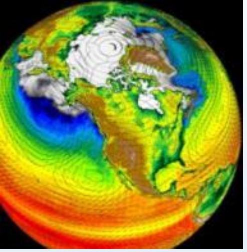
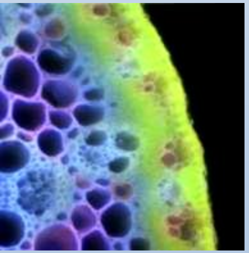
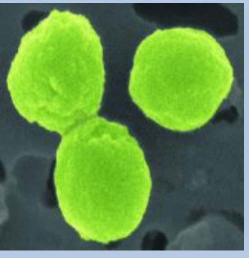


US Department of Energy
Office of Biological & Environmental Research

Cyberinfrastructure Working Group
Virtual Meeting

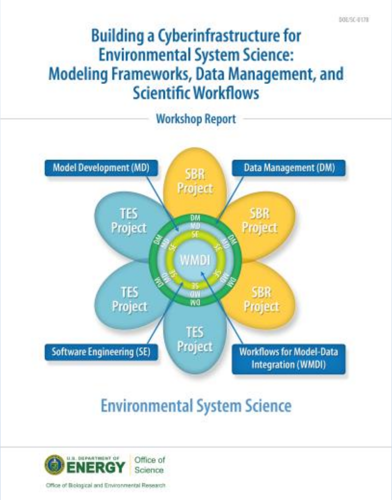
May 11, 2020

ESS Coordinators:

Jay Hnilo, Paul Bayer, Dan Stover, Jennifer Arrigo

Cyberinfrastructure Working Groups for Environmental Systems Science (ESS)

2015 Workshop – Building a Cyberinfrastructure for ESS…
DOE Organizers ESS-team
Workshop Chair David Moulton, LANL



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

CI WGs Meetings

- April 25, 2016 and annually before the ESS PI meeting (~70)
- Organized by the CI WG Executive Committee

Working Groups

- Data Management
- Model - Data Integration
- Software Engineering and Interoperability

Your interest makes ESS science better

CESD - ESS Working Groups – Google Drive

Reading File:
https://drive.google.com/drive/folders/1RmyhTnkrPVMzk_JXPjFgD545J3xkMsui?usp=sharing

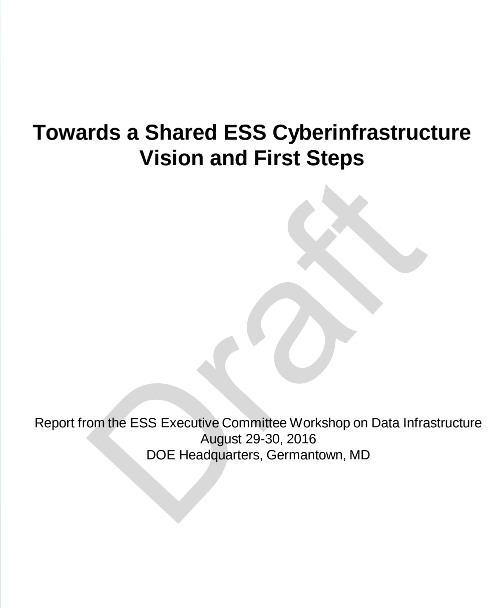
ESS CI WGs – First Product

2016 ESS CI WG Workshop Report

Product from Late August 2016 workshop

Objectives

- Requirements for a Data Center
  - Data ingest, archiving, preservation and sharing
- Thoughts on Community Tools for data-model integration
- Concepts for a Virtual Laboratory for Predictive Modeling



Towards a Shared ESS Cyberinfrastructure
Vision and First Steps

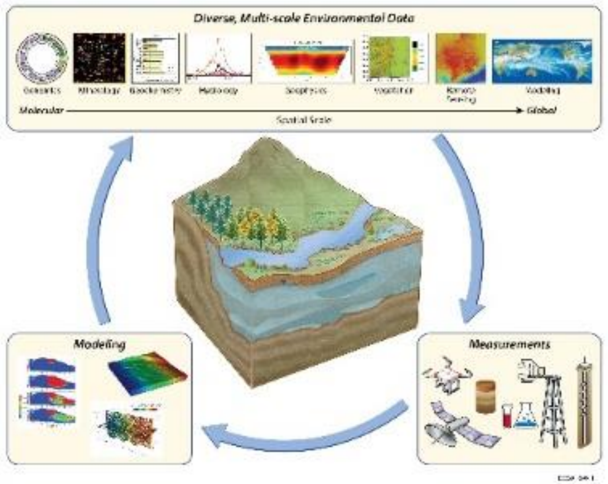
Report from the ESS Executive Committee Workshop on Data Infrastructure
August 29-30, 2016
DOE Headquarters, Germantown, MD

ESS Community Data Infrastructure
(ESS-DIVE)

Environmental Systems Sciences – Data Infrastructure for a Virtual Environment (ESS-DIVE)

- Data services launched on 1 April 2018
- Hosted by LBNL
- Established to provide long-term stewardship and enable broad usage of data from research in the DOE's Environmental System Science (ESS) domain
    - Terrestrial field and lab data
    - Watershed field and lab data
    - Subsurface/groundwater field and lab data
- Proactively engaging with the ESS scientific research community to understand their needs and to adopt or develop standards
- Designed using Findable, Accessible, Interoperable, and Reusable (FAIR) principles
- Data packages receive DOIs
- Includes all relevant data from previous ESS archive

ESGF
External Data
...
ESS-DIVE
AmeriFlux
SPRUCE
C flux Tool
Groundwater Constituent Tool
Spectral Processing Jupyter Notebook



Reference: Varadharajan, C., S. Cholia, C. Snavely, V. Hendrix, C. Procopiou, D. Swantek, W. J. Riley, and D. A. Agarwal (2019), Launching an accessible archive of environmental data, *Eos, 100,* https://doi.org/10.1029/2019EO111263. Published on 08 January 2019.

ESS-DIVE Collaborating Closely with the ESS Community

ESS-DIVE is proactively serving and engaging the ESS community (lab SFAs, projects and university projects)

- Gathers ESS project requirements through:
  - Site visits
  - Monthly webinars
  - Advisory group
- Provides regular demonstrations and tutorials
- ESS-DIVE features & priorities strongly influenced by community needs and input
- Developing inter-repository linkages (e.g., ESGF, USGS)
- Created ability to establish “project spaces”
- Funded 6 projects across the ESS community to develop standards in areas of expertise:
  - ANL – Amplicon data
  - BNL – Leaf physiology
  - ORNL – Comma separated value & File level metadata
  - PNNL – Soil respiration
  - PNNL – Soil respiration
  - SLAC - Sample chemistry

Joan Damerow presenting during the ESS-DIVE SLAC site visit

Charu Varadharajan presenting at the AGU Data Fair

http://ess-dive.lbl.gov/community-projects/

What might be needed beyond ESS-DIVE?

ESS-DIVE is serving the ESS scientific community

- Proactively engaging with the ESS community to understand their needs, adopting/developing data and metadata standards
- Advancing data management and service needs for the ESS community

✅

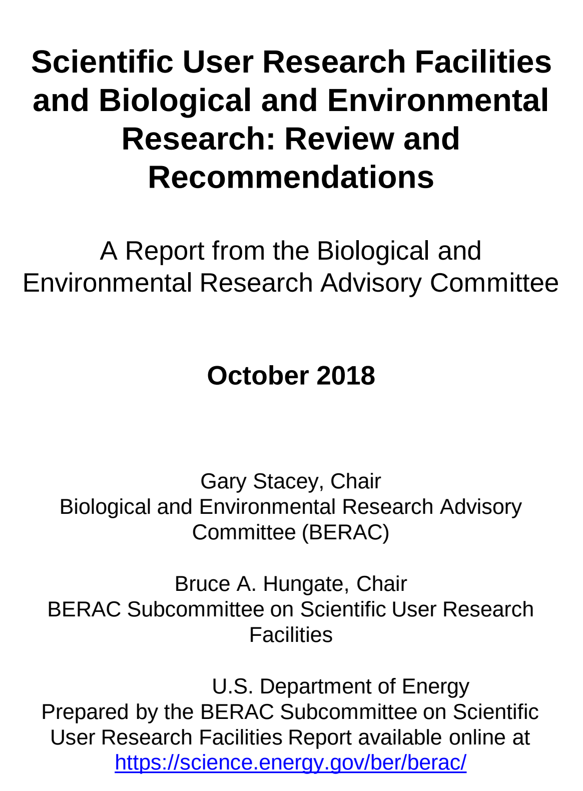
Are we done?

- Why have the CI WGs continued to meet?
- Might there be some ESS/EESSD programmatic developments or other external developments that might be relevant to CI in ESS?

BER Advisory Committee (BERAC)
Recommendations

2018 BERAC User Facilities Report

- Issued October 2018
- In the Google Docs Reading File
- Same GCs as those in Nov 2017 BERAC GC Report
  - – Earth and Environmental Systems
  - – Microbial to Earth System Pathways
  - – Computation and Data Analysis
- **Some Recommendations:**
  - – Obtain mid-range compute resources (2 GCs)
  - – Develop algorithms to harness current/future architectures to model coupled systems and analyze extreme-scale data
  - – Multi-scale modeling framework to connect/inform experimentation and modeling
  - – Promote Better Scientific Software portal (BSSw.io)
  - – Improve software development/management processes through Productivity and Sustainability Improvement Plans



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

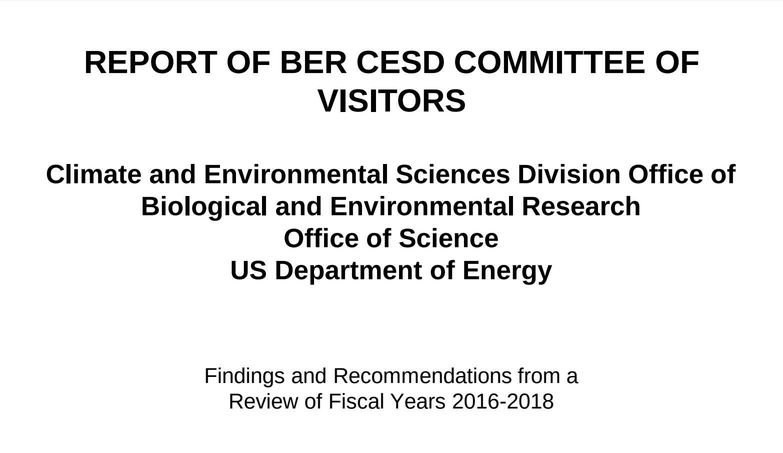
U.S. Department of Energy
Prepared by the BERAC Subcommittee on Scientific User Research Facilities Report available online at
https://science.energy.gov/ber/berac/

https://science.osti.gov/ber/berac/Reports

2019 Committee of Visitors (COV) Review of BER's Earth and Environmental Systems Sciences Division (EESSD)

2019 CESD (EESSD) COV Report

- Summer 2019 - Jim Hack
- Fall BERAC – presentation
- In Google Docs reading file
- Spring BERAC – approved draft report
- BER/EESSD preparing response
- **Some Recommendations**
  - – Strategic plan for harmonizing data collection, archiving and data access manipulation capabilities.
  - – A strategy for integrated modeling across scales.
  - – Align observational and modeling components
  - – EESSD should regularly asses computational needs & collect input from the community
  - – Streamline allocation of computational resources to funded projects



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

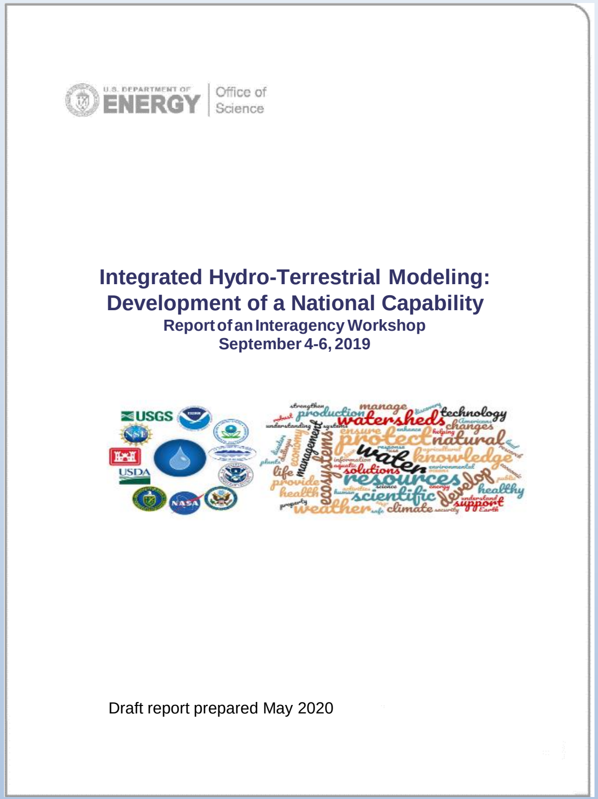
CI WG Discussion/Suggestions?

- Jay
- NGEEs, IDEAS–Watersheds
- Promote ModEx
- NERSC requests/allocations, mid-range computing at EMSL
- Mid-range compute at EMSL

Integrated Hydro-Terrestrial Modeling (IHTM) Workshop
IHTM: Development of a National Capability

2019 IHTM Workshop

- Held September 2019 at NSF Headquarters
- Interagency leads
  - – Bob Vallario (EESSD)
  - – David Lesmes (USGS)
  - – Tom Torgersen (NSF)
- Co-Chairs
  - – Tim Scheibe (PNNL)
  - – Harry Jenter (USGS)
  - – Efi Foufoula-Georgiou (UCI)
- Three “Use Cases” – Water Subcabinet
  - – Water availability in the West
  - – HABs, Hypoxia and Nutrient Loading
  - – Flooding and Extreme Weather-Related Water Hazards
- Scheibe/Lesmes Presentation at 2019 AGU



U.S. DEPARTMENT OF ENERGY | Office of Science

Integrated Hydro-Terrestrial Modeling:
Development of a National Capability
Report of an Interagency Workshop
September 4-6, 2019

Draft report prepared May 2020

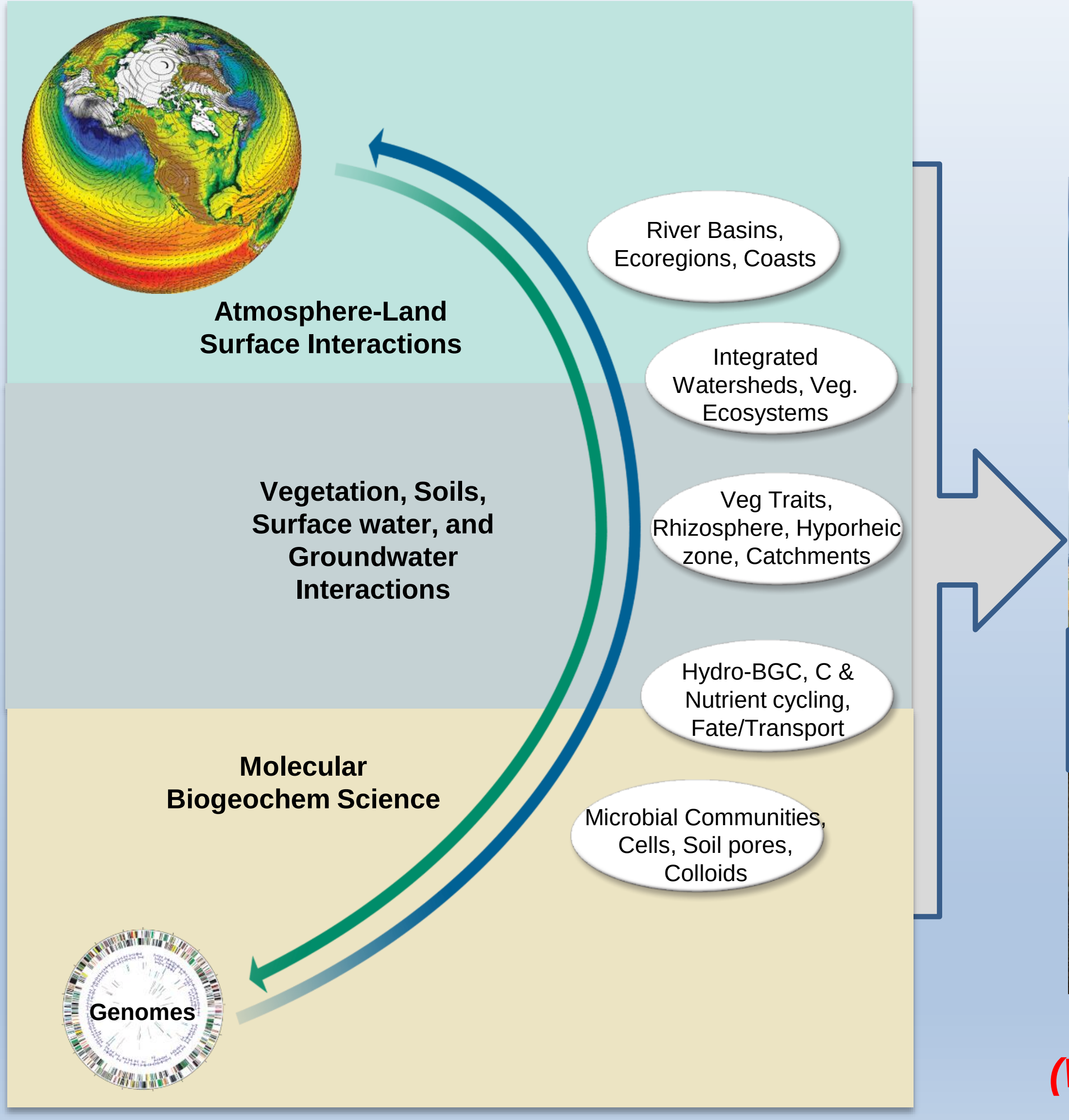
DOE, EPA, NASA, NOAA, NSF, USACE, USBR, USDA, USGCRP, USGS

SC Cloud Computing Strategy

Spring 2020 Request from Chris Fall, SC Director

- Why - Explanation for the DOE Secretary, Congress, OMB, vendors
- What – SC Strategy/Approach for Cloud Computing
- Who – Ramana Madupu is the BER POC; Jay for ESS and EESSD. Both part of the SC Data Management Working Group
- Applicable to – BER user facilities (EMSL, ARM, JGI) and E3SM
- What - How SC user facilities and research programs determine how and when to employ various computational resources (edge, cluster (physical/virtual), cloud (commercial/private) and HPC
- Data Call now:
  - Computing strategy for your facility/program? Research or production?
  - Costs for cloud vs. purchase/leasing hardware, and how current?
  - Using any commercial cloud services?
  - Challenges and opportunities for cloud for your facility/program?

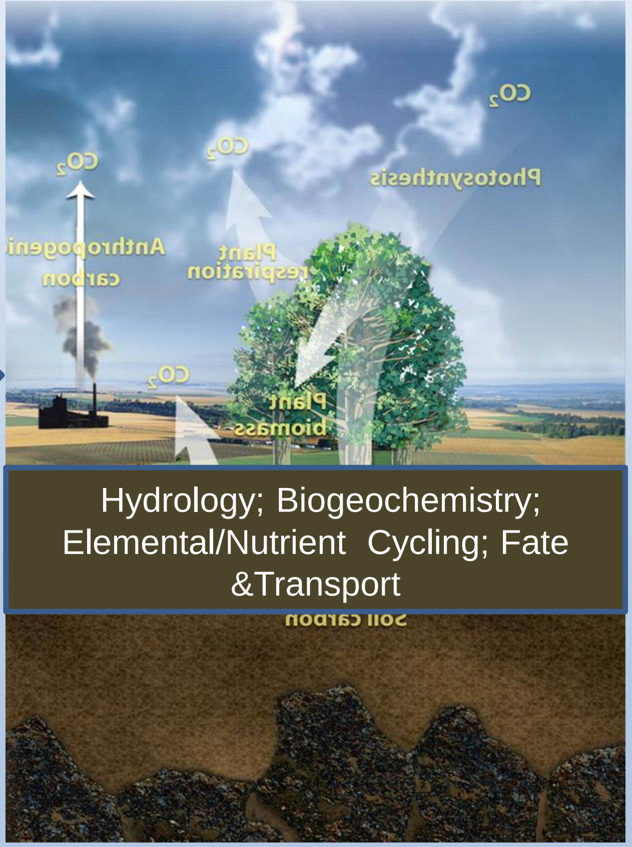
Scales of BER Systems Science & ESS Scope


Atmosphere-Land Surface Interactions
Vegetation, Soils, Surface water, and Groundwater Interactions
Molecular Biogeochem Science
Genomes
River Basins, Ecoregions, Coasts
Integrated Watersheds, Veg. Ecosystems
Veg Traits, Rhizosphere, Hyporheic zone, Catchments
Hydro-BGC, C & Nutrient cycling, Fate/Transport
Microbial Communities, Cells, Soil pores, Colloids

ESS Scope

TES

(Canopy=>Soils, Roots)



Hydrology; Biogeochemistry; Elemental/Nutrient Cycling; Fate &Transport

SBR

(Watersheds=>Soils, Sediments, Groundwater)

Earth and Environmental Systems Sciences Division Strategic Plan

Earth and Environmental Systems Sciences Division

Strategic Plan

Vision: Develop an improved capability for Earth system prediction on seasonal to multi-decadal time scales to inform the development of resilient U.S. energy strategies.

High level Grand Challenges

- **Integrated water cycle** – processes involving atmospheric, terrestrial, oceanic and human system components and their interactions and feedbacks across local, regional and global scales.

- **Biogeochemistry** – coupled biogeochemical processes and cycles across spatial and temporal scales by investigating natural and anthropogenic interactions and feedbacks.

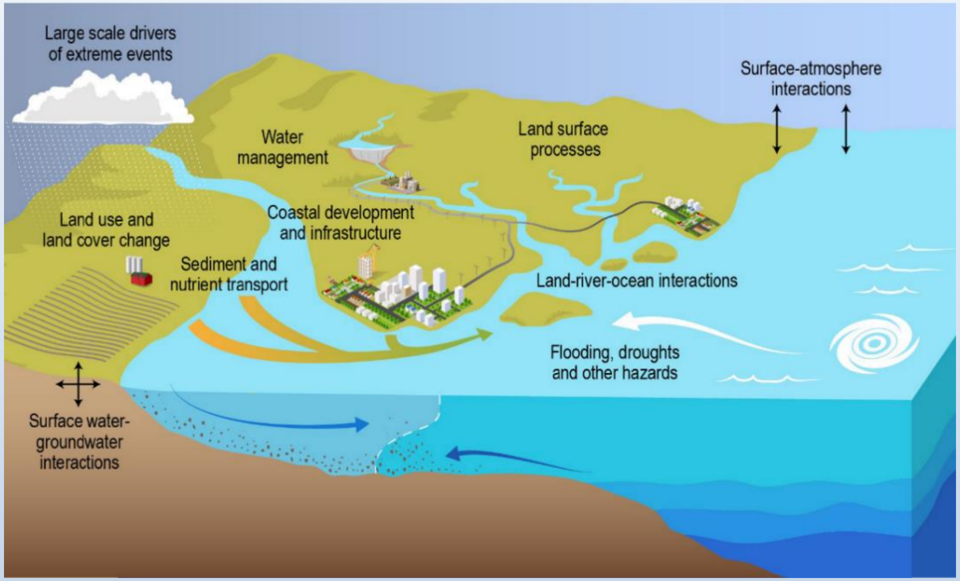
- High Latitudes – quantify the drivers, interactions and feedbacks among the high latitude components and between the high latitudes and the global system.

- Drivers/Responses in the Earth System – next generation understanding of Earth system drivers and their effects on the integrated Earth-energy-human system.

- **Data-Model Integration –** develop a broad range of interconnected infrastructure capabilities and tools that support integration and management of models, experiments, and observations across a hierarchy of scales and complexity.

Integrated Coastal Modeling (ICoM) Project

"Deliver a robust predictive understanding of coastal evolution that accounts for the complex, multiscale interactions among physical, biological, and human systems."


Large scale drivers of extreme events
Water management
Land surface processes
Surface-atmosphere interactions
Land use and land cover change
Coastal development and infrastructure
Sediment and nutrient transport
Land-river-ocean interactions
Flooding, droughts and other hazards
Surface water-groundwater interactions

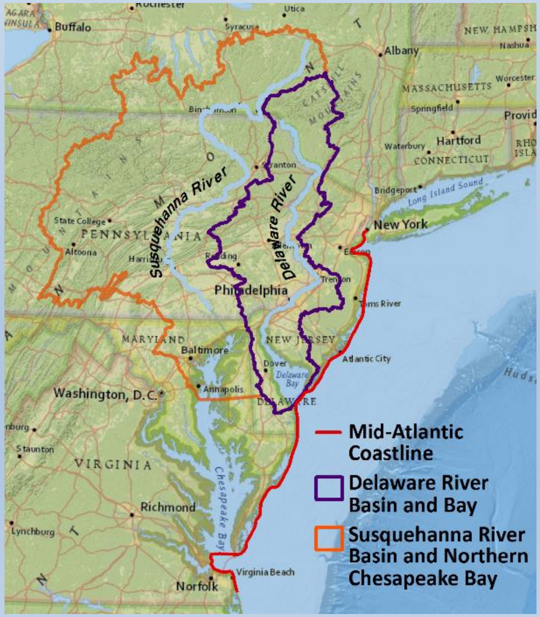
- **Pacific Northwest National Laboratory led multi-institutional team** (LANL a strong participant)… >40% funding awarded by PNNL to others

- **Mid-Atlantic regional focus** … existing DOE capabilities, complex systems interactions, extensive data, and converging interagency activities

- **$16.2M** over three years ($5.4M/yr)

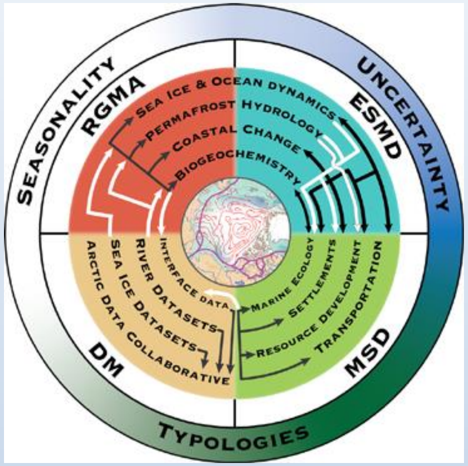
- **A "federated" approach** spanning four distinct program areas within DOE's CESD; requires foundational work in each area and substantial crosscut modeling work.

- **Informs potential follow-on observational and experimental work.**



Interdisciplinary Research for Arctic Coastal Environments (InteRFACE) Project

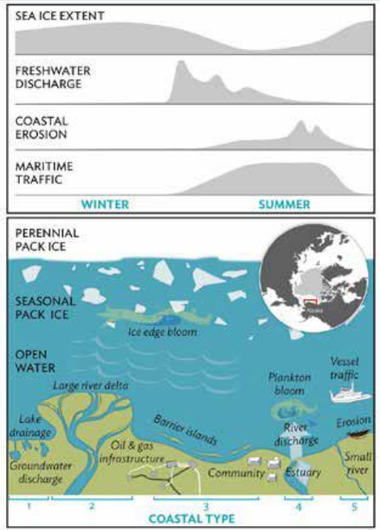
“Quantify and reduce uncertainties in the fundamental understanding of the magnitude, rate and patterns of change along the Arctic coast.”



- **Los Alamos National Laboratory led multi-institutional team** (4 labs and UA Fairbanks)

- **Developing and integrating predictive capabilities** … to simulate feedbacks that will determine the likely trajectory(ies) of the coupled land-ocean-sea Ice-atmosphere-human interface in the Arctic

- **Data plus multi-scale/multi-physics models** … to quantify the timing, rate, patterns and uncertainty in projected change in both human and natural systems of the Arctic

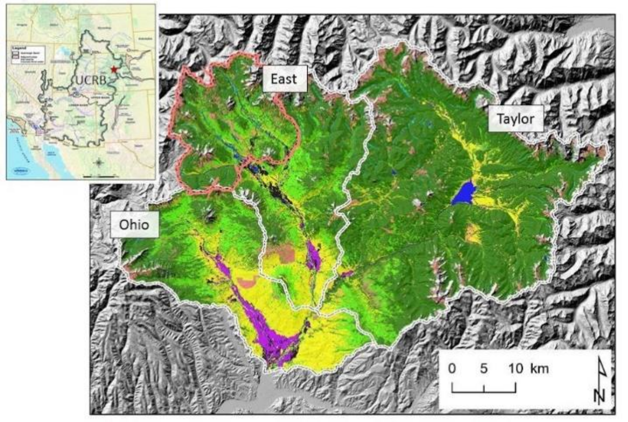


INTERFACE Los Alamos NATIONAL LABORATORY EST.1943 OAK RIDGE National Laboratory Pacific Northwest NATIONAL LABORATORY Sandia National Laboratories UAF UNIVERSITY OF ALASKA FAIRBANKS

ExaSheds Pilot Project

ExaSheds Pilot: To advance Watershed System Science using Machine Learning and Data-Intensive Extreme-Scale Simulation

- Pilot initiated in 2019, 4 labs
- Co-led by C. Steefel (LBNL) and Scott Painter (ORNL)
- Exploring strategies for learning-assisted simulation
  - Development of model inputs from sparse, coarse and indirectly-related information
  - Hybridization of process-resolving simulations and ML
- Working with data from:
  - East River, Colorado
  - Upper Colorado Water Resources Region
  - Continental U.S.
- Adapting DOE-developed watershed simulation tools to leadership-class computer architectures



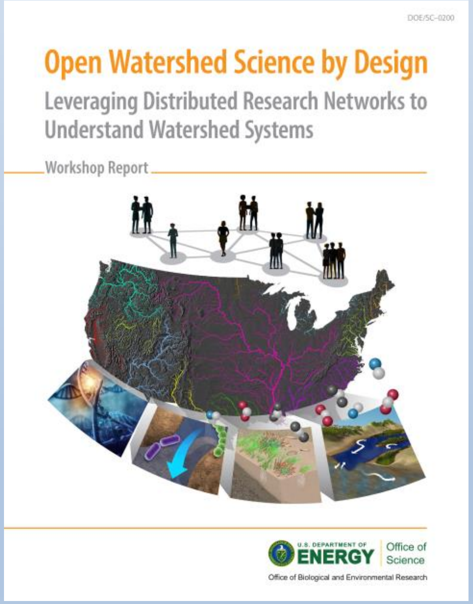
ExaSheds

Open Watershed Science by Design & WHONDRS

- SBR Workshop – Jan 2019
- Leveraging Distributed Research Networks to Understand Watershed Systems
- James Stegen, Kelly Wrighton, Eoin Brodie
- In Google Docs reading file
- Presented at the Spring 2019 BERAC meeting & 2019 AGU
- Recommendations
  - Promoting FAIR data principles
  - Introduced ICON data principles
    - Integrated
    - Coordinated
    - Open
    - Networked
- Combines open science data principles with design thinking techniques
- Five use cases to do together what is not possible alone
  - WHONDRS example
  - Reaction-scale
  - Watershed-scale
  - Basin-scale
  - Multi-scale

WHONDRS

Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems



DOE/SC-0200

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

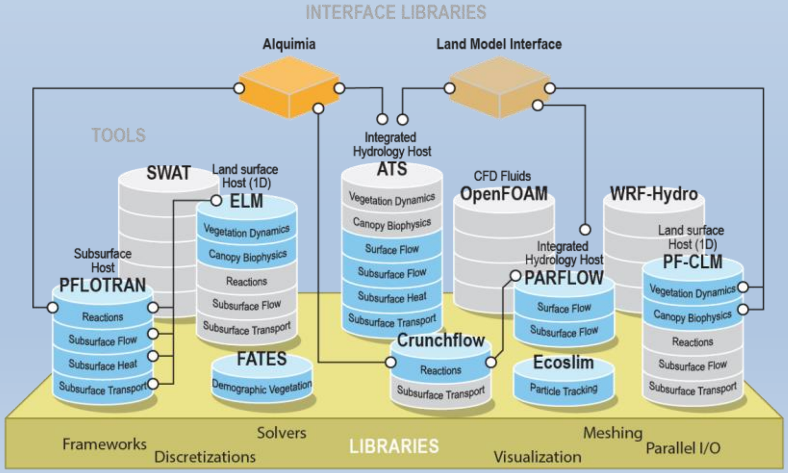
IDEAS – Watersheds Project

IDEAS – Interoperable Design of Extreme-scale Application Software

- Joint SBR/ASCR funding
- 2014-2018
- A software ecosystem of interoperable components, built on 3 SBR/TES use cases, started with "libraries"
- Spawned IDEAS-ECP, xSDK4ECP and IDEAS-Watersheds
- SW Productivity & Sustainability Plans

IDEAS – Watersheds

- Reviewed in 2019, 5 labs and 2 universities
- Led by D. Moulton (LANL)
- From Silos to an interoperable systems of codes/code components
- Enable integration of hydro-biogeochemical process modeling across scales
- Designed to address fragmented codes and changing HPC architectures
- Focused on specific several scales
  - Reaction-scale
  - Watershed-scale
  - CONUS-scale
  - Multi-scale
- In partnership with all 6 SBR SFAs

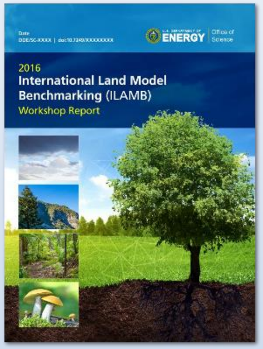


ILAMB, IOMB & PEcAn

ILAMB – International Land Model Benchmarking Project

A community coordination activity created to:

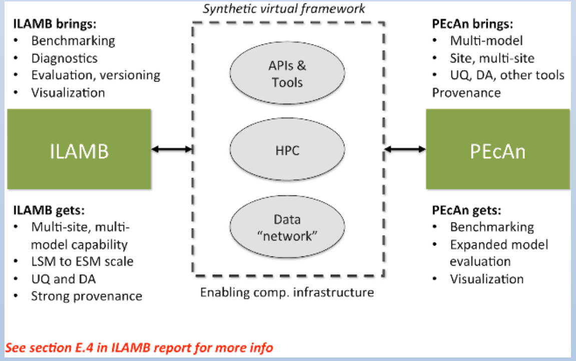
- **Develop internationally accepted benchmarks** for land model performance by drawing upon collaborative expertise
- **Promote the use of these benchmarks** for model intercomparison
- **Strengthen linkages between experimental, remote sensing, and Earth system modeling communities** in the design of new model tests and new measurement programs
- **Support the design and development of open source benchmarking tools**
- https://www.ilamb.org/



Date
DOE/SC-XXXX | doi:10.7249/XXXXXXXXX
U.S. DEPARTMENT OF ENERGY | Office of Science
2016
International Land Model Benchmarking (ILAMB)
Workshop Report

PEcAn – Predictive Ecosystem Analyzer

- An integrated ecological bioinformatics toolbox and data assimilation system that synthesizes information contained in ecological models, data, and expert knowledge
- Enables users to run ecosystem models, as well as a suite of R packages for model-data fusion
- http://pecanproject.github.io/



See section E.4 in ILAMB report for more info

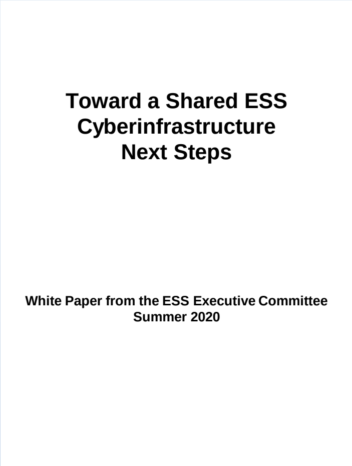
IOMB - International Ocean Model Benchmarking package

- Evaluates ocean biogeochemistry results compared with observations (global, point, ship tracks)
- Scores model performance across a wide range of independent benchmark data
- Leverages ILAMB code base, also runs in parallel
- Built on python and open standards
- Is also open source and will be released soon

A White Paper from the ESS CI WGs is Needed

Requesting an ESS CI WGs White Paper

- Continue to advance data management for ESS
  - Approaches to enhance ESS-DIVE functionality/use and connectivity/federation with other agency data systems
  - Consider open science by design broadly within ESS
- Thoughts on enhancing model-data integration
  - Virtually bringing data to models – PEcAn, CD-MII
  - And vice versa?
- Software Engineering
  - IDEAS-Watersheds, ExaSheds plus NGEEs (including FATES) – cross-lab projects
  - ICoM and InteRFACE – cross-program projects
- Inclusion of hardware/cloud needs
  - Mid-range compute (and data storage/cloud?)
- Consider recommendations from BERAC report &, EESSD (CESD) COV report
- Mid- to late-summer 2020



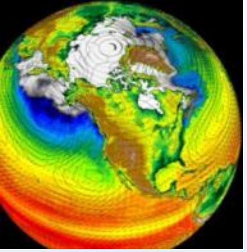
Toward a Shared ESS
Cyberinfrastructure
Next Steps

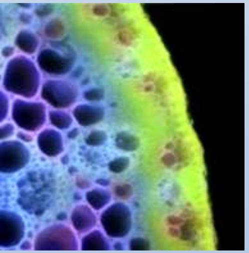
White Paper from the ESS Executive Committee
Summer 2020

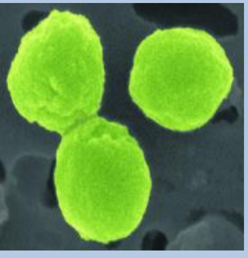
Questions?

Jay Hnilo, EESSD
Justin.hnilo@science.doe.gov

Paul Bayer, EESSD
paul.bayer@science.doe.gov





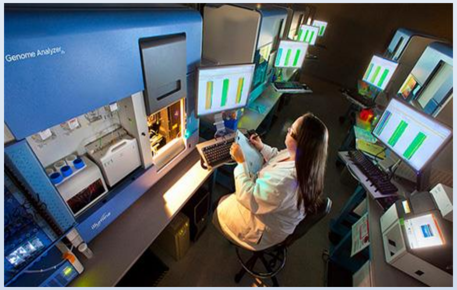


Backup Slides

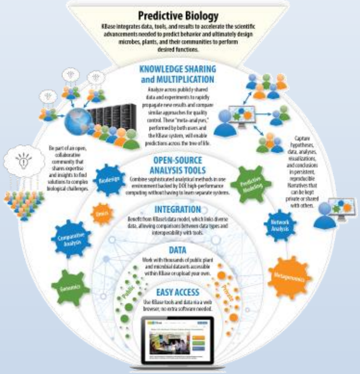
BER User Facilities/Community Resources

EMSL, JGI, KBase

EMSL enables scientists to use multiple combinations of premier experimental and modeling capabilities to obtain a mechanistic understanding of physical, chemical, and intra- and inter-cellular processes and interactions.



JGI provides genome sequencing, genome data acquisition, and genome analyses in support of DOE mission needs in bioenergy, carbon cycling and biosequestration, and biogeochemical processes.



Predictive Biology

KBase integrates data, tools, and results to accelerate the scientific advancements needed to predict behavior and ultimately design microbes, plants, and their communities to perform desired functions.

KNOWLEDGE SHARING and MULTIPLICATION

Analyze across publicly shared data and experiments to rapidly propagate new results and compare similar approaches for quality control. These "meta-analyses," performed by both users and the KBase system, will enable predictions across the tree of life.

Be part of an open, collaborative community that shares expertise and insights to find solutions to complex biological challenges.

Capture hypotheses, data, analyses, visualizations, and conclusions in persistent, reproducible Narratives that can be kept private or shared with others.

OPEN-SOURCE ANALYSIS TOOLS

Combine sophisticated analytical methods in one environment backed by DOE high-performance computing without having to learn separate systems.

INTEGRATION

Benefit from KBase's data model, which links diverse data, allowing comparisons between data types and interoperability with tools.

DATA

Work with thousands of public plant and microbial datasets accessible within KBase or upload your own.

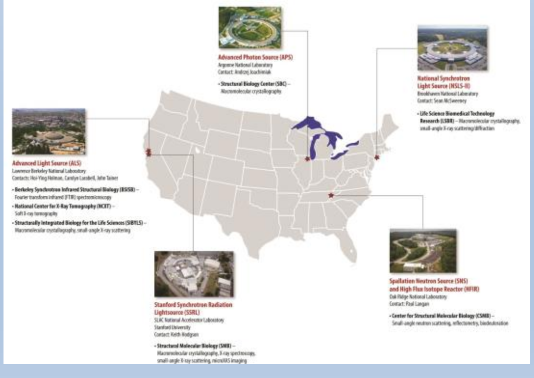
EASY ACCESS

Use KBase tools and data via a web browser, no extra software needed.

KBase (Systems Biology Knowledgebase) is a large-scale bioinformatics system that enables users to upload their own data, analyze it (along with collaborator and public data), build increasingly realistic models, and share and publish their workflows and conclusions.

FICUS
Facilities Integrating
Collaborations for User Science

FICUS is a mechanism that enables scientists to access two Office of Science User Facilities via a single proposal (e.g., EMSL and JGI).



Advanced Photon Source (APS)
Argonne National Laboratory
Contact: Andrzej Joachimiak

- **Structural Biology Center (SBC)** – Macromolecular crystallography

National Synchrotron Light Source (NSLS-II)
Brookhaven National Laboratory
Contact: Sean McSweeney

- **Life Science Biomedical Technology Research (LSBR)** – Macromolecular crystallography, small-angle X-ray scattering/diffraction

Advanced Light Source (ALS)
Lawrence Berkeley National Laboratory
Contacts: Hoi-Ying Holman, Carolyn Larabell, John Tainer

- **Berkeley Synchrotron Infrared Structural Biology (BSISB)** – Fourier transform infrared (FTIR) spectromicroscopy
- **National Center for X-Ray Tomography (NCXT)** – Soft X-ray tomography
- **Structurally Integrated Biology for the Life Sciences (SIBYLS)** – Macromolecular crystallography, small-angle X-ray scattering

Stanford Synchrotron Radiation Lightsource (SSRL)
SLAC National Accelerator Laboratory
Stanford University
Contact: Keith Hodgson

- **Structural Molecular Biology (SMB)** – Macromolecular crystallography, X-ray spectroscopy, small-angle X-ray scattering, microXAS imaging

Spallation Neutron Source (SNS) and High Flux Isotope Reactor (HFIR)
Oak Ridge National Laboratory
Contact: Paul Langan

- **Center for Structural Molecular Biology (CSMB)** – Small-angle neutron scattering, reflectometry, biodeuteration

- See BERStructuralBioPortal.org

BER supports capabilities at the DOE synchrotron and neutron User Facilities. Free access is available for molecular- through tissue-level spectroscopic, scattering and imaging capabilities.