IHESD

INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

The GCAM Ecosystem – Status and Future Directions

November 6, 2019

Chris R. Vernon

On behalf of the GCAM team

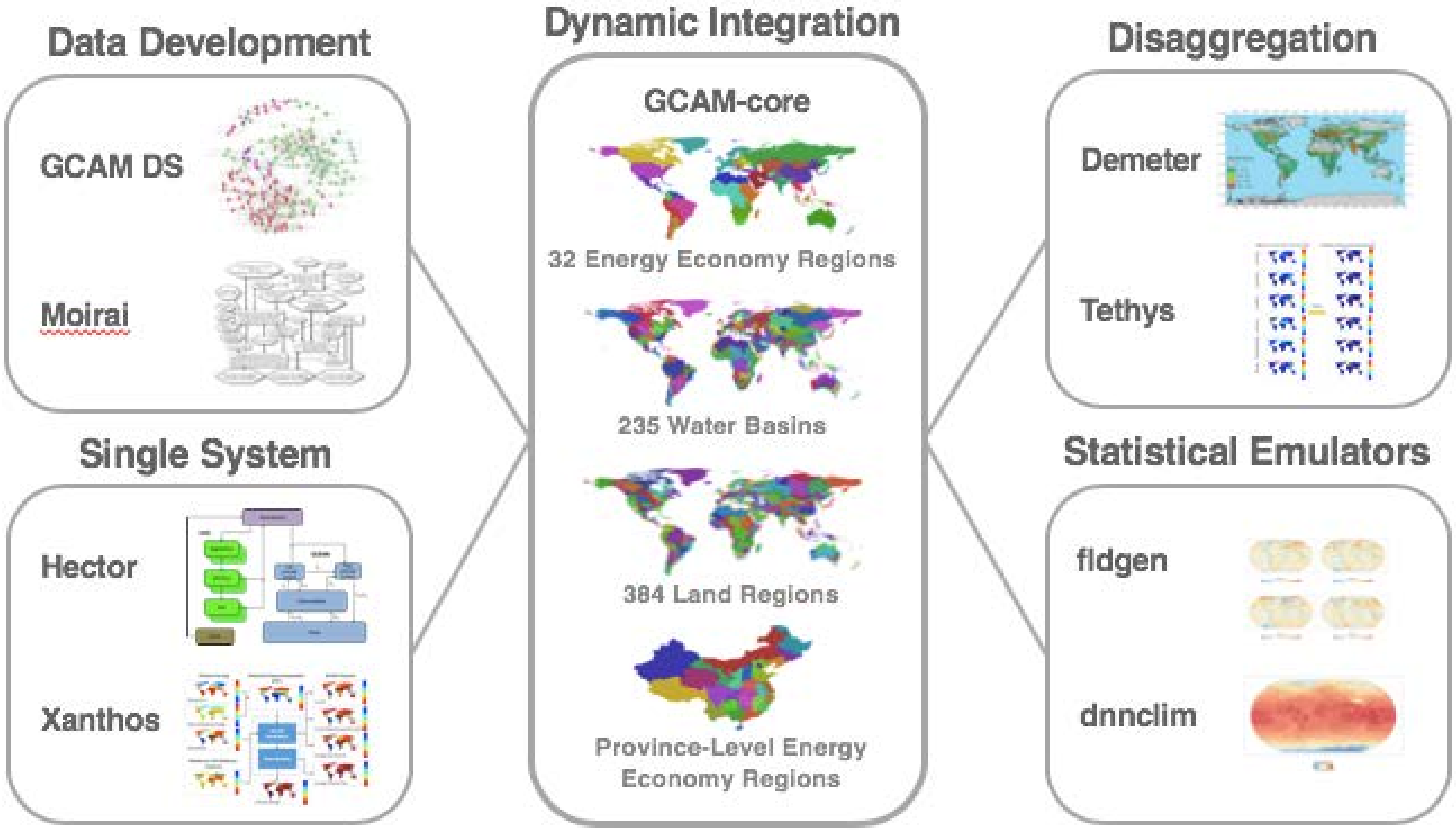
Pacific Northwest NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy

The GCAM Ecosystem



The GCAM Ecosystem

- **Exposure** – strive to maximize use of resources and capabilities for all collaborators
- **Usability** – generate functional, well documented models and tools
- **Efficiency** – identify bottlenecks and improvement areas for experimental modeling
- **Robustness** – support multiple scales, sectors, and platforms
- **Reproducibility** – utilize versioned models, tools, data, and workflows in modeling efforts

The GCAM Ecosystem

- **Exposure** – strive to maximize use of resources and capabilities for all collaborators
- **Usability** – generate functional, well documented models and tools
- **Efficiency** – identify bottlenecks and improvement areas for experimental modeling
- **Robustness** – support multiple scales, sectors, and platforms
- **Reproducibility** – utilize versioned models, tools, data, and workflows in modeling efforts
- **Interoperability** – set standards and support mechanisms for integration

IHESD
INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

The GCAM Ecosystem: Metrics

- 25+ models and tools in active development
- ALL open-source, publicly available
- 6 software publications, 1 in-press, 1 in-review since last meeting
- 23 citations (software publications)
- 2,453 unique reads (only JORS)
- 629 downloads (only JORS)

IHESD
INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

The GCAM Ecosystem: Status

- We are capitalizing on previous investments
  - Packages are being used to expedite development (e.g., rgcam, pygis)
  - Ensure consistency
  - Make updates single source

The GCAM Ecosystem: Status

- We are capitalizing on previous investments
  - Packages are being used to expedite development (e.g., rgcam, pygis)
  - Ensure consistency
  - Make updates single source

- More of our models are adopting a test-driven development approach and implementing CI
  - When we make changes to our code, we want to make sure those changes are not disruptive
  - Stability

IHESD
INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

The GCAM Ecosystem: Status

- We are capitalizing on previous investments
  - Packages are being used to expedite development (e.g., rgcam, pygis)
  - Ensure consistency
  - Make updates single source

- More of our models are adopting a test-driven development approach and implementing CI
  - When we make changes to our code, we want to make sure those changes are not disruptive
  - Stability

- Building in-memory transfer mechanisms
  - Traditionally models were heavily dependent upon file-based transfer with manual steps in-between to reformat data
  - Now in-memory transfer of data which cuts down room for error and increases performance

The GCAM Ecosystem: Status

“Be stubborn on vision and flexible on journey”
– Noramay Cadena

The GCAM Ecosystem: Status

“Be stubborn on vision and flexible on journey” – Noramay Cadena

- Replacement of GitLFS (large file storage) with remote retrieval and install functionality
- Sets us up to be an early adopter of MSD-LIVE!

IHESD
INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

Future Directions

- GitHub repositories for publication supplemental code used to make figures and conduct analysis. The README provides a road map to reproducibility:
  - Code reference for the version used
  - Journal reference for the publication
  - Data reference for the contributing or resulting data
  - Contributing model references

DOI 10.5281/zenodo.2035720

vernon_etal_jors_2019

A Global Hydrologic Framework to Accelerate Scientific Discovery

Abstract

With the ability to simulate historical and future global water availability on a monthly time step at a spatial resolution of 0.5 geographic degree, the Python package Xanthos version 1 provided a solid foundation for continuing advancements in global water dynamics science. The goal of Xanthos version 2 was to build upon previous investments by creating a Python framework where core components of the model (potential evapotranspiration (PET), runoff generation, and river routing) could be interchanged or extended without having to start from scratch. Xanthos 2 utilizes a component-style architecture which enables researchers to quickly incorporate and test cutting-edge research in a stable modeling environment prebuilt with diagnostics. Major advancements for Xanthos 2 were also achieved by the creation of a robust default configuration with a calibration module, hydropower modules, and new PET modules, which are now available to the scientific community.

Code reference

Vernon, C.R., Hejazi, M.I., Turner, S.W.D., Liu, Y., Braun, C.J., Li, X. and Link, R.P., 2019. A Global Hydrologic Framework to Accelerate Scientific Discovery. Zenodo. https://doi.org/10.5281/zenodo.2035720

Journal reference

Vernon, C.R., Hejazi, M.I., Turner, S.W.D., Liu, Y., Braun, C.J., Li, X. and Link, R.P., 2019. A Global Hydrologic Framework to Accelerate Scientific Discovery. Journal of Open Research Software, 7(1), p.1. DOI: http://doi.org/10.5334/jors.245

Data reference

Vernon, C.R., Hejazi, M.I., Turner, S.W.D., Liu, Y., Braun, C.J., Li, X. and Link, R.P., 2019. Supporting Data: A Global Hydrologic Framework to Accelerate Scientific Discovery. Zenodo. https://zenodo.org/record/2578287/files/example.zip?download=1

Contributing models

Model	Version	Repository Link	DOI
Xanthos	v2.0.0	https://github.com/JGCRI/xanthos/tree/v2.0.0	https://doi.org/10.5281/z

Future Directions

- Focus on high cohesion in software development
  - Easier to profile and identify performance bottlenecks
  - Increase reusability

Future Directions

- Focus on high cohesion in software development
  - Easier to profile and identify performance bottlenecks
  - Increase reusability
- Autogenerate scale conversion and boundary products
  - Spatial products (e.g., boundaries) from Moirai

Future Directions

- Focus on high cohesion in software development
  - Easier to profile and identify performance bottlenecks
  - Increase reusability
- Autogenerate scale conversion and boundary products
  - Spatial products (e.g., boundaries) from Moirai
- Incorporate FAIR data practices into our development workflow (e.g., release our calibration and validation datasets so the process can be reproduced)

Future Directions

- Focus on high cohesion in software development
  - Easier to profile and identify performance bottlenecks
  - Increase reusability
- Autogenerate scale conversion and boundary products
  - Spatial products (e.g., boundaries) from Moirai
- Incorporate FAIR data practices into our development workflow (e.g., release our calibration and validation datasets so the process can be reproduced)
- Modules to read and process data from source
  - No more derivative products without provenance
  - Read in data from source datasets (ISIMIP, etc.) where preprocessing code is called within the model

Future Directions

- Visual analytics
  - Interactive and static discovery, comparative, relational, composition, distribution, and trend products
  - Utilize existing web services

Future Directions

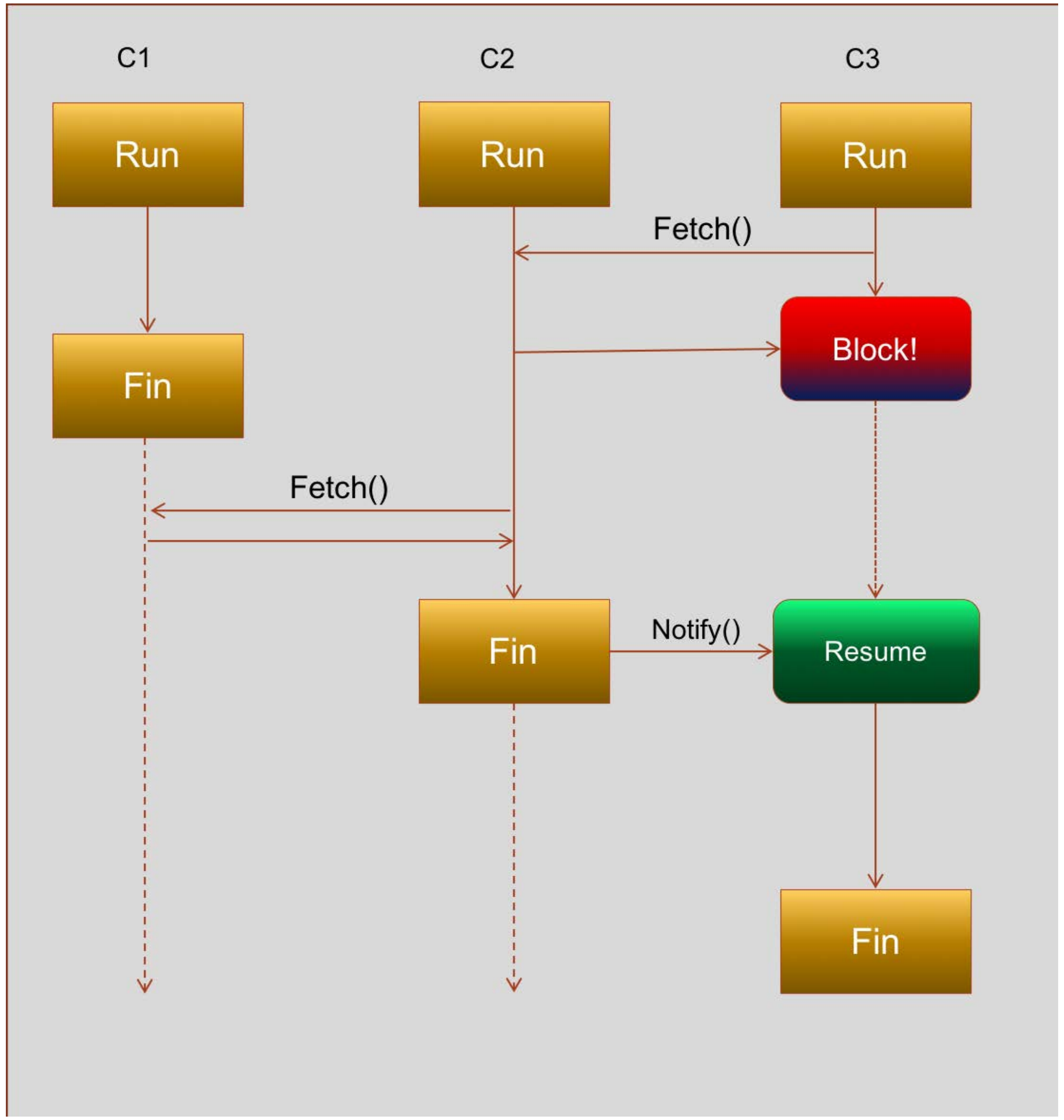
- Visual analytics
  - Interactive and static discovery, comparative, relational, composition, distribution, and trend products
  - Utilize existing web services
- Standardized model configurations setup to run what you need on a HPCC
  - Most people ask for very similar runs
  - Catalog of exiting runs with an associated DOI, metadata, and citable reference that may fit the bill

Future Directions

- Visual analytics
  - Interactive and static discovery, comparative, relational, composition, distribution, and trend products
  - Utilize existing web services
- Standardized model configurations setup to run what you need on a HPCC
  - Most people ask for very similar runs
  - Catalog of exiting runs with an associated DOI, metadata, and citable reference that may fit the bill
- From forcing to coupling with Cassandra
  - Will provide a mechanism to help us address increasingly complex science objectives
  - Promotes the rapid configuration of models to address specific questions; not an uber model where everything is on all of the time

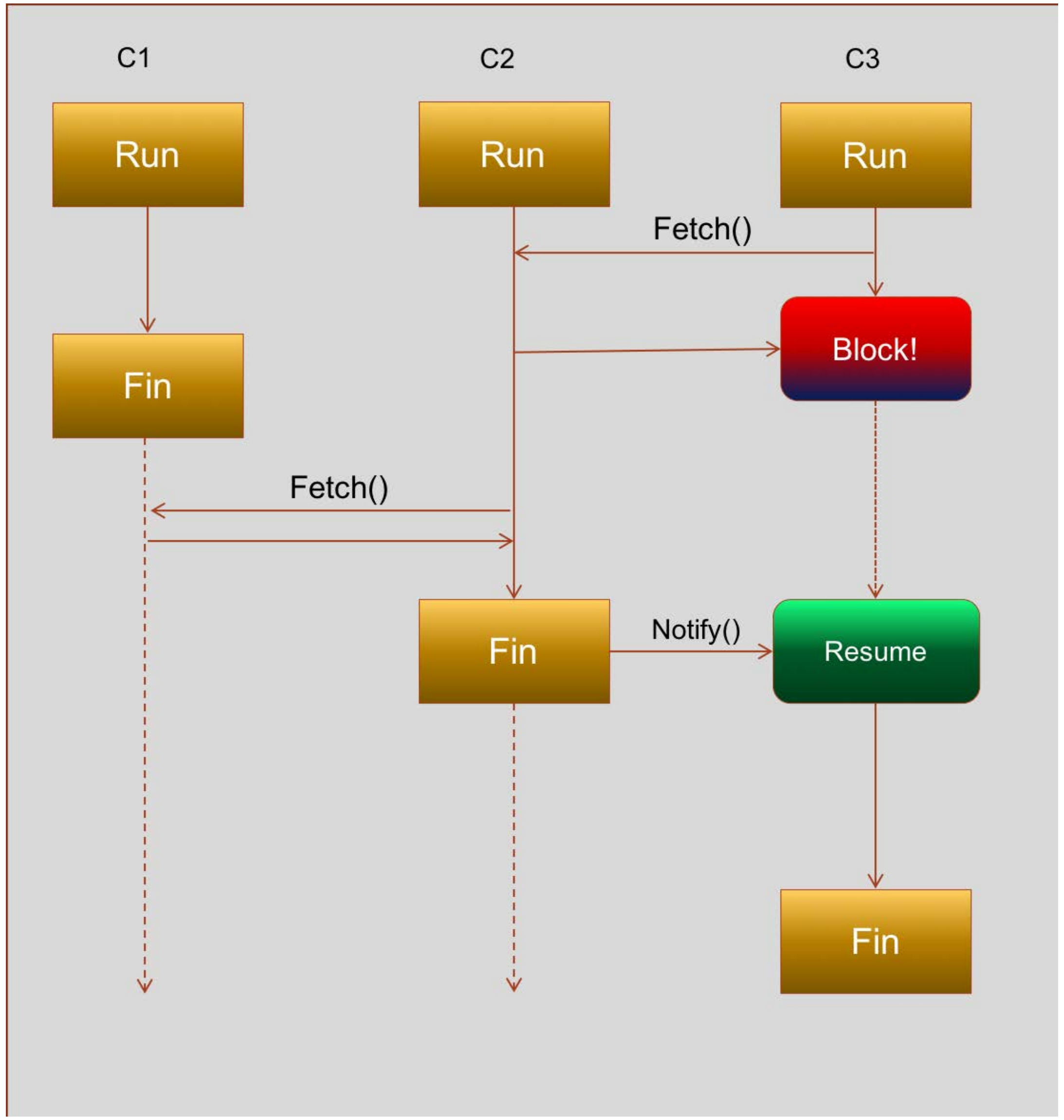
Cassandra

- Model coupling framework
- Component-based (customizable)
- Components provide capabilities to the overall framework for other components to use
- The configuration of models is setup by a config file
- Models can be ran on a single node, or distributed across nodes
- Currently supports one-way forcing/coupling



Cassandra: Development

- Setup models to participate (e.g., Python API for GCAM)
- Creation of components for existing models and tools
- In-memory data exchange per time-step (iterative feedback capability)
- Explore standard naming conventions



The GCAM Ecosystem

- Status and Future Directions
- Advancements: Fldgen, Robert Link
- Advancements: Demeter, Min Chen
- Advancements: Xanthos, Chris Vernon
- New Developments: Metis, Zarrar Khan
- Afternoon Overview: Steve Smith
- Breakout Session!

IHESD
INTEGRATED HUMAN-EARTH SYSTEM DYNAMICS

Thank you

Pacific Northwest
NATIONAL LABORATORY