EESSD - ESS Cyberinfrastructure Working Groups

Annual Meeting Highlights

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)

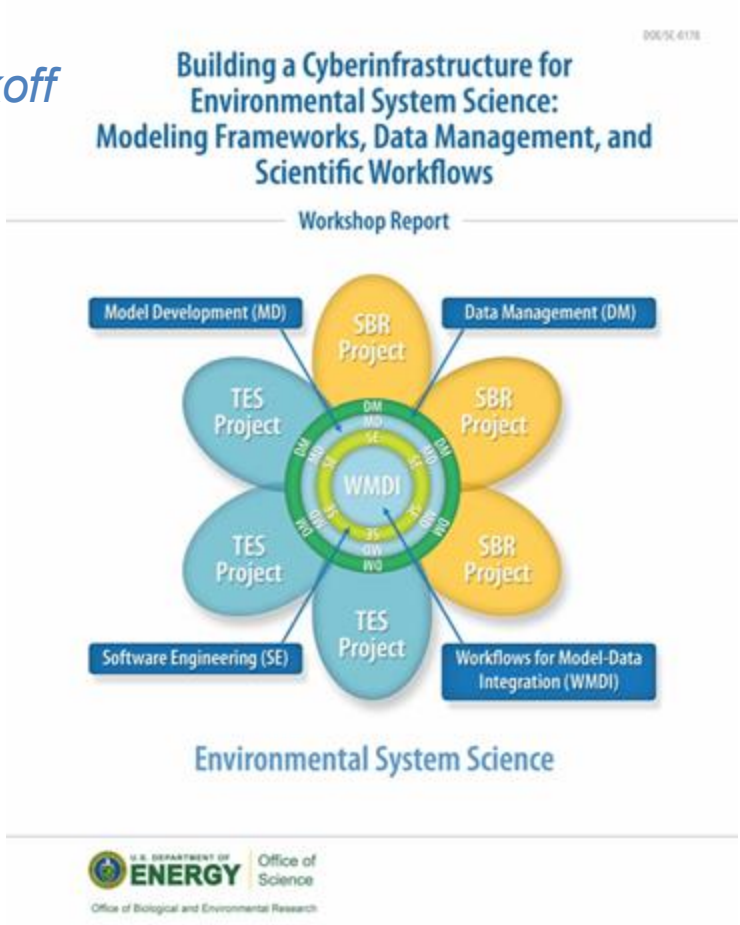

Xingyuan Chen (PNNL)
Forrest Hoffman (ORNL)

Software Engineering and Interoperability (SEI)

David Moulton (LANL)

Computing Infrastructure

Lee Ann McCue (PNNL)



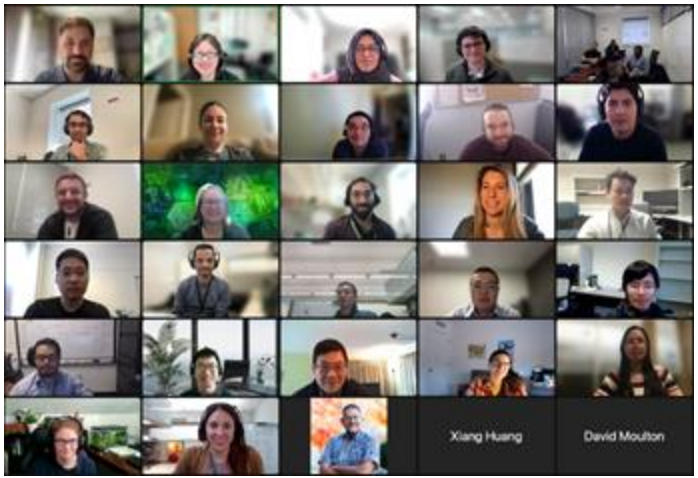
https://doesbr.org/ESS-WorkingGroups/ESSWG_WorkshopReport-final.pdf

Working Groups Annual Meeting

- Participation:
  - 30 in-person, 20 virtual on Zoom
  - 4 BER representatives
- Hybrid Logistics:
  - Google Docs and Jam Boards
- Agenda Highlights:
  - Quick recap of recent activities
  - ESS-DIVE Updates/Planning
  - Community Flash talks
  - Generative AI: Panel and Group Discussion
  - Hands-on and Breakout Sessions

Working Group Activity Highlights

- Data Management has quarterly meetings, topics include
  - Data integration brainstorming for BERAC RFI
  - NMDC submission portal & EMSL metadata template
  - BERAC Unified Data Report & ESS-DIVE proposal input
- Model Data Integration coordinated two hack-a-thons,
  - ELM (April 13, 2023) led by Dan Ricciuto and Faria Zahura
  - ILAMB (April 27, 2023) led by Jitu Kumar and Nathan Collier



Working Group Activity Highlights (continued)

Model Data Integration also partnered with CUAHSI on a Reactive Transport Short Course

CUAHSI ABOUT STUDENTS FACULTY COMMUNITY DATA SERVICES & SOFTWARE DONATE

REACTIVE TRANSPORT WITH SOPHISTICATED REACTION NETWORKS: HANDS-ON EXPERIENCE WITH THE PFLOTRAN REACTION SANDBOX

NOVEMBER 6 - 9, 2023

REGISTRATION DEADLINE 9.14.2023

Location: Pacific Northwest National Laboratory, Richland, WA

REGISTRATION WEBSITE

Instructors

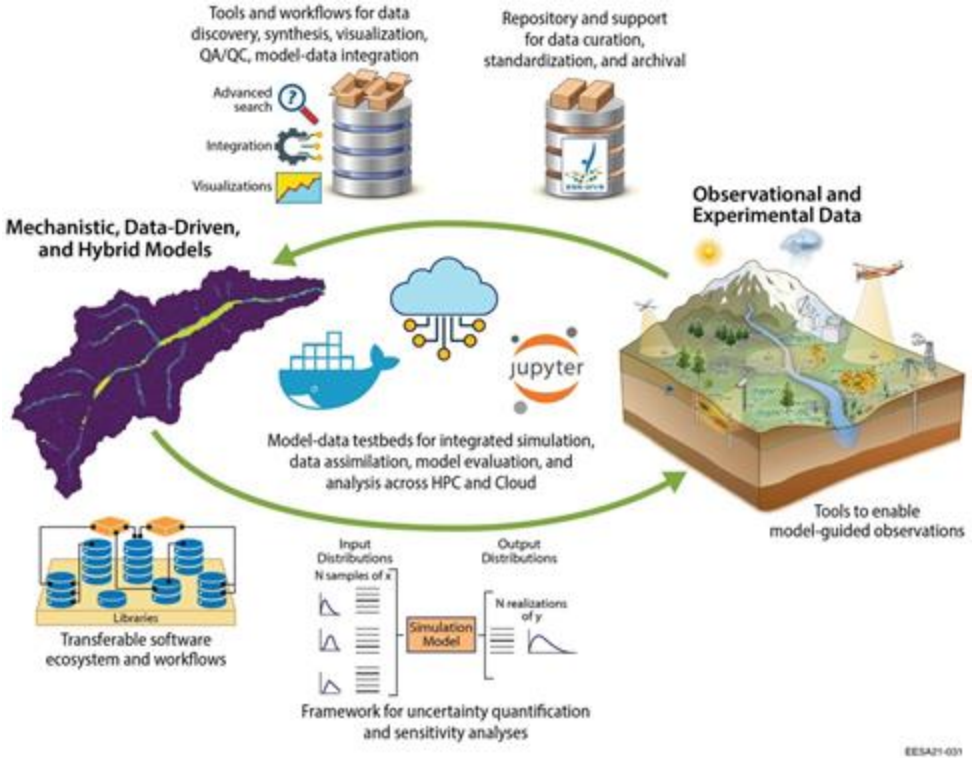
Glenn Hammond	Pacific Northwest National Laboratories
Xingyuan Chen	Pacific Northwest National Laboratories

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.



Flash Talks

Highlights from the community on progress in key areas:

- Community and Facilities
- Data Management
- ModEx
- AI/ML
- Toolsq

BERAC

nmdc National Microbiome Data Collaborative

ModEx

AI4CH$_4$

jupyter

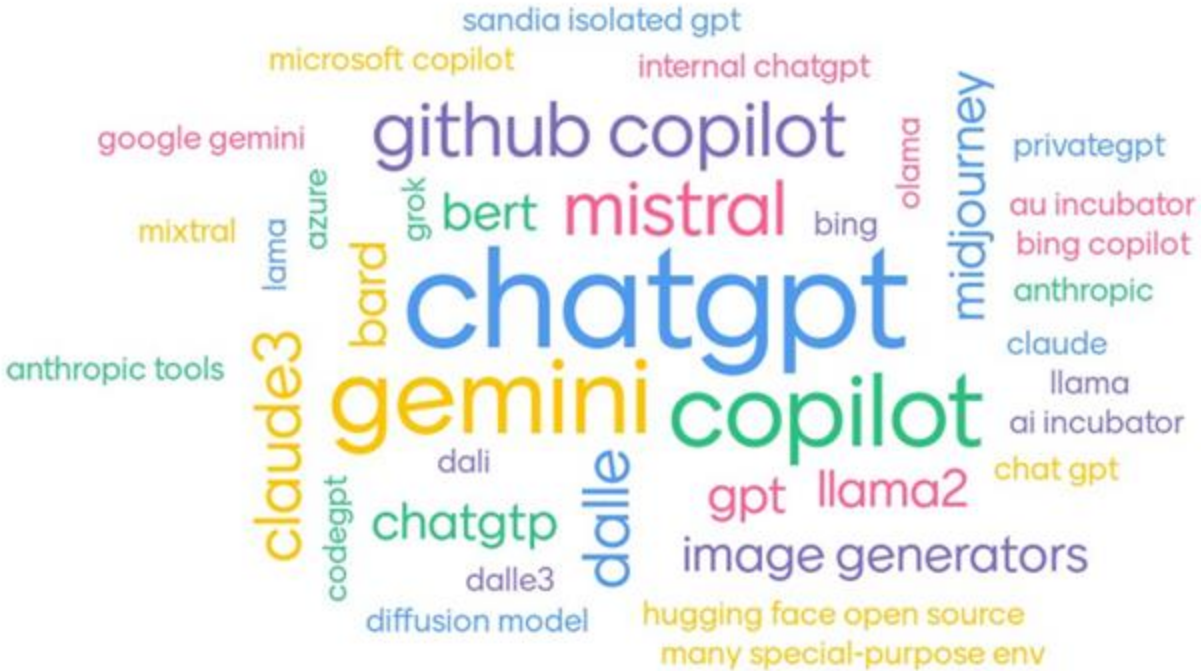
COMPASS

ODMX

Making possible what we couldn't do alone.

Generative AI: Panel and Group Discussion

What generative AI tools are you aware of?



Generative AI: Panel and Group Discussion

Tim Scheibe
(PNNL)

Mike Heroux
(SNL)

Tasneem Ahmadullah
(PNNL)

Tyson Swetnam
(U of Arizona)

Edwin Skidmore
(U of Arizona)

Panelists described their experience(s) with generative AI:

- Generating initial draft of a science highlight
- Creating a custom ChatGPT instance to help students accelerate selection of a research project.
- Accelerating learning a new language, python
- Creating a completely open-source (and free) generative AI system (Chatur), as part of NSF Cyberinfrastructure Project (cyverse) at University of Arizona

Generative AI: Panel and Group Discussion

Tim Scheibe
(PNNL)

Mike Heroux
(SNL)

Tasneem Ahmadullah
(PNNL)

Tyson Swetnam
(U of Arizona)

Edwin Skidmore
(U of Arizona)

Take away messages:

- It's can accelerate some facets of work, but its not a final solution
- It's iterative, and it takes time and effort to use responsibly
  - Ensure students are truly learning
  - Ensure text is consistent and accurate
- If equity is a priority, open-source (and free) tools are critical
  - Question of hardware resources remains
- Tap into your imagination

Breakout Sessions: Take away messages

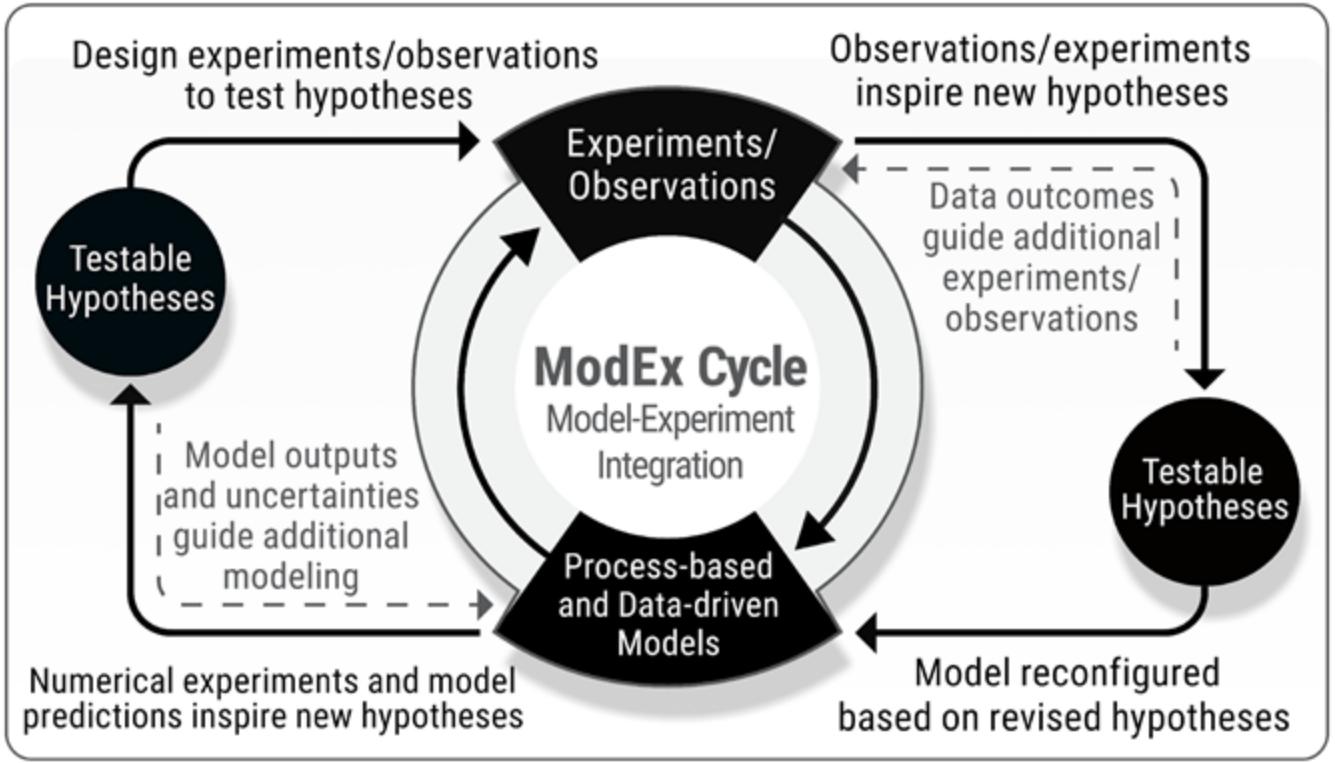
Model Intercomparisons – Beyond Benchmarking

- There is high interest in using model intercomparisons as a diagnostic tool to understand the importance of process components and for scenario analyses.
- Integration of metrics and comparisons through, e.g., ILAMB with other workflow flow tools such as Basin3D and Watershed Workflow tools would help the users.

Modeling code management and packaging of model data

- Important to archive model data outputs and we need to have a way to standardize how that is done across the spectrum of models used in ESS science
- Archive reusable model code or workflows as software with DOE code, and data specific processing scripts as part of the dataset.

Reframing MoDEx: Hypothesis-Driven



Inner Loop: Using models to design experiments or observations, using data to revise models.

Outer Loop: Explicit generation and testing of hypotheses.

Thank You!
Any Questions?