EESSD - ESS
Cyberinfrastructure Working Groups

Current Scope and White Paper Plans

Working Groups Structure

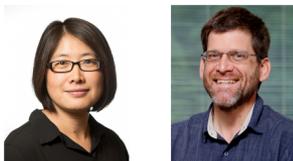*2015 AGU Town Hall & 2016 Working Groups Kickoff*

Executive Committee (EC) & Working Groups:

Data Management (DM)

Danielle Christianson (LBNL)
Terri Velliquette (ORNL)

Model–Data Integration (MDI)

Xingyuan Chen (PNNL)
Forrest Hoffman (ORNL)

Software Engineering and Interoperability (SEI)

David Moulton (LANL)
Ethan Coon (ORNL)

DOE/SC-0178

Building a Cyberinfrastructure for Environmental System Science: Modeling Frameworks, Data Management, and Scientific Workflows

Workshop Report

Environmental System Science

U.S. DEPARTMENT OF ENERGY | Office of Science

Office of Biological and Environmental Research
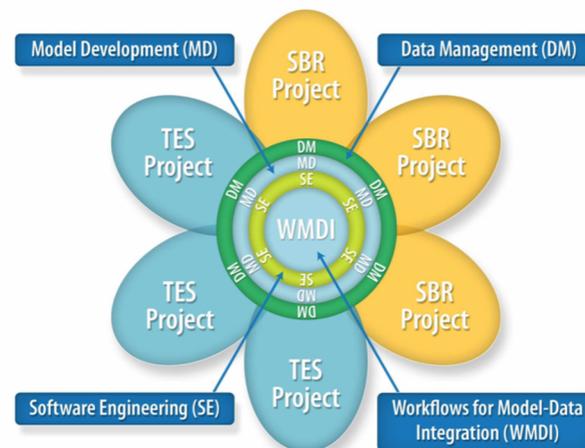
Activity Highlights

- ESS-DIVE Community Projects (Standards development)
- RUBISCO-AmeriFlux Working Group
- Model–Data Integration WG Survey on workflows and tools
- DOE CMIP6 Hackathon for RGMA Projects
- 2019 AGU Fall Meeting Sessions:
  - "Innovation and Exploration of Observations and Earth System Models Using Machine Learning and Big Data Analysis"
- Geospatial analysis and remote sensing
  - Shiklomanov, et al. (2019), Enhancing Global Change Experiments through Integration of Remote-sensing Techniques, *Front. Ecol. Environ.*, 17(4):215–224, doi:10.1002/fee.2031.

Our big highlight each year is our face-to-face Annual Meeting *the day before the ESS-PI Meeting*

Working Group (Virtual) Annual Meeting

½ Day 1:

- New projects
  - IDEAS-Watersheds, ExaSheds, ICoM, InteRFACE
- Early Career
  - Charuleka Varadharajan, James Stegen
- EES-DIVE updates and community projects
- Community Updates
  - IHTM workshop report, CD-MII, ILAMB/IOMB, FATES/NGEE Tropics
- BER Perspectives
  - Community of Visitor Report
  - Challenges with HPC Resources
- Working Group status updates and goals

½ Day 2:

- Breakout sessions ...

ZOOM!

Participation was awesome!

- 70+ total most of the time
- 12 BER representatives
- Several external participants
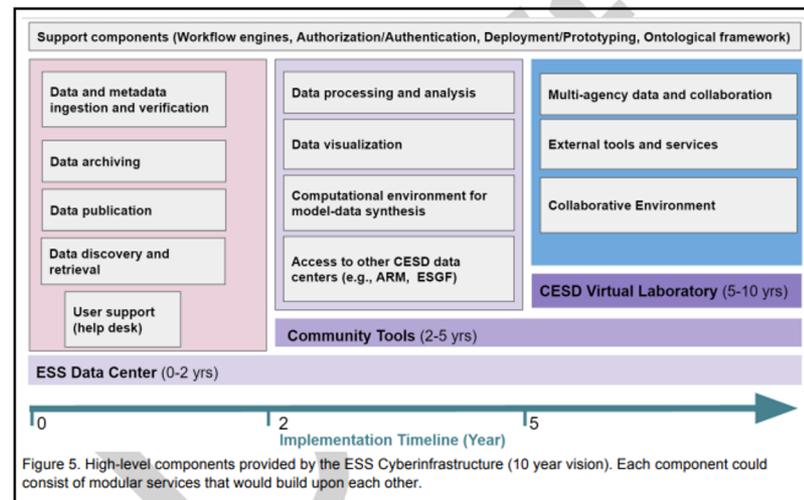
White Paper on Data Infrastructure

Motivation

- *"the innovation most needed is a framework that allows seamless integration of multiscale observations, experiments, theory, and process understanding into predictive models for knowledge discovery"* (BERAC 2013)
- Exponential growth in the amount, variety and complexity of scientific data.
- Significant fragmentation across projects and disciplines still remains.
- A need for a data center that would be a foundational part of a community cyberinfrastructure.

Outcome

- Report used as a reference in the call that led to ESS-DIVE, and contained a phased vision for CI.

Towards a Shared ESS Cyberinfrastructure:
Vision and First Steps

Figure 5. High-level components provided by the ESS Cyberinfrastructure (10 year vision). Each component could consist of modular services that would build upon each other.

https://doesbr.org/BERfiles/Towards_a_Shared_ESS_Cyberinfrastructure.pdf

BER Perspectives - Time to update the White Paper

Open Watershed Science by Design
Leveraging Distributed Research Networks to Understand Watershed Systems

Workshop Report

U.S. DEPARTMENT OF ENERGY | Office of Science
Office of Biological and Environmental Research

DOE/SC-0200

Scientific User Research Facilities and Biological and Environmental Research: Review and Recommendations

A Report from the Biological and Environmental Research Advisory Committee

October 2018

Gary Stacey, Chair
Biological and Environmental Research Advisory Committee (BERAC)

Bruce A. Hungate, Chair
BERAC Subcommittee on Scientific User Research Facilities

U.S. Department of Energy
Prepared by the BERAC Subcommittee on Scientific User Research Facilities Report available online at
https://science.energy.gov/ber/berac/

REPORT OF BER CESD COMMITTEE OF VISITORS

Climate and Environmental Sciences Division Office of Biological and Environmental Research
Office of Science
US Department of Energy

Findings and Recommendations from a Review of Fiscal Years 2016-2018

- HPC Resource Challenges - what is the right mix and approach to ensure BER scientists needs are met
- Software release processes - opportunities to streamline and improve sustainability and productivity

Data Management (DM)

Management and Archival of DOE climate and environmental datasets

- Data **Preservation, Sharing, and Publication**
- Common Data and Metadata **Standards**
- Data **Citation and Attribution**
- Data **Federation** across different data catalogs

Data **Synthesis** across ESS and other relevant Datasets

Development of common **Tools** for data usage

QA/QC, processing, analysis, mining and visualization data to prepare them for use in new research projects.

Leads: Danielle Christianson (LNBL)   Terri Velliquette (ORNL)

Data Management (DM) - Breakout Outcomes

Standards for data interoperability
- Leverage existing, Identify / Prioritize ESS needs, Reusable development workflows, FAIR compliance, Linking persistent identifiers, Implementation (old & new)

Resources for small and large projects
- DOE DM team, Software toolkits, Funding opportunities for data management, Proposal requirements

Enabling infrastructure to discover, access, visualize data
- Federation, Data use and re-use, Software tools for provenance, Governance

**Cultural Shift / Training** to encourage adoption of good DM and data sharing

Compliance with DOE requirements to ensure data sharing and usability

Can we develop cross-WG use cases to drive development and integration?
- Identify / Leverage dependencies across Working Groups
- Examples: Watershed intercomparison, HydroFrame

Model–Data Integration

- Model–data comparison, model evaluation and benchmarking
- Uncertainty quantification (UQ) and data assimilation (DA)
- Management of model results and observational data (with Data Management Working Group)
- Geospatial and remote sensing data analysis
- Data analytics methods and techniques, e.g.,
  - Data mining
  - Machine learning, neural network models
  - Genetic algorithms
  - Visual analytics
- Hybrid ML/process-based models & data-driven models
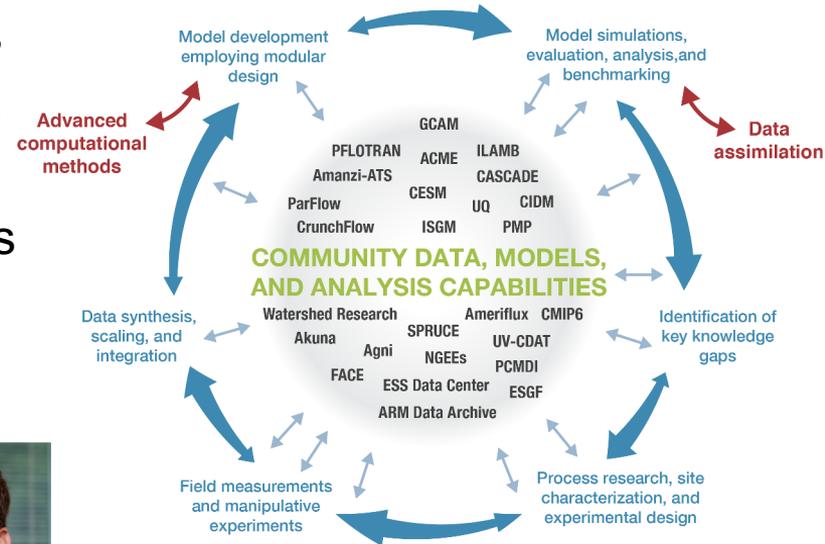
Co-Leads:
Xingyuan Chen (PNNL)

Forrest M. Hoffman (ORNL)

Model–Data Integration - Breakout Outcomes

Survey ➜ Seminars ➜ Workshop

Understand the tools and resources ESS projects find most useful
- Design a new survey about commonly used tools, models, and frameworks

Start up a Model–Data Integration seminar series
- Bring together modelers and measurements experts
- Topics include framework demos, success stories and *failures*, best practices, and *communications*

Cross-Working Group interests and needs
- Topics include archiving model output, versioning data, licensing, standards, software infrastructure
- Cross-Working Group activity, e.g., *Watershed MIP* to drive CI design, produce science papers

Workshops on Model–Data Integration Frameworks and on Scaling
- Modelers and field/lab/remote sensing experts to survey and design toolboxes for ESS activities
- Address scaling challenges and remote sensing with lessons learned and ESS-DIVE engagement

Model–Data Integration Guide
- Document best practices with many use cases

Education/Outreach
- Seminars, tutorial videos for running models, hackathons on data/models, bring models to empiricists

Software Engineering & Interoperability

A high quality community driven open-source software ecosystem of interoperable components that can be assembled in flexible configurations within a common framework supporting ModEx and the Virtual Laboratory:

- integration of legacy and new capabilities across projects
- rigorous – but rapid – testing and validation of model-data integration capabilities
- changing architectures and programming models
- complex multiscale models (coupling, interoperability)
- performing quantitative and formalized UQ
- diverse interdisciplinary teams, and training
- increased scientific productivity
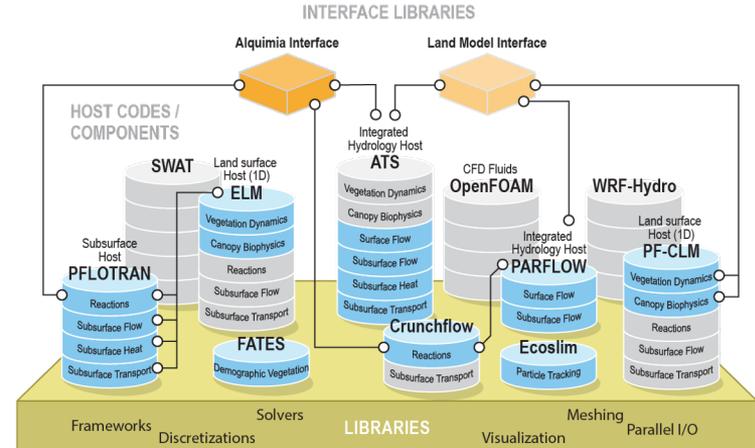
Co-Leads: David Moulton (LANL)   Ethan Coon (ORNL)

Software Engineering & Interoperability - Breakout Outcomes

- **Land Model Interface & Interoperability Workshop**
  - Discussion: identify key users (from both sides of an interface)
  - Are top-down (coarse-grain) and bottom-up (fine-grain) view of process coalescing?

- **Workflows and Software Engineering**
  - Engage with Jupyter developers
  - Explore working with DMI and use cases

- **HPC Resources and Cloud**
  - Create a survey on needs and experience with Cloud
  - Explore containerization as useful tool independent of Cloud
  - Clarify workflows suitable with Cloud

- **Open Source Software Development and Release**
  - Create guidance for communicating release information to PMs, accelerating release
  - Guidance on the code maintainer role, sustainability and business model
  - User community and Training

Looking Forward

Working Groups
- Consider joining, it could seed your next great collaboration

White Paper Timeline
- Draft to share with program managers in mid-summer
- Interested in helping out contact us.

Recordings from the virtual meeting are available to the working groups
- **HINT:** join the working groups and checkout the videos

Questions?