

# CESD Data Activities

Justin (Jay) Hnilo

Program Manager - Data Informatics

CESD/BER/SC

February 2015



U.S. DEPARTMENT OF  
**ENERGY**

Office  
of Science

Office of Biological  
and Environmental Research

## BER supports diverse research programs

- These programs represent state of the art in several disciplines
- Emerging forefront science is multi-disciplinary
  - See covers of Science and Nature
- Opportunity to empower these advances by integrating our high quality data streams

## CESD data resources

- Carbon Dioxide Information Analysis Center
- Earth System Grid Federation
- Environmental Molecular Sciences Laboratory
- The Atmospheric Radiation Measurement
- Ultra-scale Visualization – Climate Data Analysis Tools



# The Data Challenge

- BER data resides within programs, facilities and ongoing community research projects
- Need for a unified capability to cross talk data outside of individual research domains
- Exponential growth in the volume, acquisition rate, variety, and complexity of scientific data
- Cross-disciplinary science working across techniques, integrating simulation, experimental/observational results complicates data management, analysis, and visualization
- Metadata standards are varied or non-existent
- Too much time is spent rewriting data to a usable form
- Analytic and visualization capabilities not harmonized across user groups

**These all act to complicate the accessibility, availability and usefulness of high quality research data to address multi-disciplinary problems.**

# The solution

A sophisticated data environment as a “one stop shop” to access multiple archives of observed and/or modeled data, common analytical tools, and visualization.

## Features

- Data access and computations are coupled
- Integrate observational, modeled and experimental data
- Offers a spectrum of compute platforms that can be tailored to specific needs
- Data fusion, discovery and intelligent search capabilities
- Data mining and knowledge generation
- Comprehensive visualization and analytic engines
- Modular and scalable in design

# Capability components

## *Data Integration*

- 1) — Integrating complex data generating systems
- 2) — High throughput networks
- 3) — Data collection and management

## *Computational Environment*

- 4) — Data analytics
- 5) — Human computer interface
- 6) — Decision control and knowledge discovery

## Data Integration

Integrating complex data generating systems for simultaneous cross-search and easy access

(BER modeling efforts, field observations, and laboratory experimental results)

- Develop consistent metadata
- Develop libraries to allow cross talk between data.

Input for this component provided by data experts at National Laboratories and DOE Program Managers

# Computational Environment

## Data Analytics

- Leverage existing and future DOE leadership class facilities
- Implement an analysis platform
- Develop visualization/intercomparison tools
- Provide Provenance, automation, and human-computer interaction

Input for this component will require extensive community involvement, via a workshop to be held late in 2015, involving university and Laboratory scientists across the climate and environmental sciences.



## Benefits to the Community

- Open and ready access to CESD integrated data (known quality and format)
- User-friendly access
- Ability to rapidly prototype, run and assess new process algorithms
  - Science across scales
  - Science that merges large multi-dimensional data

# Community input: opportunities, feedback, and defining user requirements

- **CESM Advisory Board Meeting**, June 2014
- Townhall presentation at the **Annual American Meteorological Society Meeting**. December 2015
- The **ESGF governance** structure defined for DOE to be the permanent steering committee chair. January 2015
- A review of the **Annual ESGF & UV-CDAT F2F Conference Report**:  
[http://aims-group.github.io/pdf/2014-ESGF UV-DAT Conference Report.pdf](http://aims-group.github.io/pdf/2014-ESGF_UV-DAT_Conference_Report.pdf)
  - Highlights broad community involvement in this effort
- A white paper was written by data experts at National Laboratories and submitted to BER, on April 29, 2014. This paper and ideas within it will be used to foster community input as we move forward.  
<http://dx.doi.org/10.1109/BigData.2014.7004304>.
- Initial planning for a workshop to be held in late 2015, to describe capability requirements for the computational component of the data environment facility

# Timeline

- Current efforts to extend Earth System Grid Federation's data capabilities to better represent BER's diverse portfolio
- 2015-2017 - focus on metadata interoperability and connectivity, for the ARM facility and the Carbon Dioxide Information Analysis Center, yet within the existing Earth System Grid Federation
- Leveraging the late 2015 workshop we will begin the analysis/computing platform development

## Recent accomplishments

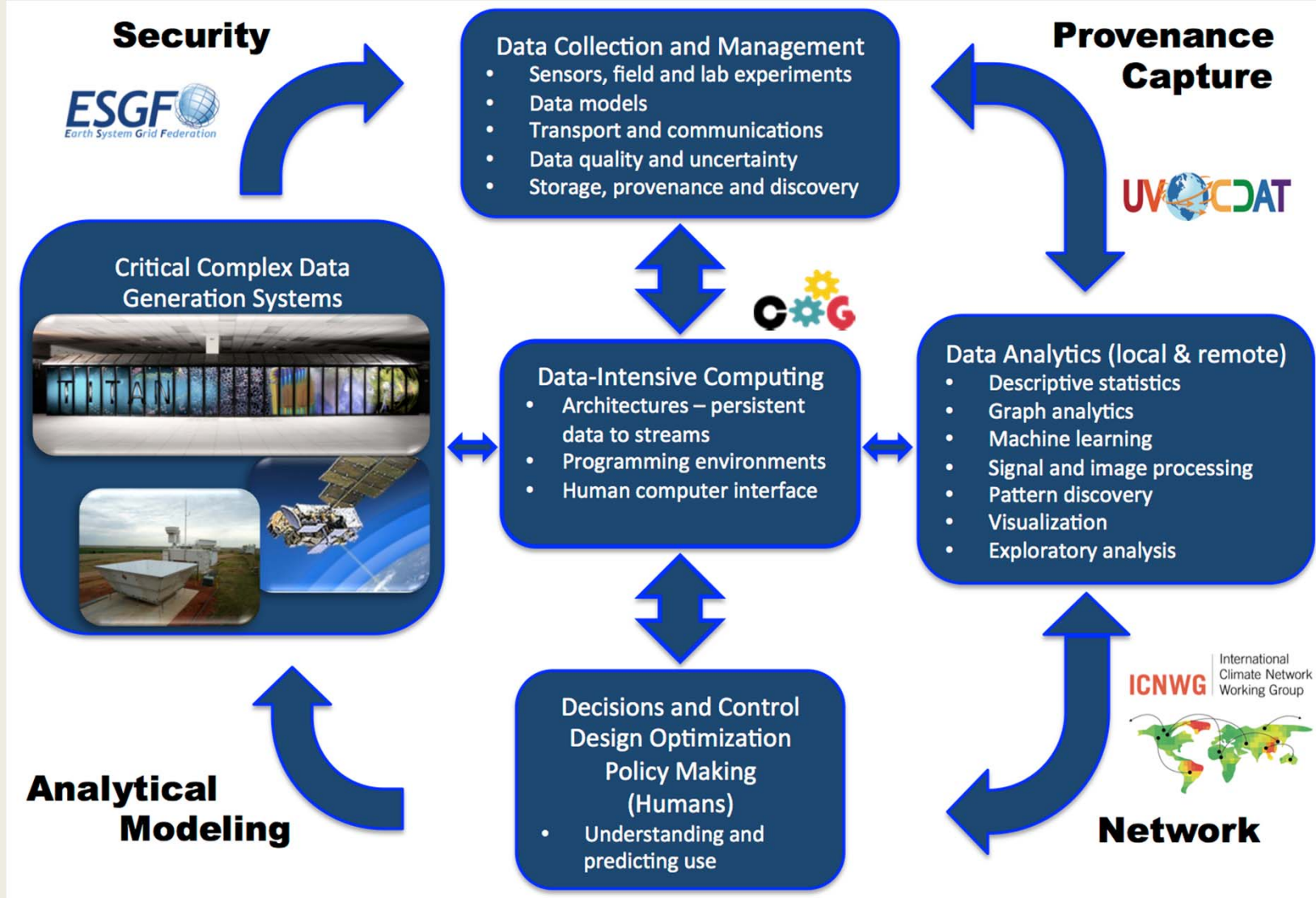
- \* ESGF's User Interface migrated to a content management system and wiki for scientific projects
- \* ESGF's data transfers now incorporate GLOBUS secure data transfers.
- \* ESGF can now store and access multiple data forms
- \* Initiated the development of server side analysis and visualization within ESGF.

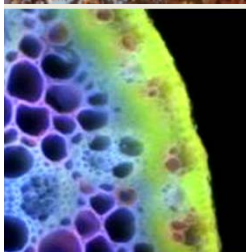
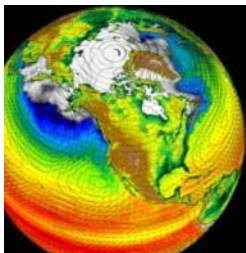
### Awards:

ESGF & UV-CDAT have won three consecutive **Federal Laboratory Consortium (FLC)** awards for "Outstanding Technical Partnership" (2013, 2014, 2015)

ESGF awarded, Enlighten your research global, EYR-global (2014), represent 8 international network organizations.

# CESD's Integrated Data Ecosystem and Workflow





# Thank you

[http://science.energy.gov/~media/ber/berac/pdf/20130221/BERACVirtualLaboratory\\_Feb-18-2013.pdf](http://science.energy.gov/~media/ber/berac/pdf/20130221/BERACVirtualLaboratory_Feb-18-2013.pdf)

[http://science.energy.gov/~media/ber/pdf/ber\\_ltv\\_report.pdf](http://science.energy.gov/~media/ber/pdf/ber_ltv_report.pdf)

[justin.hnilo@science.doe.gov](mailto:justin.hnilo@science.doe.gov)



U.S. DEPARTMENT OF  
**ENERGY**

Office  
of Science

Office of Biological  
and Environmental Research