The Cyberinfrastructure Working Groups: *Community Engagement on Data Models, and Workflows for Catalyzing ModEx*

Annual Meeting Highlights

EESSD - ESS Cyberinfrastructure Working Groups Annual Meeting
April 14 , 2025.

Working Groups Structure

2015 AGU Town Hall & 2016 Working Groups Kickoff

Executive Committee (EC) and Working Groups:

Data Management (DM)


Danielle Christianson (LBNL)
Terri Velliquette (ORNL)

Model–Data Integration (MDI)

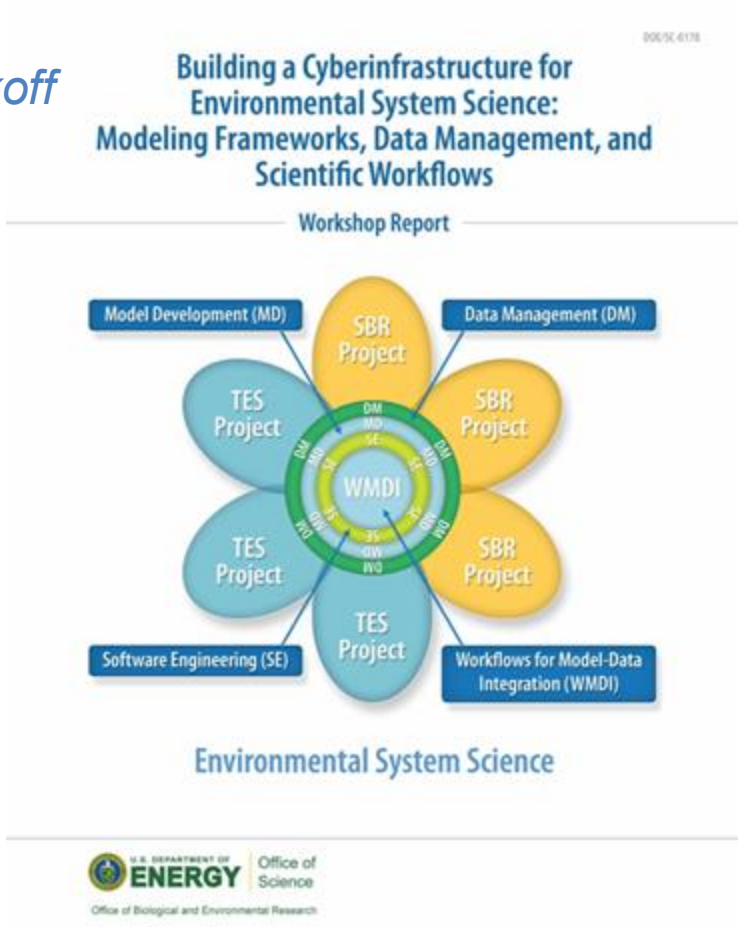

Xingyuan Chen (PNNL)
Forrest Hoffman (ORNL)

Software Engineering and Interoperability (SEI)


David Moulton (LANL)
Greg Lemieux (LBNL)

Computing Infrastructure

Lee Ann McCue (PNNL)



https://doesbr.org/ESS-WorkingGroups/ESSWG_WorkshopReport-final.pdf

Working Groups Annual Meeting

- Participation:
  - 40 in-person, 15 virtual on Zoom
  - 5 BER representatives
- Hybrid Logistics:
  - Google Docs
- Agenda Highlights:
  - Quick recap of recent activities
  - ESGF and ESS-DIVE Updates
  - Community Flash talks
  - What's already possible
  - Data integration across BSSD/EESSD
  - Accelerating ModEx, co-design breakout

Working Group Activity Highlights

- Data Management has quarterly meetings, topics include
  - Tools for implementing Reporting Formats
  - ESS-DIVE Annual Data Workshop
  - Experiences & Improvements to Archiving Model Data Products
- Model Data Integration helped support two activities,

2024 ESS-DIVE Partner Projects to Support Data Standardization and Curation

ESS-DIVE Data Curation Support

Kim Ely
LBNL

Harmonization and Usability of Hydrologic Monitoring and Soil, Sediment, and Water Chemistry Reporting Formats

Amy Goldman
PNNL

Brieanne Forbes
PNNL

Improving Advanced Terrestrial Simulator (ATS) Model Data Managing and Archiving Standards

Ethan Coon
ORNL

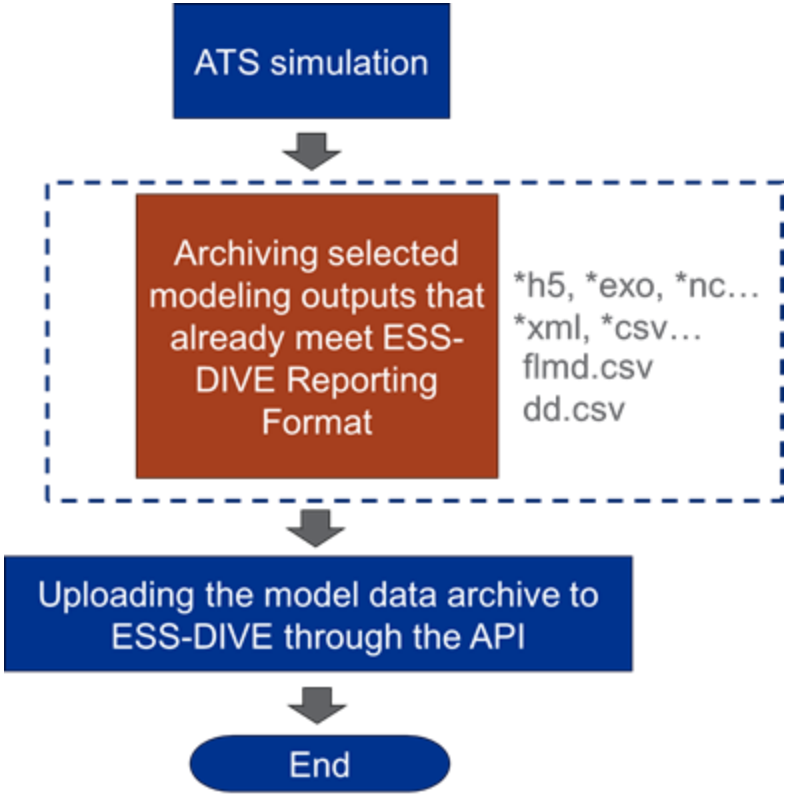
Zhi Li
PNNL

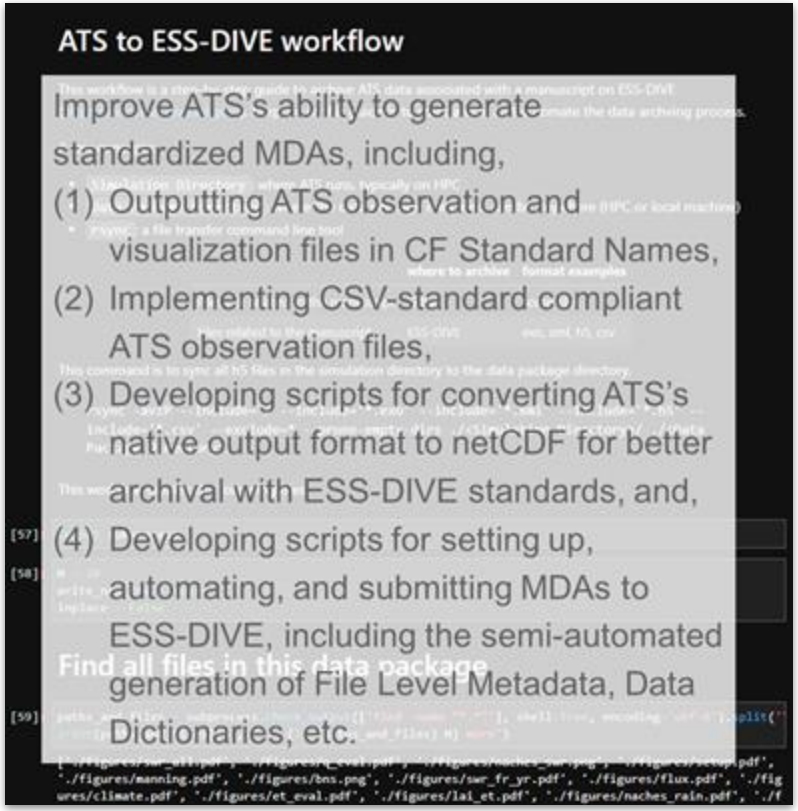
A Workflow and Reporting Format for Processing Environmental Sensor Data and Automated Generation of ESS-DIVE Compliant Metadata; Revision of Soil Respiration Reporting Format

Stephanie Pennington
PNNL

Ben Bond-Lamberty
PNNL

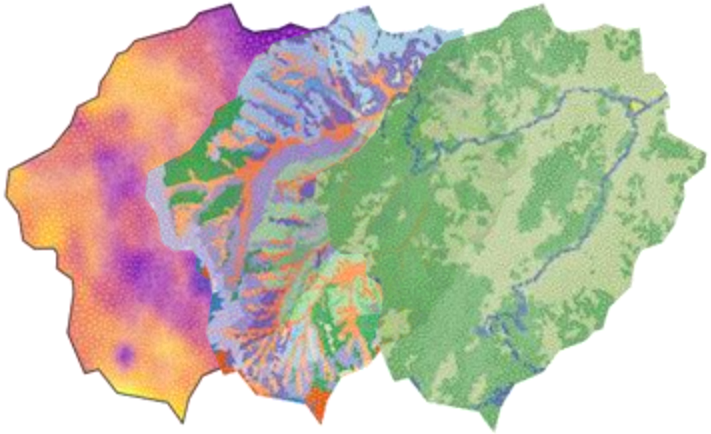
Proposed method: Generating Reporting Format compliant ATS outputs



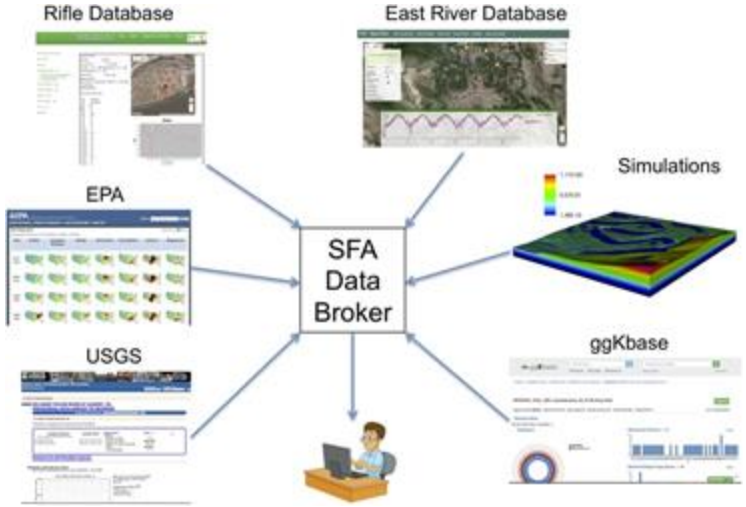


Watershed Workflow + BASIN–3D

A python-based workflow tool for downloading and adapting diverse data into a single 3D integrated hydrologic model.



A python-based brokering tool for querying and accessing data across a variety of sources, including ESS-DIVE.



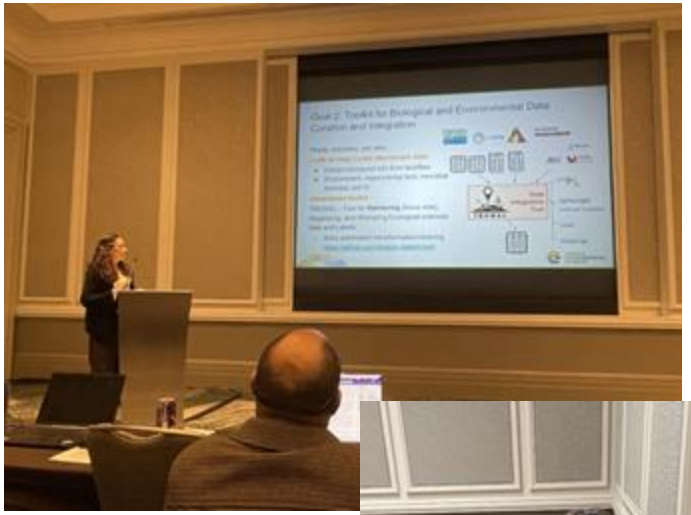
Hypothetical Use Cases:

- Automate the discovery of ESS data within ESS-modeled watersheds
- Generate ATS simulated observations to compare to ESS-DIVE archived field observations
- Accelerate ModEx studies by connecting models and data!

Flash Talks

Highlights from the community on progress in key areas:

- Models and Model Coupling
- Generative AI
- Physics-informed ML
- Tools



ELM

PFLOTRAN

Alquimia

ModEx

jupyter

Making possible what we couldn't do alone.

Bio-Eco Data Integration: 5 BER Resources Collaborating to Prototype a Unified Data Access Layer




Unifying Access Layer Components

BERtron - Global Search supported by common APIs - find and reuse data

Data Transfer Service - maintain provenance, propagate credit

Sample metadata, standardized data products

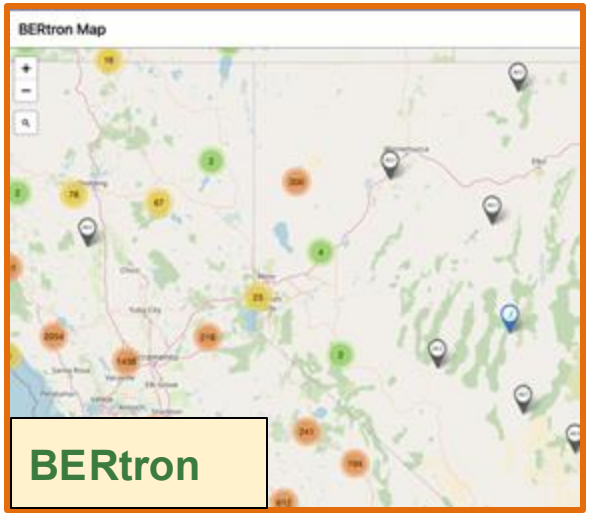
Biogeochemical measurements, sensor data

Proteomics, metabolomics, imaging

(meta)genomics, -transcriptomics, metadata

Bio-Eco Data Integration: Progress Highlight

BERtron map to find ESS-DIVE data co-located with biological data.



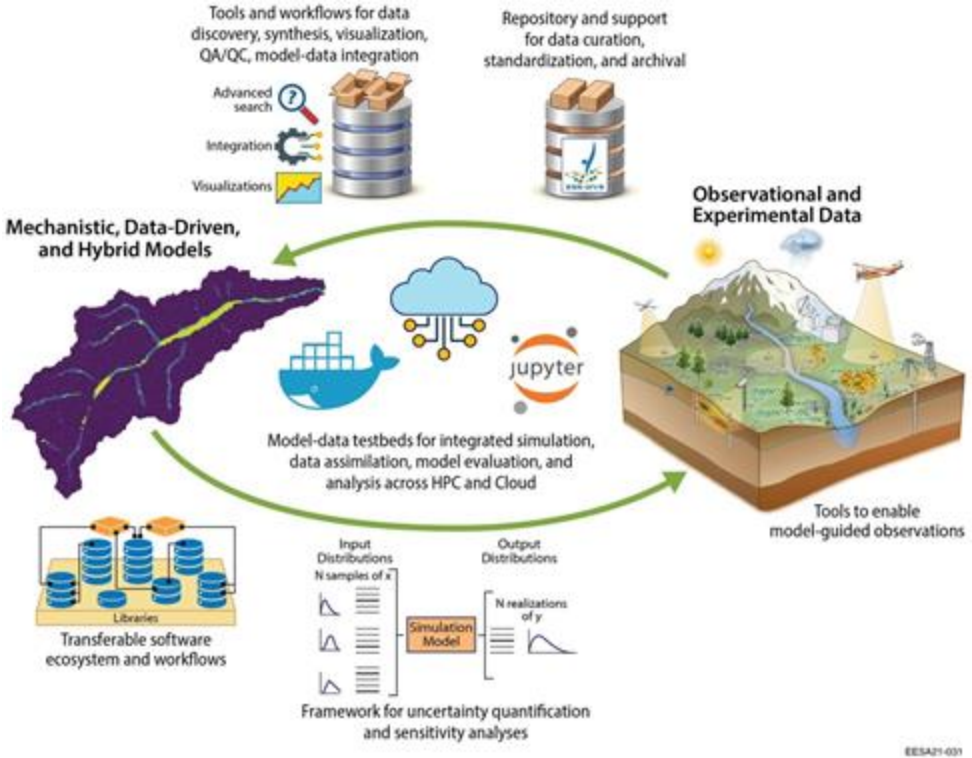
https://ber-data.github.io/bertron/map/index.html

ModEx cycle frames challenges and opportunities

Identify how your work fits into the ModEx cycle:

- What do you need or use from the ModEx cycle.
- Where does your science contribute to ModEx for the community.

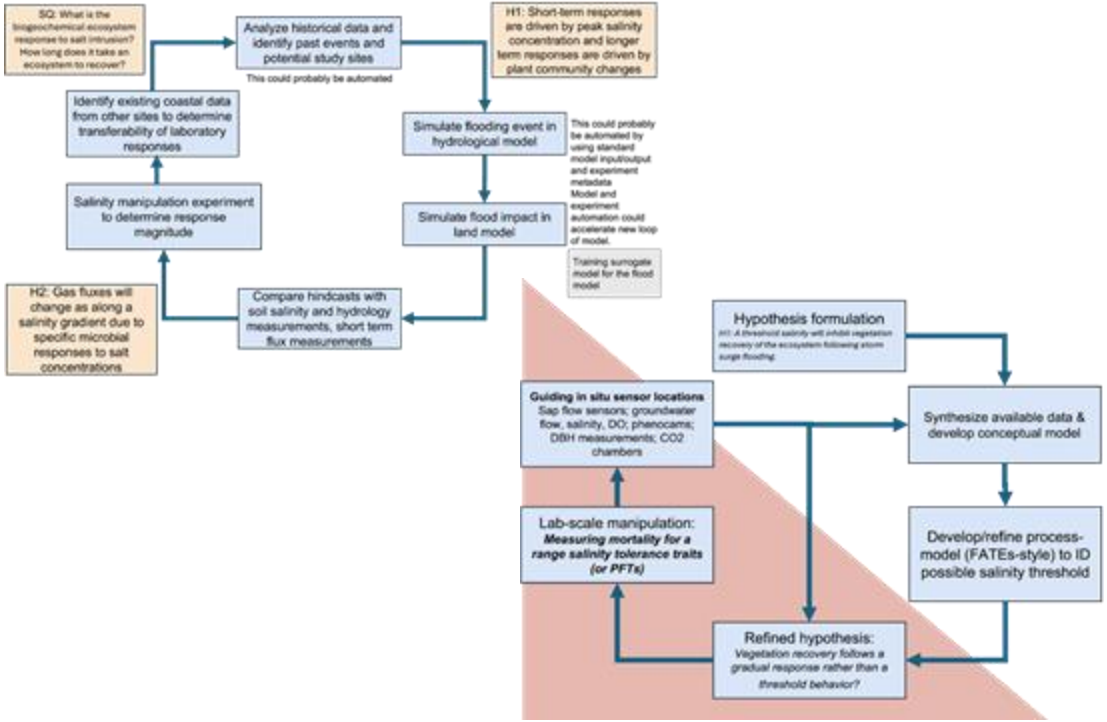
Imagine what standards, tools, interfaces, resources, etc. would help the community work more efficiently.

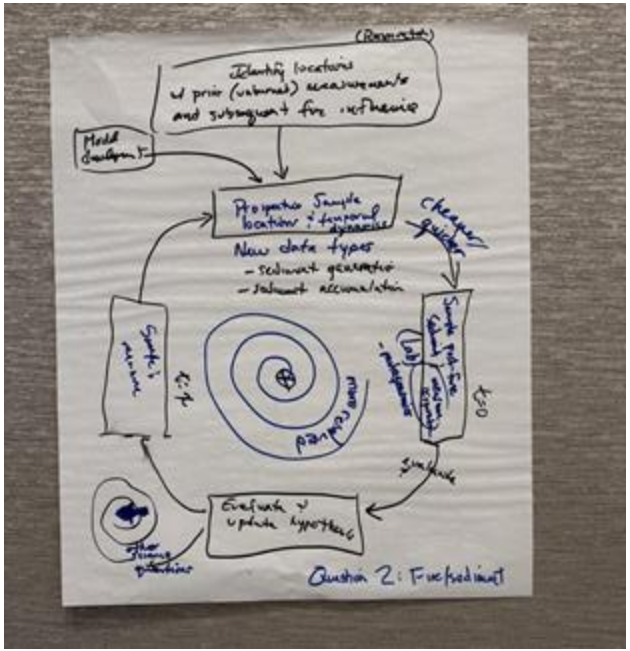


Model/Data Acclerating ModEx: Breakout Sessions

Model/Data Accelerating ModEx: Report Outs

We had a mix of PowerPoint and Hand Crafted ModEx Loops





Accelerating ModEx: Take Aways

- **Co-design** concept is very helpful, having modelers and empiricists working together helps *identify conceptual issues, understand each other needs, identify opportunities*, etc.
- **Iterating/cycling** moves team from scientific question to hypothesis, and then to refine or add hypotheses.
- Starting with imperfect and incomplete data, and simple models, and iterate together to improve data, and add complexity to the models.
- A single ModEx loop seems constraining, some groups had additional side loops, decision points/forks, etc.
- There are **many opportunities for AI/ML to accelerate** the iteration.

Conclusions

Annual Meeting is always collaborative, productive and energizing ...

Find new ways to keep our momentum throughout the year:

- Jump start more regular meetings through the Data Management Working Groups Quarterly meeting with an AI Topic for all.
- Join the “Accelerating ESS ModEx Cycles via AI/ML Integration” (Wednesday Breakout Session III) and team up for a short white paper
- Explore options for an AI/ML hackathon (e.g., Agentic AI workflows).

Consider joining the working groups, reach out to any of the leads or send me email (moulton@lanl.gov).

Thank You!
Any Questions?