Level 1, QC Level 2, QC Level 3, Digital Object Identifiers (DOIs)
- Websites and Web Portal Development
  - ✓ Data, Metadata, Journal Publication Application
- Notifications, Monitoring, Metrics
- Security
- Product Services
  - ✓ UV-CDAT[2]



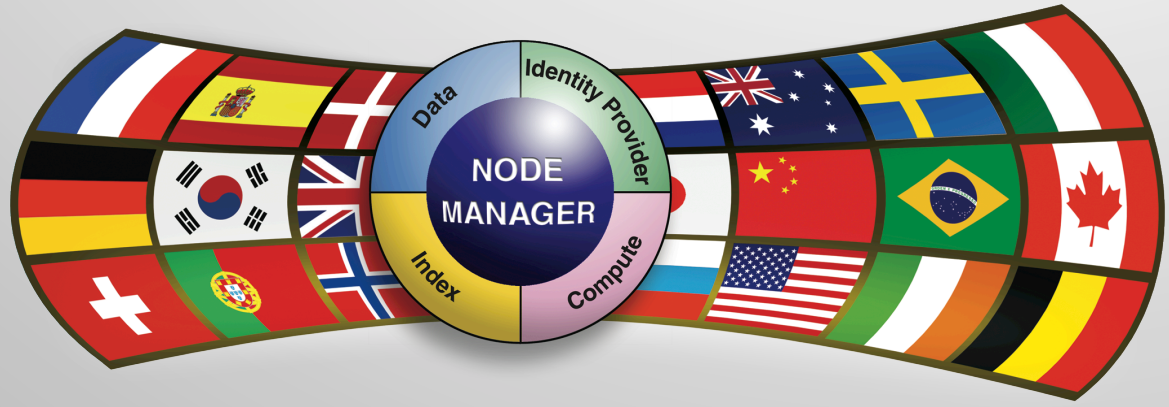
[2] Dean N. Williams, Timo Bremer, Charles Doutriaux, John Patchett, Sean Williams, Galen Shipman, Ross Miller, David R. Pugmire, Brian Smith, Chad Steed, E. Wes Bethel, Hank Childs, Harinarayan Krishnan, Prabhat, Michael Wehner, Claudio T. Silva, Emanuele Santos, David Koop, Tommy Ellqvist, Jorge Poco, Berk Geveci, Aashish Chaudhary, Andy Bauer, Alexander Pletzer, David Kindig, Gerald L. Potter, Thomas P. Maxwell, "Ultrascale Visualization of Climate Data," IEEE *Computer Magazine*, September 2013, **46** (9), 68−76, DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/MC.2013.119.

New ESGF technologies developed and deployed by LLNL and our partners

- Peer-to-peer
- Search services
- Security services
- User interface development
- Data publisher
- Live access server (compute node)
- Data transfer (WGET, BDM, GridFTP, Globus Online, BeStMan)
- DAP services (THREDDS Data Servers [TDS], OPeNDAP)
- Really simple syndication (RSS) feed
- UV-CDAT (client analysis tool access)
- Dashboard (system monitor service)
- Replication and versioning
- Installation script



Current CMIP federated portals (LLNL's portal: pcmdi9.llnl.gov)



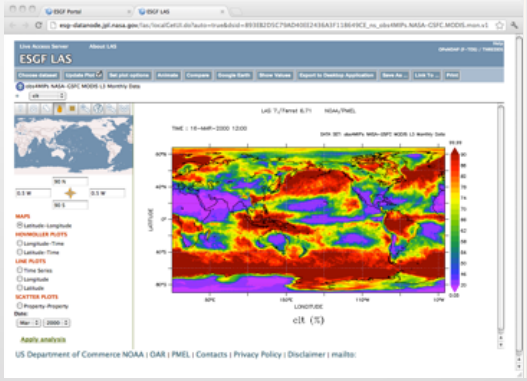
Peer Node Status

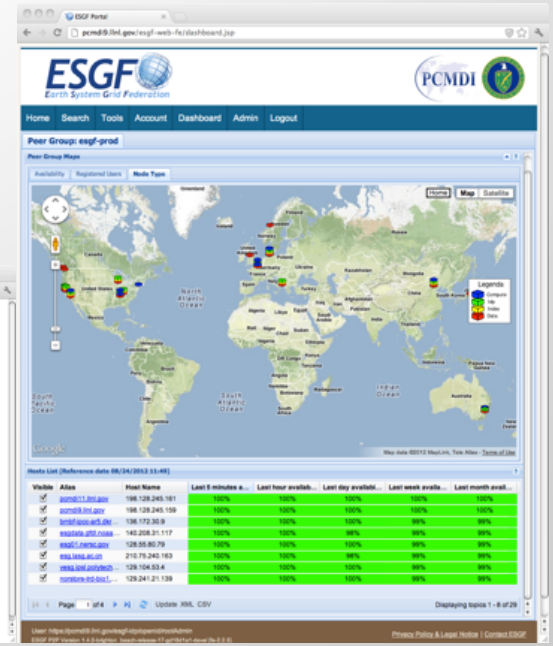
Page History | Clone URL

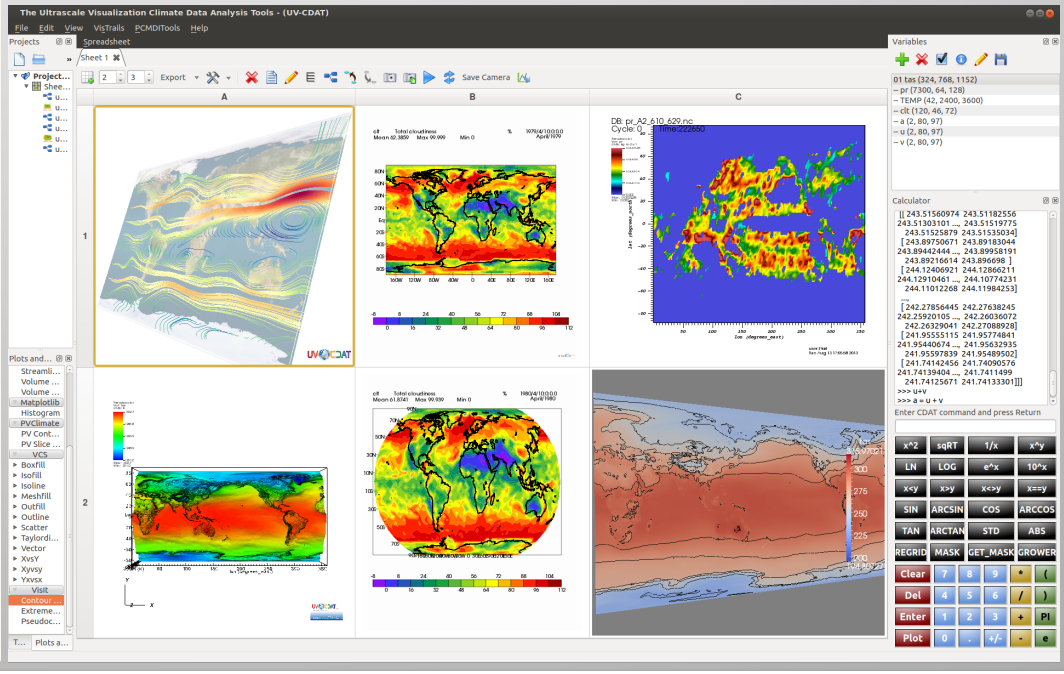
Node	URL	Admin	Status
ANL	http://dev.esg.anl.gov	Eric Blau	up
BADC	http://esgf-index1.ceda.ac.uk	Stephen Pascoe	up
BADC	http://cmip-dn1.badc.rl.ac.uk	Stephen Pascoe	up
BADC	http://esgf-data1.ceda.ac.uk	Stephen Pascoe	up
BCC	https://bcccsm.cma.gov.cn	Min Wei	up
BNU	http://esg.bnu.edu.cn	Qizhong Wu	up
CCCma	http://dapp2p.cccma.ec.gc.ca	Mike Berkley	up
CEA (IPSL)	http://dods1.extra.cea.fr	Sebastien Denvil	up
CMCC	http://adm07.cmcc.it	Sandro Fiore	up
CNRM	http://esg.cnrm-game-meteo.fr	Eric Gerbier	up
DIAS (MIROC/MRI/NICAM)	http://dias-esg-nd.tkl.iis.u-tokyo.ac.jp	Toshihiro Nemoto	up
DKRZ	http://euclipse1.dkrz.de	Hans Ramthun	up
DKRZ (replica)	http://albedo2.dkrz.de	Stephan Kindermann	up
DKRZ (MPI)	http://bmbf-ipcc-ar5.dkrz.de	Stephan Kindermann	up
DKRZ (replica)	http://cmip3.dkrz.de	Stephan Kindermann	up
DKRZ CMIP5	http://esgf-data.dkrz.de	Stephan Kindermann	up
DMI	http://cordexesg.dmi.dk	Ole Bøssing Christensen	up
EC-EARTH	http://esg2.e-inis.ie	Alastair McKinstry	up
FIO	http://cmip5.fio.org.cn	Zhenya Song	up
IPSL	http://esgf-node.ipsl.fr	Sebastien Denvil	up
IPSL	http://vesg.ipsl.fr	Sebastien Denvil	up
IPSL	http://vesg.ipsl.polytechnique.fr	Sebastien Denvil	up
LASG	http://esg.lasg.ac.cn	Kangjun Chen	up
LLNL	http://pcmdi9.llnl.gov	Gavin Bell	up
LLNL	http://pcmdi11.llnl.gov	Gavin Bell	up
NASA-GISS		Laura Carriere	-
NASA-GMAO-COLA		Laura Carriere	-
NASA-GSFC	http://esgdata1.nccs.nasa.gov	Zed Pobre	up
NASA-JPL	http://esg-datanode.jpl.nasa.gov	Luca Cinquini	up
NCC		Ingo Bethke	-
NCEP		Laura Carriere	-
NCDC/NOMADS			-
NCI	http://esg2.nci.org.au	Ben Evans	up
NERSC	http://esg01.nersc.gov	Shreyas Cholia	up
NIMR/KMA		Young Hwa Byun	-
NOAA-ESRL	http://hydra.fsl.noaa.gov	Luca Cinquini	up
NOAA-GFDL	http://esgdata.gfdl.noaa.gov	Serguei Nikonov	up
NSF-DOE-NCAR		Shreyas Cholia	-
ORNL	http://esg.ccs.ornl.gov	John Harney	up
PIK	http://esg.pik-potsdam.de	Matthias Büchner	up
SMHI-LIU-NSC	http://esg-dn1.nsc.liu.se	Prashanth Dwarakanath	up
UIO	http://noresg.norstore.uio.no		up
UIO	http://noresg.norstore.uio.no	Ingo Bethke (CMIP5) / Martin King (CORDEX)	up

last updated: 01/16/14 : 05:00:02

New ESGF browser and desktop clients

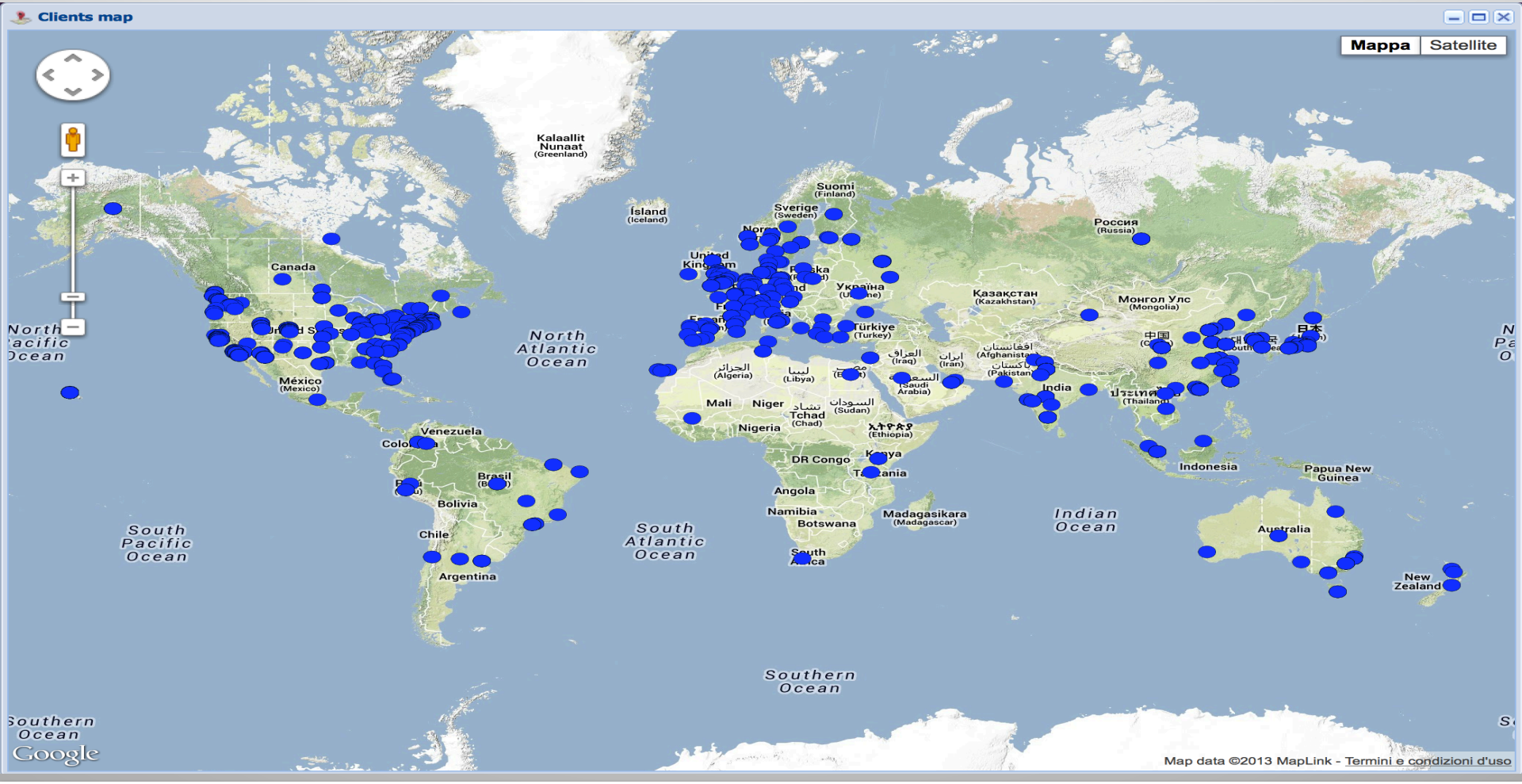




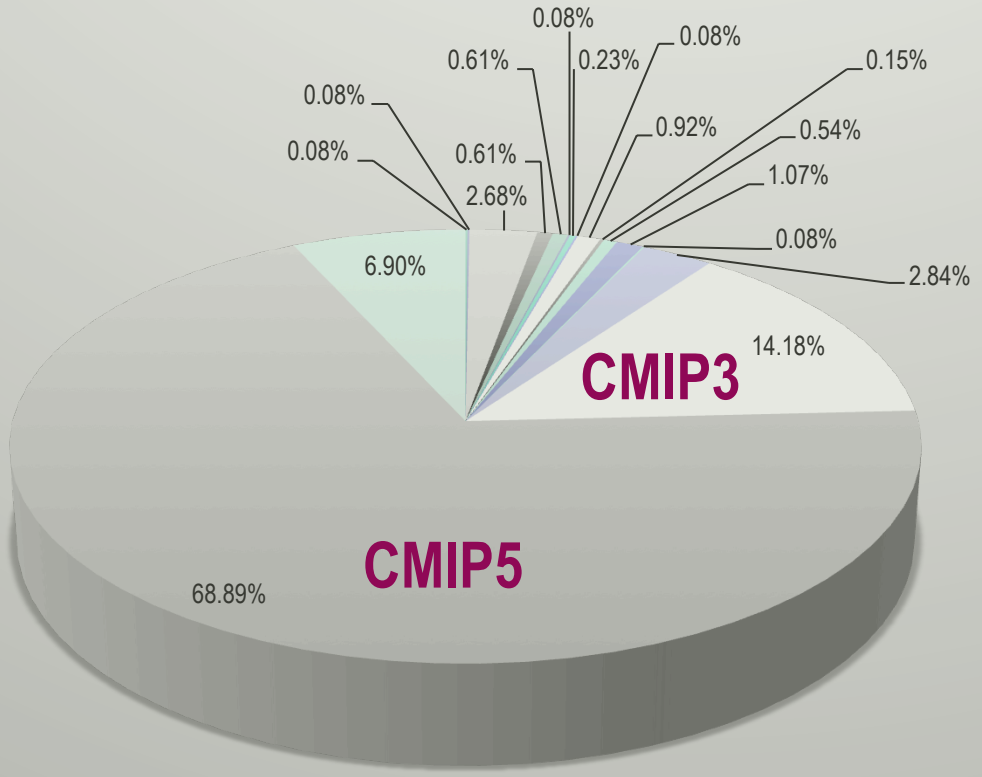


ESGF nodes are deployed internationally

Geo-distribution related to the users that have downloaded CMIP5 data sets from pcmdi9.llnl.gov (over 2,000 publications generated from archive)



Users distributed by identity provider



- euclipse1.dkrz.de
- esg-dn1.nsc.liu.se
- ceda.ac.uk
- dev.esg.anl.gov
- esg-datanode.jpl.nasa.gov
- esg-gateway.jpl.nasa.gov
- esg-vm-demo02.ccs.ornl.gov
- esg.ccs.ornl.gov
- esg.nci.org.au
- esg1-gw.pnl.gov
- esg2.nci.org.au
- esgf-data.dkrz.de
- esgf-node.ipsl.fr
- ipcc-ar5.dkrz.de
- pcmdi3.llnl.gov
- pcmdi9.llnl.gov
- www.earthsystemgrid.org

ESGF climate data holdings and growing

- Phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5)
- Coordinated Regional Climate Downscaling Experiment (CORDEX)
- Climate Science for a Sustainable Energy Future (CSSEF)
- European Union Cloud Intercomparison, Process Study and Evaluation Project (EUCLIPSE)
- Geo-engineering Model Intercomparison Project (GeoMIP)
- Land-Use and Climate, Identification of Robust Impacts (LUCID)
- Paleoclimate Modeling Intercomparison Project (PMIP)
- Transpose-Atmospheric Model Intercomparison Project (TAMIP)
- Clouds and Cryosphere (cloud-cryo)
- Observational Products More Accessible for Coupled Model Intercomparison (obs4MIPs)
- Reanalysis for the coupled model intercomparison (ANA4MIPs)
- Dynamical Core Model Intercomparison Project (DCMIP)
- Community Climate System Model (CCSM)
- Parallel Ocean Program (POP)
- North American Regional Climate Change Assessment Program (NARCCAP)
- Carbon Land Model Intercomparison Project (C-LAMP)
- Atmospheric Infrared Sounder (AIS)
- Microwave Limb Sounder (MLS)

27,000 users; >1,500 users active per month; >2.5 PB served

ESGF is an award winner of the Enlighten Your Research Global international networking program

Global International Networks: ESnet (U.S.); Janet (U.K.); DFN (Germany); AARnet (Australia); SURFnet (Netherlands)

The Federal Laboratory Consortium (FLC) recognized ESGF for outstanding partnership

FEDERAL LABORATORY CONSORTIUM FOR TECHNOLOGY TRANSFER

FAR-WEST REGION

Outstanding Partnership

Earth System Grid Federation

Lawrence Livermore National Laboratory, Jet Propulsion Laboratory, Oak Ridge National Laboratory, Pacific Marine Environmental Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, German Climate Computing Center, Euro-Mediterranean Center on Climate Change, British Atmospheric Data Centre, Goddard Space Flight Center, Institute Pierre Simon Laplace, Geophysical Fluid Dynamics Laboratory

October 21, 2013

For additional information, please visit:

- http://esgf.org
- http://uv-cdat.org
- http://aims.llnl.gov
- http://climate.llnl.gov
- http://www-pcmdi.llnl.gov

Lawrence Livermore
National Laboratory