Lawrence Livermore National Laboratory

Analytics, Informatics, and Management Systems

California State University: Chico

Chico, USA, 12 November, 2014

Sam Fries : Matthew Harris

Lawrence Livermore National Laboratory

LLNL-PRES-657120

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

Data integrated systems for science

Multidisciplinary team, technology, and engineering

Expanding to other scientific domains

Critical Complex Data Generation Systems

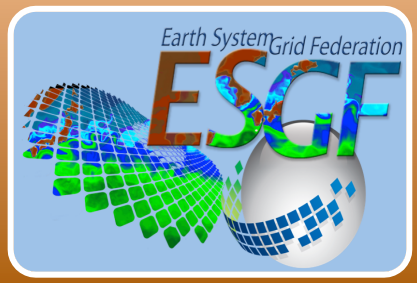
aims.llnl.gov

Networks

ICNWG International Climate Network Working Group

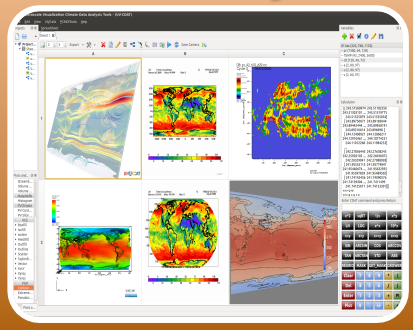
icnwg.llnl.gov

Data Collection and Management

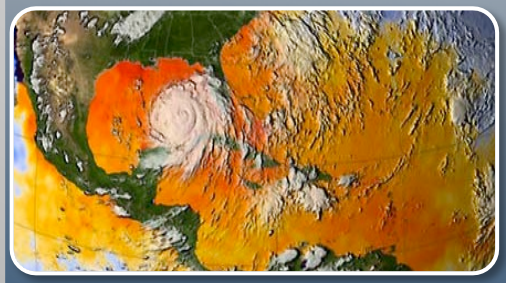


esgf.org

Data Analytics

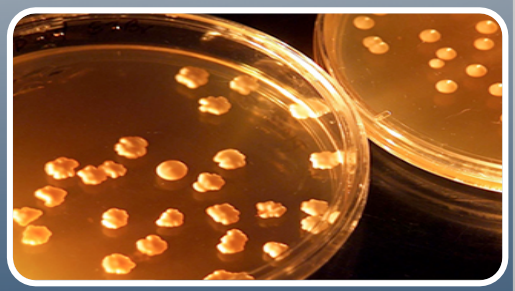


uv-cdat.llnl.gov



Environment

Energy



Biology

Climate Integrated Data Ecosystem and Workflow

Security

ESGF Earth System Grid Federation

Data Collection and Management
- Sensors, field and lab experiments
- Data models
- Transport and communications
- Data quality and uncertainty
- Storage, provenance and discovery

Provenance Capture

UV-CDAT

Critical Complex Data Generation Systems

Data-Intensive Computing
- Architectures – persistent data to streams
- Programming environments
- Human computer interface

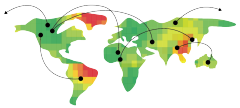
Data Analytics (local & remote)
- Descriptive statistics
- Graph analytics
- Machine learning
- Signal and image processing
- Pattern discovery
- Visualization
- Exploratory analysis

ACME Accelerated Climate Model for Energy

Analytical Modeling

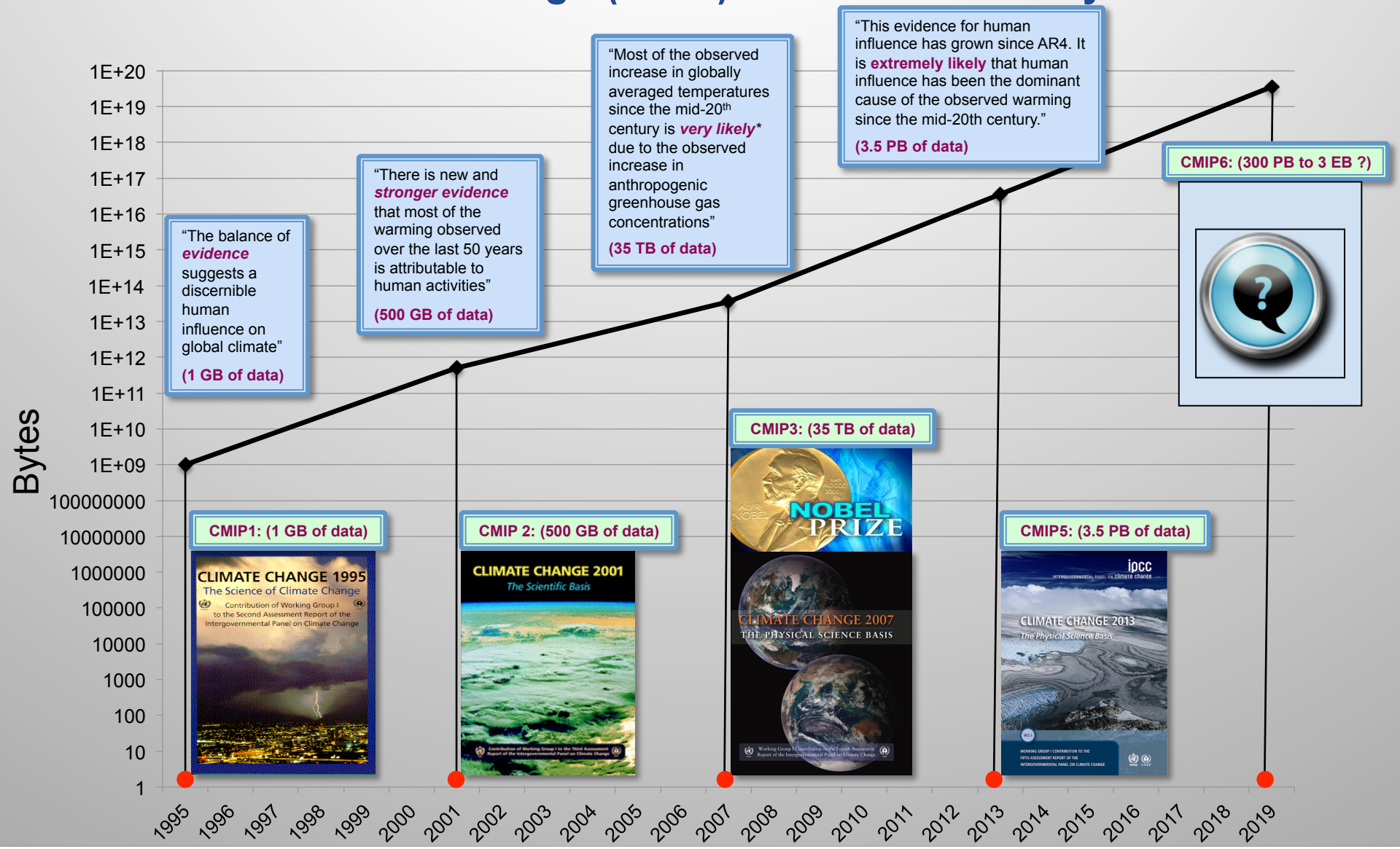
Decisions and Control Design Optimization Policy Making (Humans)
- Understanding and predicting use

ICNWG International Climate Network Working Group

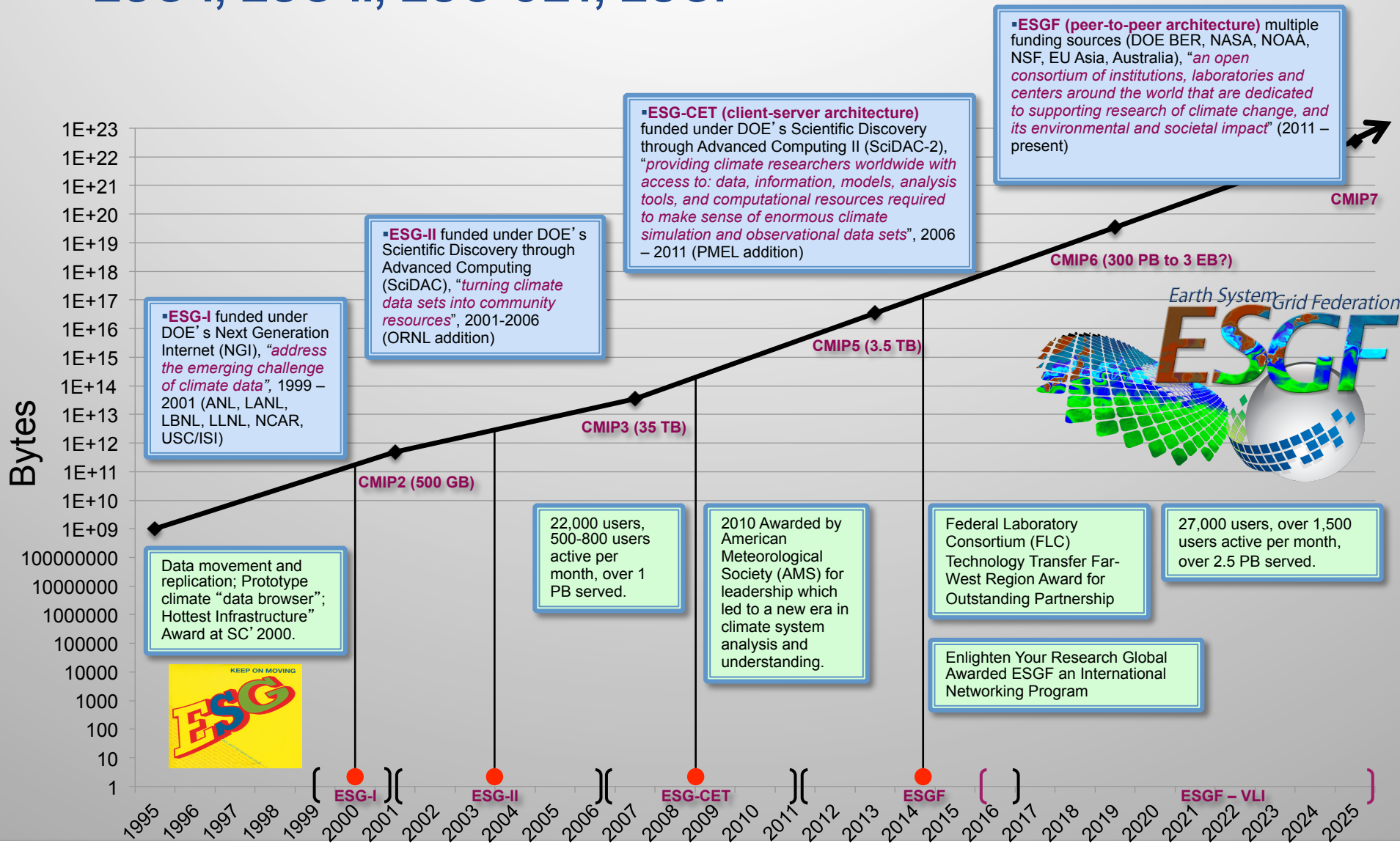


Network

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

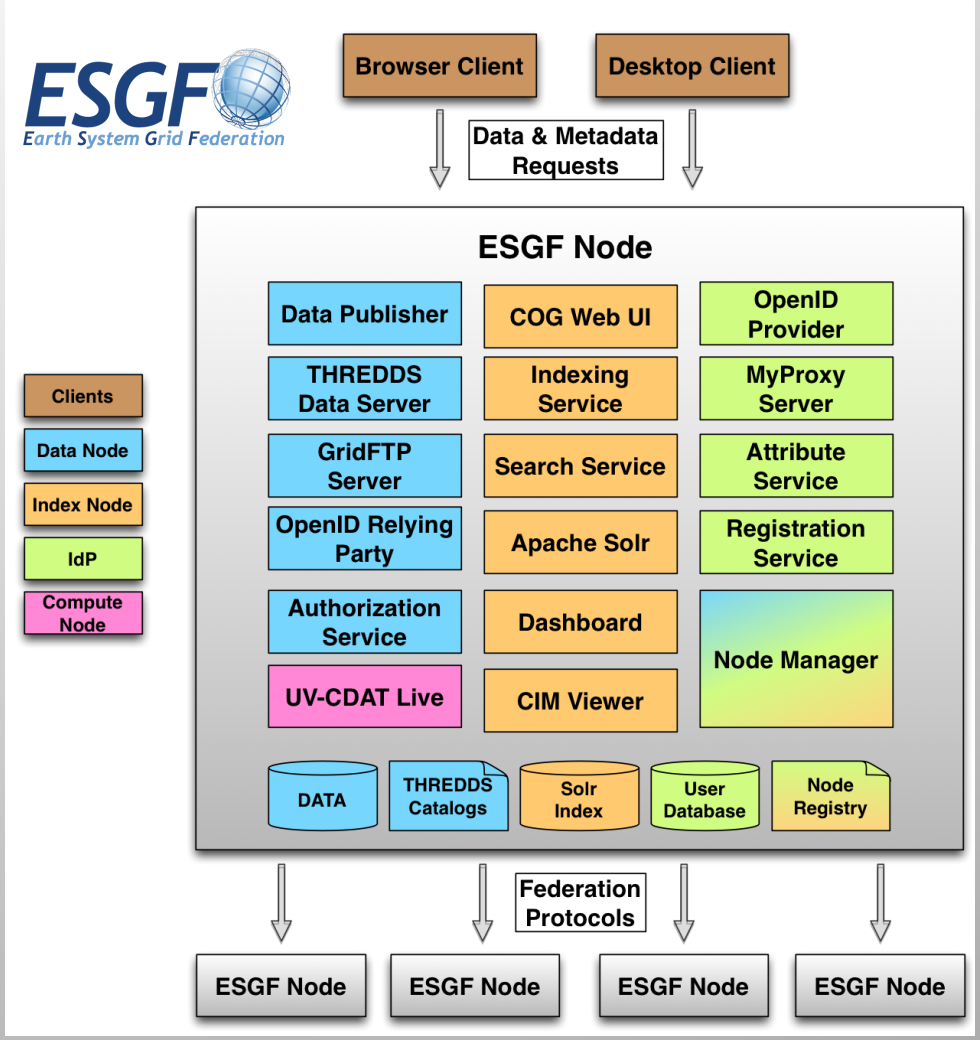


History and community of the Earth System Grid (ESG): ESG-I, ESG-II, ESG-CET, ESGF

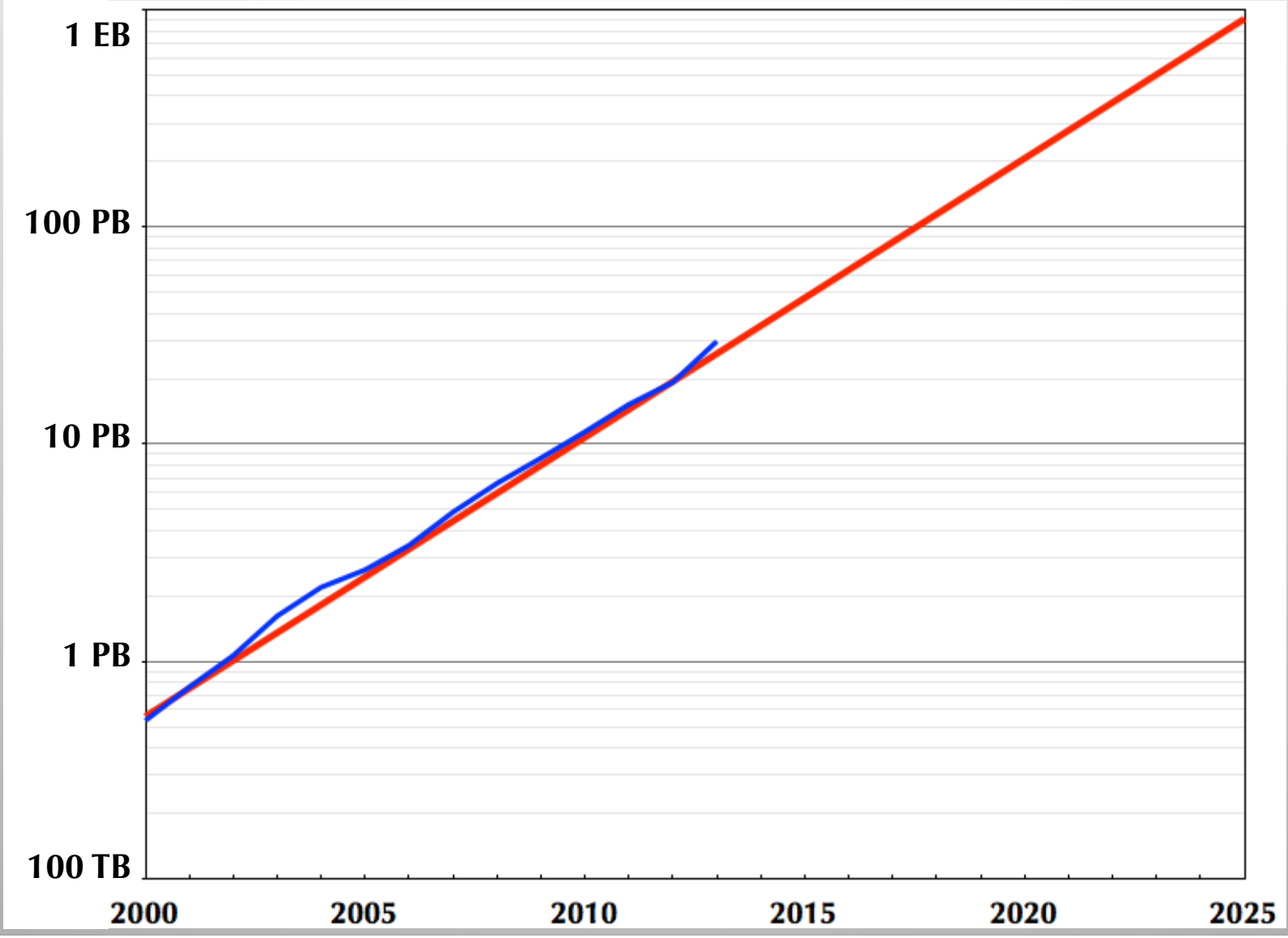


New ESGF technologies developed and deployed by LLNL and our partners

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

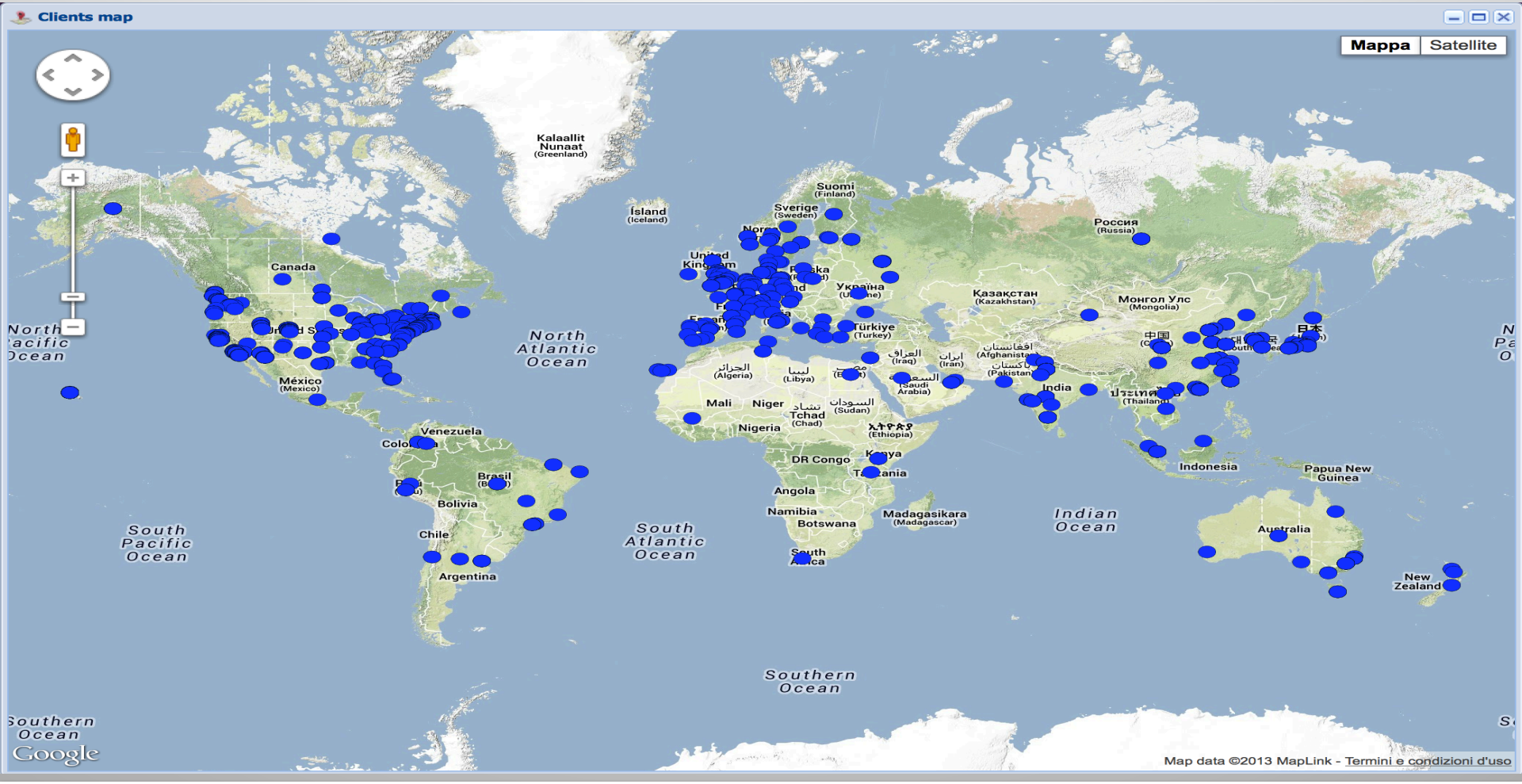


Data challenges: NSF and DOE Community Earth System Model (CESM) archival volume (image courtesy of Gary Strand)



ESGF nodes are deployed internationally

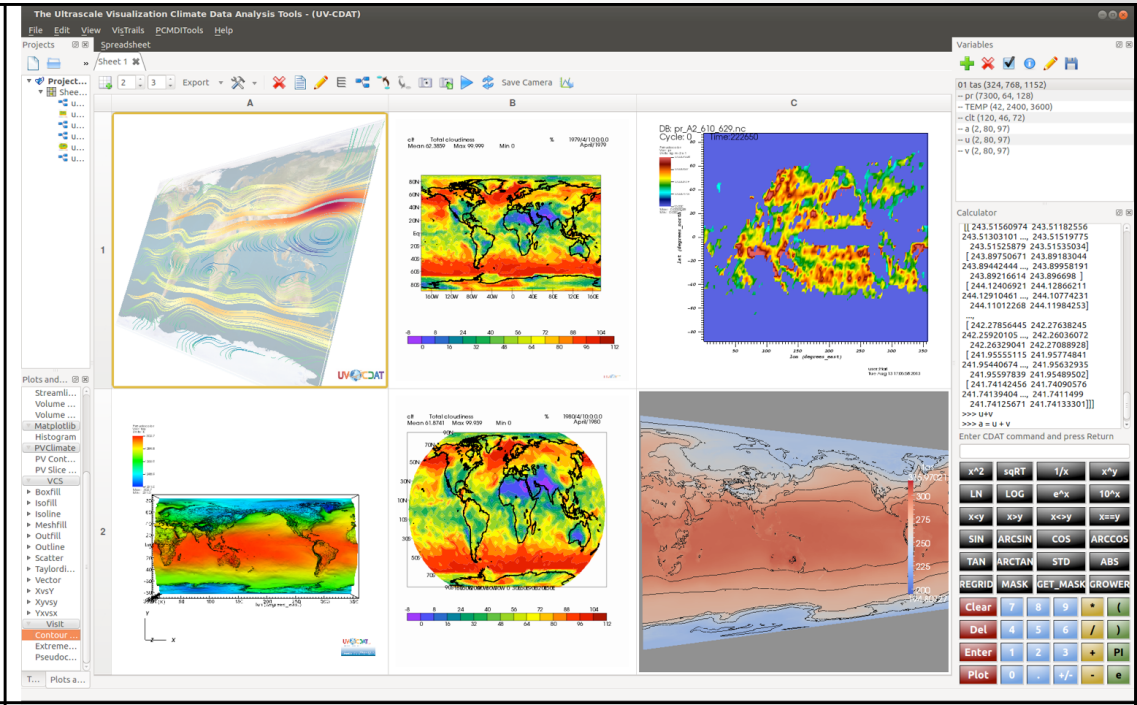
Geo-distribution related to the users that have downloaded CMIP5 data sets from pcmdi9.llnl.gov (over 2,000 publications generated from archive)



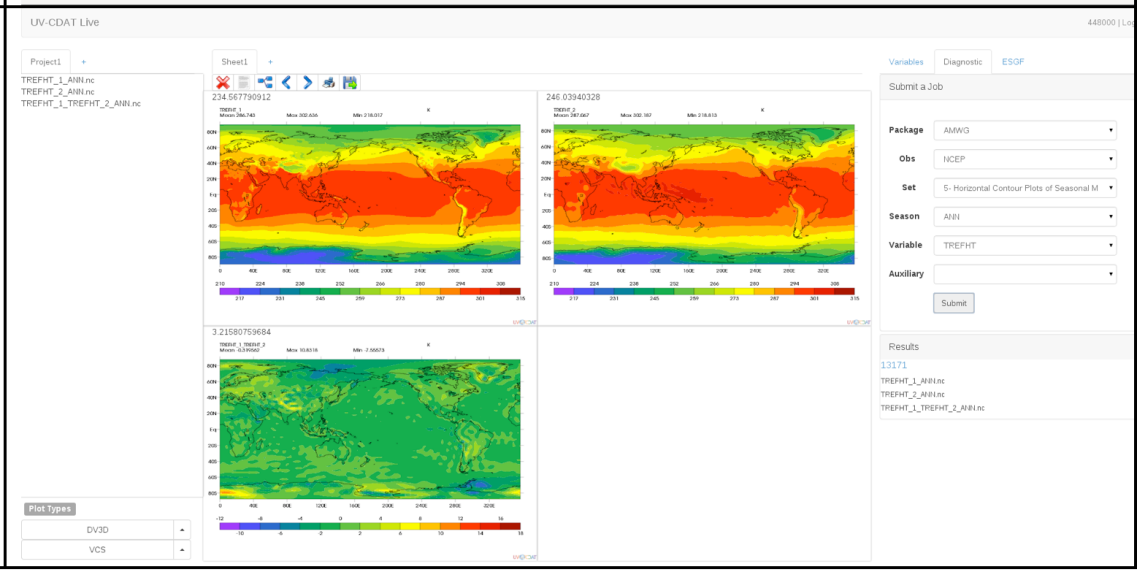
UV-CDAT big data access and displays

- Client analysis tool access
- Thick and smart client analysis tool
- Seamlessly integrate of new components and packages into the framework
- Quickly explore massive amounts of data in unique ways to form new hypotheses and verify simulation data against observational data

UV-CDAT Application (thick-client)



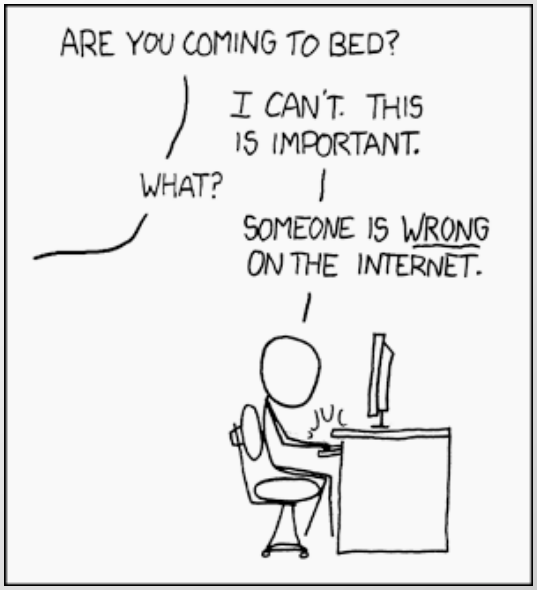
UV-CDAT Live web browser (smart-client)



Future directions and challenges

- **Scale with "Big Data" produced by higher resolution models, satellites, and instruments**
- **Expand server-side functionality**
  - Server-side processing through WPS (climate indexes, custom algorithms); GIS mapping services (for climate change impact studies at regional and local scale); Facilitate model to observations inter-comparison
- **Expand direct client access capabilities**
  - Increased support for OPeNDAP based access; Track provenance of complex processing workflows for reproducibility and repeatability; Controlled Vocabularies
- **Cloud deployment**
  - Instantiate ESGF nodes on demand for short lifetime projects; Environment with elastic allocation of back-end storage and computing resources

Internships



- **Co-op (6+ months Jan – Aug)**
- **Summer (3 months May – Aug)**
- **Work Tasks**
  - **Maintain and update project websites**
  - **Continue development on webshooter**
  - **Standup a user question answer forum (Askbot)**
  - **Continue PCMDI website migration from PHP to Hyde**
  - **Maintain user manuals and API documentation**
- **"Requirements"**
  - **Python, Linux, Git, JavaScript, Jekyll / Hyde, Github, Markdown / HTML / CSS, SQL, QT …**
  - **GPA >3.0, CSCI111, 211, 311, 370(db), 465(web)**
  - **Resume, GitHub account, online work username.github.io, ecst.csuchico.edu/~username …**
  - **https://careers.llnl.gov/students**

Questions?

- **Matthew Harris** harris112@llnl.gov https://github.com/mattben
- **Sam Fries** fries2@llnl.gov https://github.com/chaosphere2112

http://aims.llnl.gov
https://github.com/aims-group
http://uvcdat.llnl.gov
https://github.com/UV-CDAT
http://esgf.org
https://github.com/ESGF
http://cfconventions.org
https://github.com/cf-convention
http://webshootertk.github.io
https://github.com/webshootertk

http://llnl.gov
https://careers.llnl.gov