The concept of NCAR's community software facility

Thomas Hauser, Glen Romine, Tricia O'Keefe

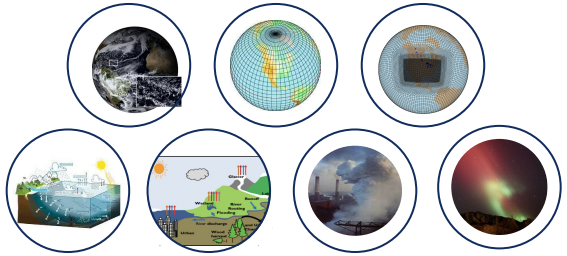
October, 10, 2023

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

NCAR's computational science strategy

- Enabling and supporting the community through a digital Earth System Science lab
- Democratizing Research Capabilities in support of Earth System Science

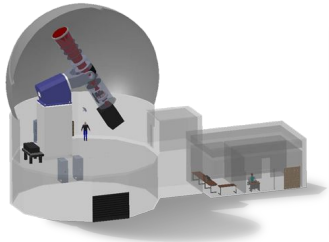
Community Software Facility

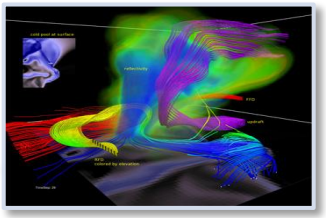


NCAR Compute Infrastructure

Translational Computational Science

NCAR Research Data Commons





Disruptive trends → Opportunity to rethink our approach

Accelerating Earth system science knowledge and discovery through computational science

Societal Need

- Understanding local, extreme hazards and events
- Enable place-based science across natural and human systems
- Create climate resilience and mitigation solutions

Increasing Model Complexity

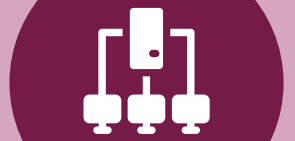
- Democratizing access to our models and model outputs
- Earth system coupled models across time and space scales
- Data assimilation across component models
- Finer spatial and temporal resolution simulations
- Large number of ensembles

Rapid Changes in Compute & Data Technology

- Energy efficiency
- Physical limits drive us towards heterogeneous hardware
- Barriers of entry to advanced computing systems
- AI/ML revolution
- Data silos & inaccessibility of individual PI's data sets

Integration and engagement across labs, community and stakeholders

NCAR's computational science strategy



NCAR Compute Infrastructure

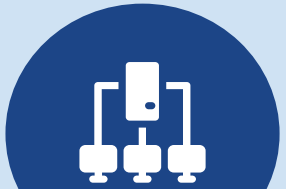
- Data intensive architecture
- Integrate and federate with the NSF CISE funded computing ecosystem
- Expand our resources through cross-agency partnerships

Community Software Facility

- Opportunity to rethink how we develop and support our models
- Unified support across NCAR
- Partnerships to focus on what is most valuable to our community
- Democratize access to our models
- Co-Design with computational facility

Translational Computational Science

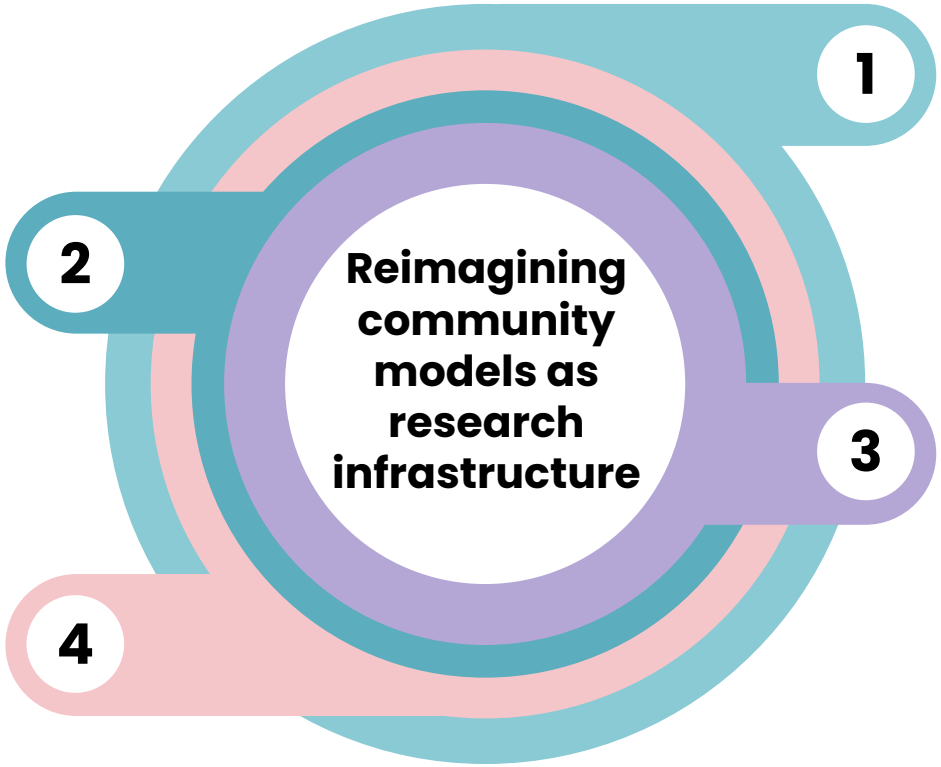
- Innovate across all areas of the computational science strategy
- Investment into applied computational science research
- Extending our capability through connections and partnerships



NCAR Research Data Commons

- Unified platform for NCAR and community data
- Sustainable business model
- Cloud platform services for data analysis, AI/ML, and visualization

CONCEPT: Community Software Facility



1 SCIENCE DRIVER

Earth System Science

2 COMMUNITY DRIVER

Democratization of Science

3 TECHNOLOGY DRIVERS

Compute and Data Technology

Software Development

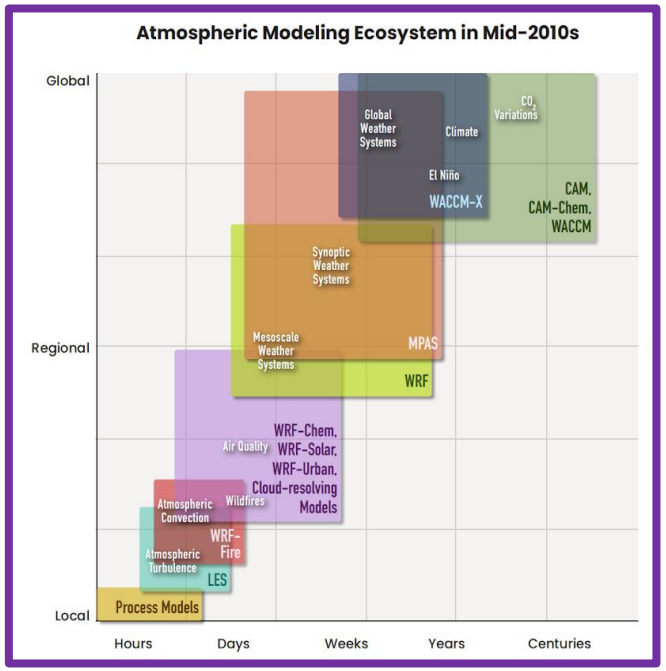
4 WORKFORCE DRIVER

Workforce of the Future

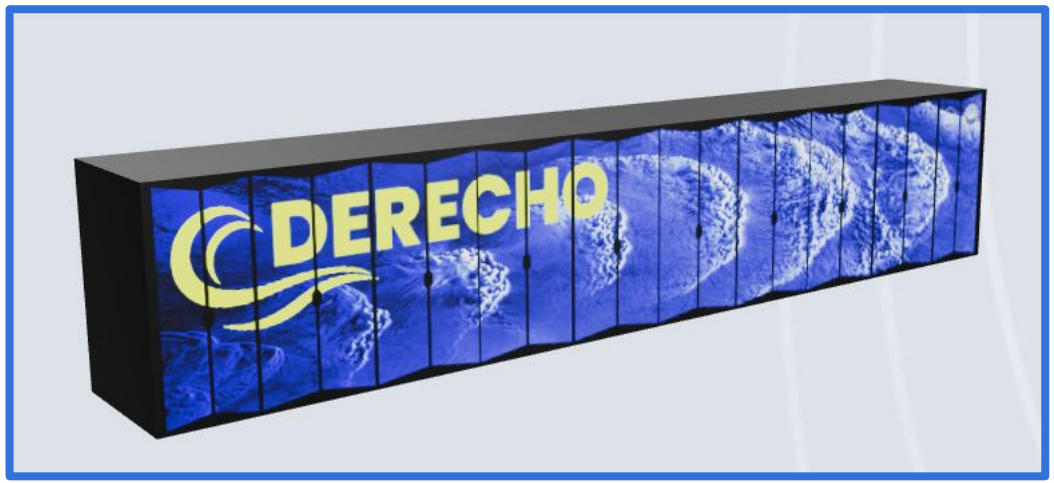
Disruptive Trends

- Societal need
- Increasing model complexity, resolution & ensembles
- Rapid changes in software & technology

Research infrastructure at NCAR



Models as scientific instruments



Cyberinfrastructure

We are at a breaking point → change is necessary to support the community

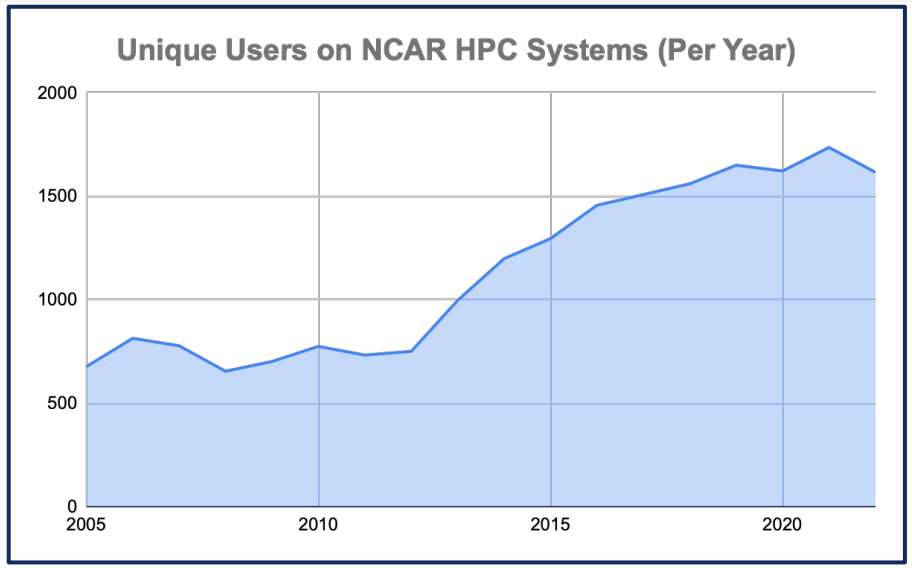
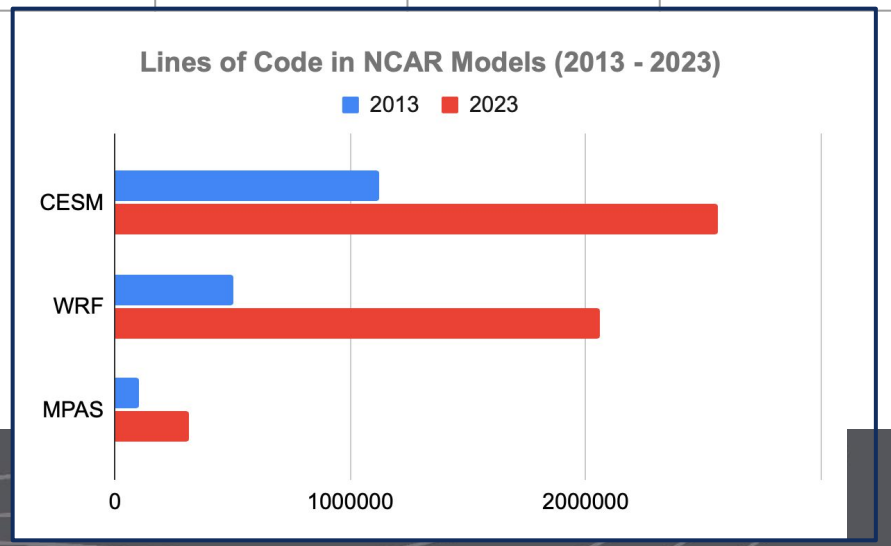


Image from: Nowogrodzki, Anna. 2019. "How to Support Open-Source Software and Stay Sane." *Nature* 571 (7763): 133–34. https://doi.org/10.1038/d41586-019-02046-0.

New Users, Threads and Messages on the CESM Forums (per year)

	2020-2021	2021-2022	2022-2023
Users	420	486	564
Threads	741	869	896
Messages	3609	4182	4630

Model	Open Issues	Closed Issues	Open PRs	Closed PRs
CESM	1190	4003	66	4961
WRF	121	177	10	1513
MPAS	59	127	57	787



NCAR | UCAR

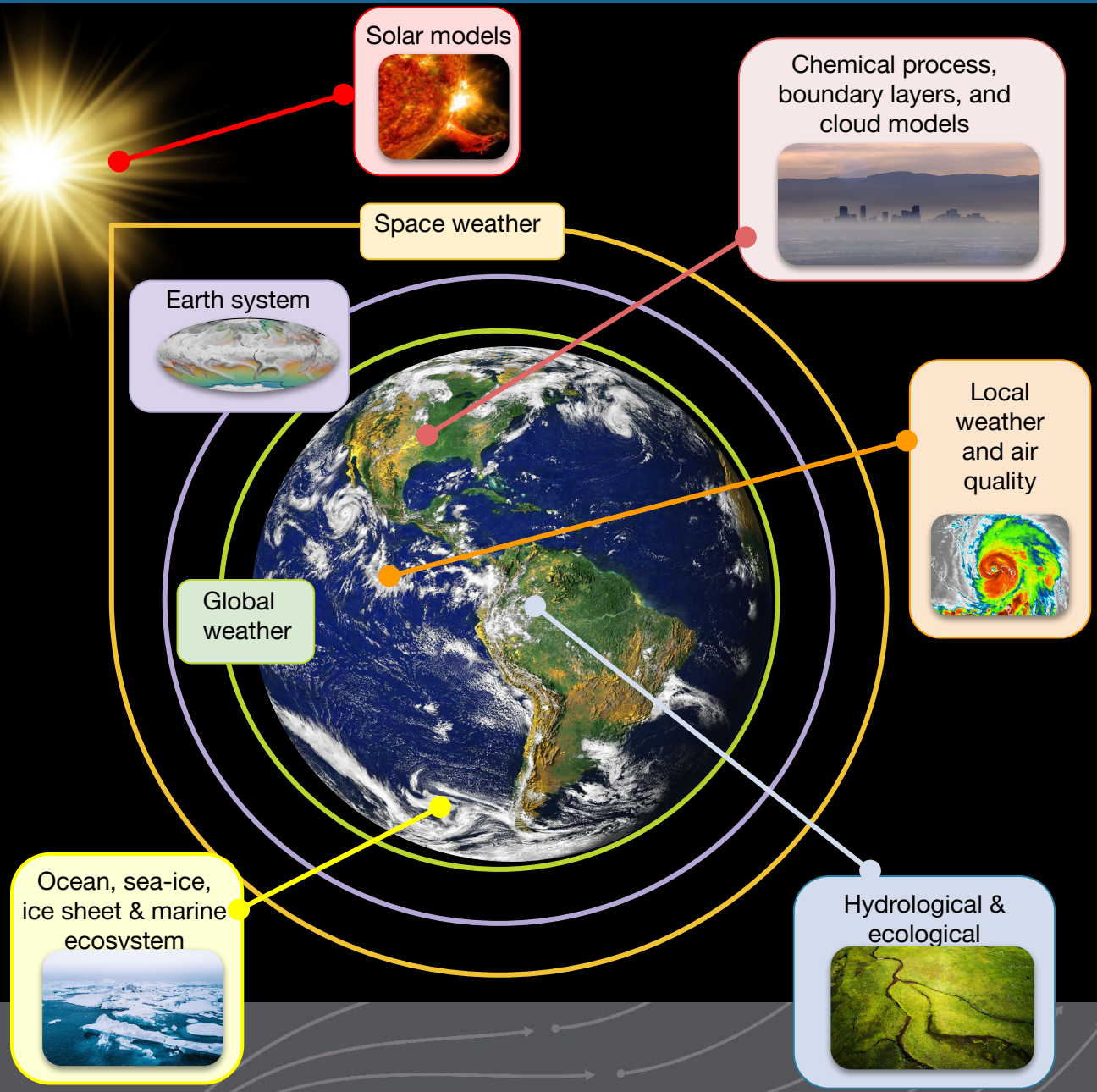
Graphs credit: Brian Dobbins, CGD

NCAR'S Modeling Ecosystem

The collection of models, informed by NCAR's deep scientific expertise, represents a diverse, flexible, and powerful *modeling ecosystem*

Enables interdisciplinary Earth system research across

- global to local scales
- array of complex Earth system interactions



- Risk of doing nothing
- **Culture change is difficult**
  - Team Science
  - Software engineering practices
  - Partnerships
  - Project management
  - Workforce composition
- How do we manage that change?

Community Software Facility Roadmap

Year 1
- Create unified support across NCAR/UCAR services (Tier 1-3)
- Deploy a unified testing framework

Year 3
- Organizational structure
- Technology selection
- Automation and AI/ML
- SE standard practices
- Common architecture

Year 5
- Code modernization
  - Ease of use
  - Modularization
- Reasonable portable performance
- NCAR's models can compete for Gordon Bell award in climate science

Assess capability
Workhorse CESM
EarthWorks
SIMA
CTSM

Assess SE practices

Assess Partnership Opportunities

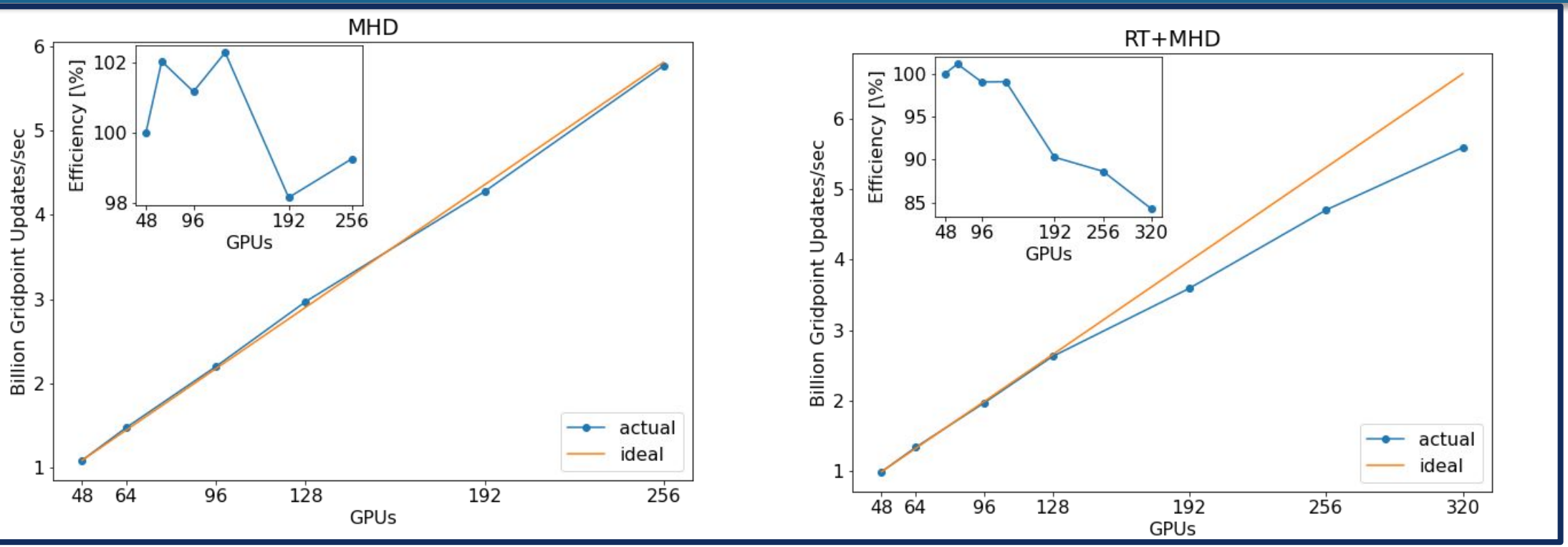
Community workshops and stakeholder engagements, then assess and revise roadmap to meet community needs; engage community in development

Formalize partnerships to exchange knowledge and adopt practices with partners, grow capacity

Develop sustainable model for adopting community code contributions

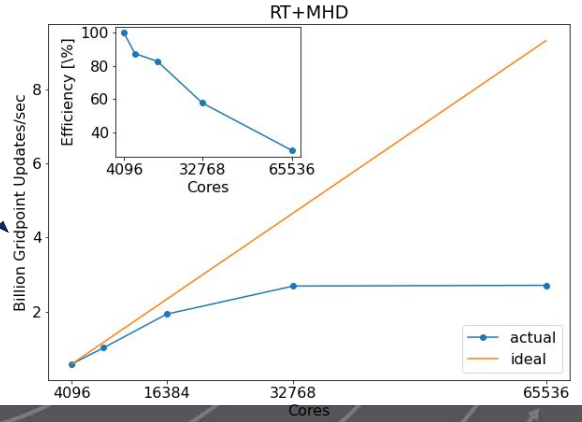
Dependent on securing additional resources!

Diverse Team Success Story - MURaM



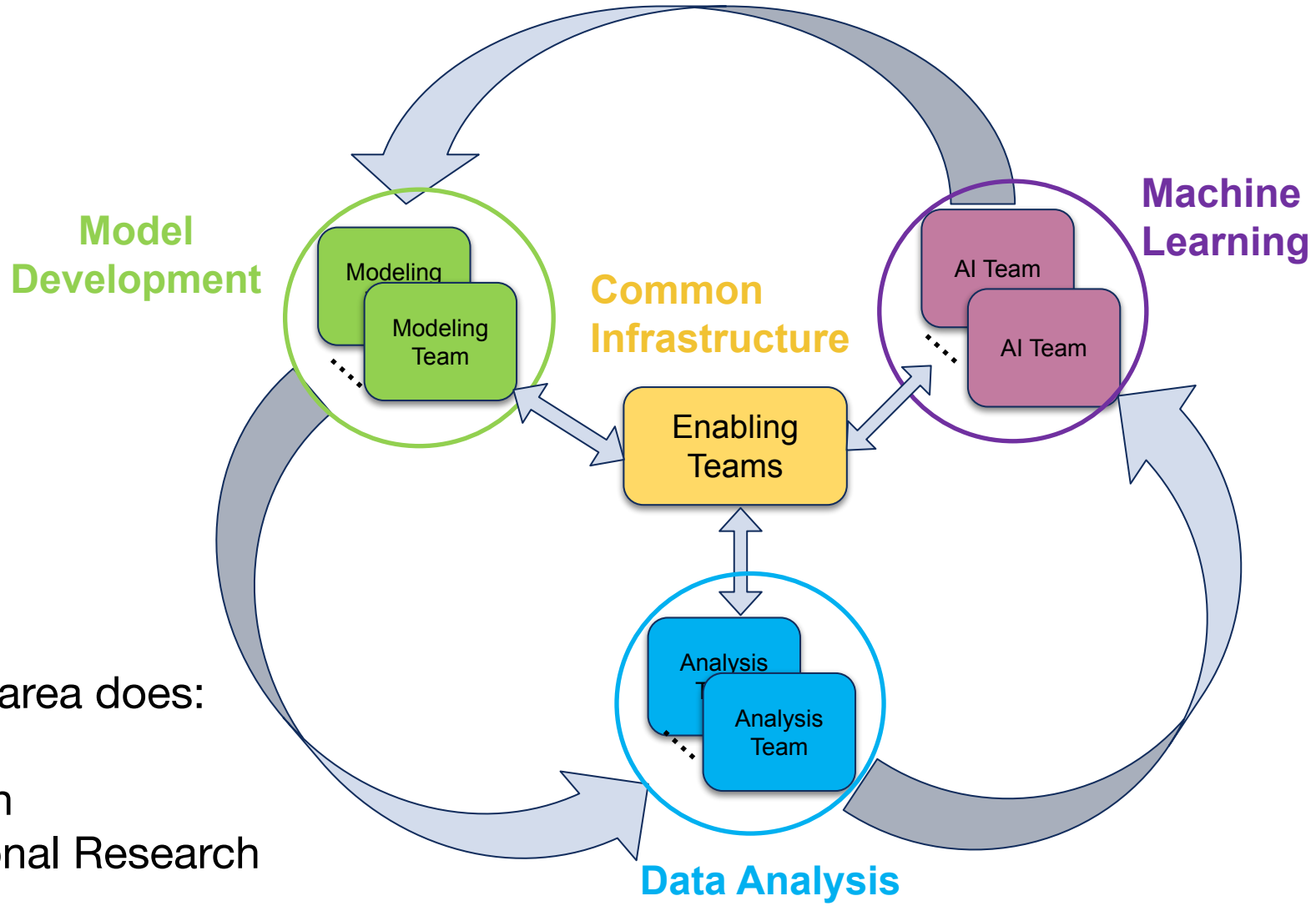
Strong scaling for MHD and MHD+RT physics within MURaM

CPU scaling for same case falls off above 16K cores, cannot catch up!!



***MURaM can run the same problem efficiently on 192+ of Derecho's GPUs than currently possible on any number of CPU cores**, courtesy Matthias Rempel et al.*

Software Facility Organizational Structure: A Teams of Teams Approach



- **Conway's law:** Any organization that designs a system will produce a design whose structure is a copy of the organization's communication structure. (and every team that worked before on that)

Each focus area does:

- Training
- Education
- Translational Research

[1]Raybourn E.M., et al. (2019). **Scaling Productivity and Innovation on the Path to Exascale with a "Team of Teams" Approach,** Springer, Cham.

Community software facility - immediate opportunities

LEAP

Learning the Earth with Artificial Intelligence and Physics (LEAP)

Columbia University in the City of New York

LEAP is an NSF STC focused on developing AI-based solutions to represent complex physical processes at scale within CESM.

NCAR does not have processes and funding to incorporate external innovations.

Approach:

- Develop and fund new approaches to incorporate and support community developments
- Increase NCAR's AI capacity

EarthWorks is an NSF supported project to extensively modify CESM to enable ultra-resolution simulations of the Earth system.

Demonstrates the capabilities of CESM infrastructure but lacks a sustainable pathway for developments.

Approach:
Enable software innovations through greater interoperability among components and community support services

Raijin is an NSF-Earthcube project to enhance analysis and visualization tools that are community-owned, sustainable, and scalable.

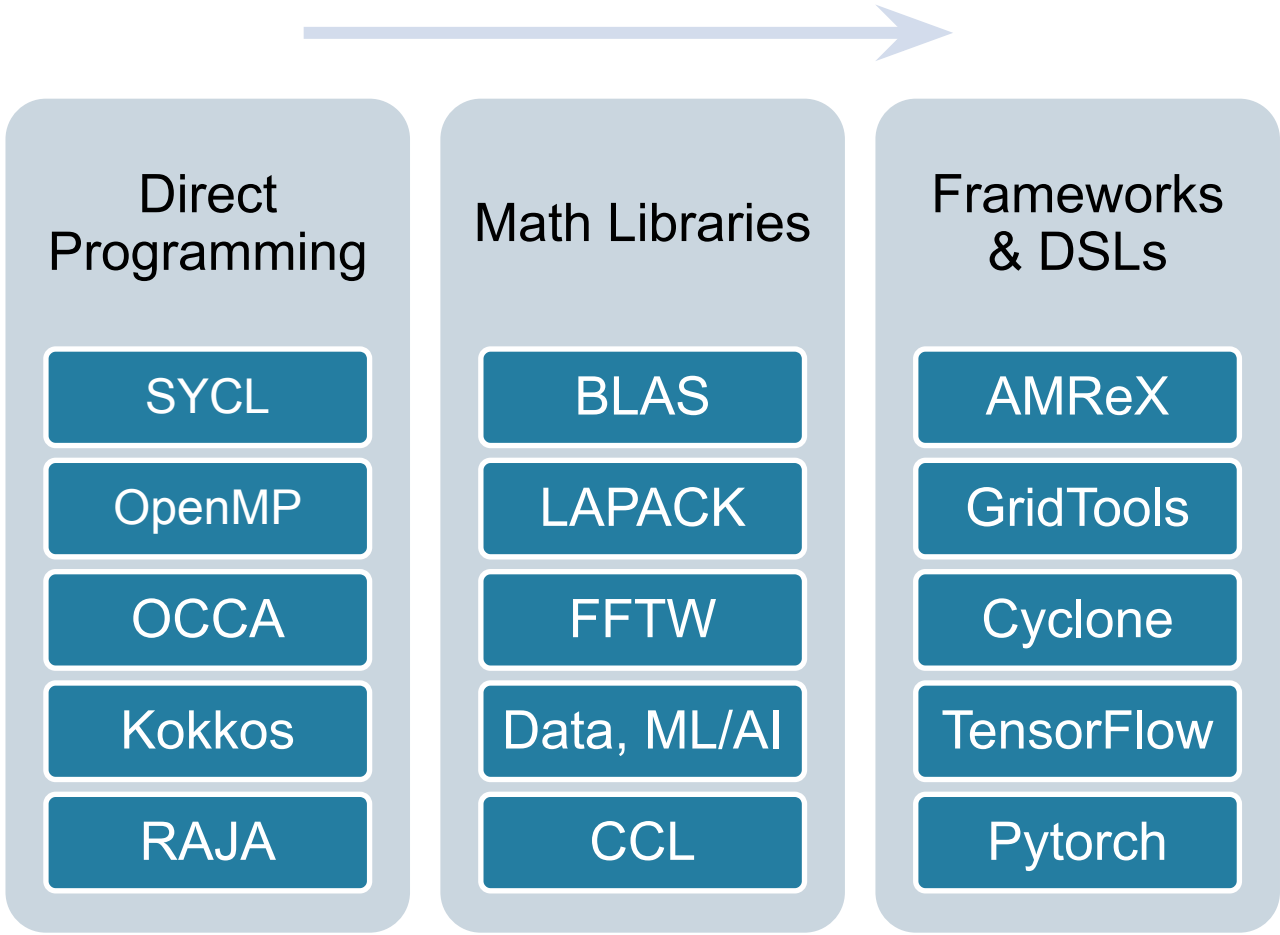
Exemplar of NCAR funded with the community to adopt and support project outcomes.

Approach:

- NCAR and universities co-develop community software
- Integration of community software in our ecosystem

Tension between Portability, Productivity and Performance

- Portable software can run on
  - Different types of hardware
  - Different vendors' hardware
  - What are the right abstractions?
- Goal is to minimize
  - Lines of source code needed to achieve portability
  - Effort to run existing code on new and future types of hardware
- Want turnkey performance
  - Otherwise with minimal (automatic) parameter tuning



Our Challenges are not not unique - Pursue Partnerships

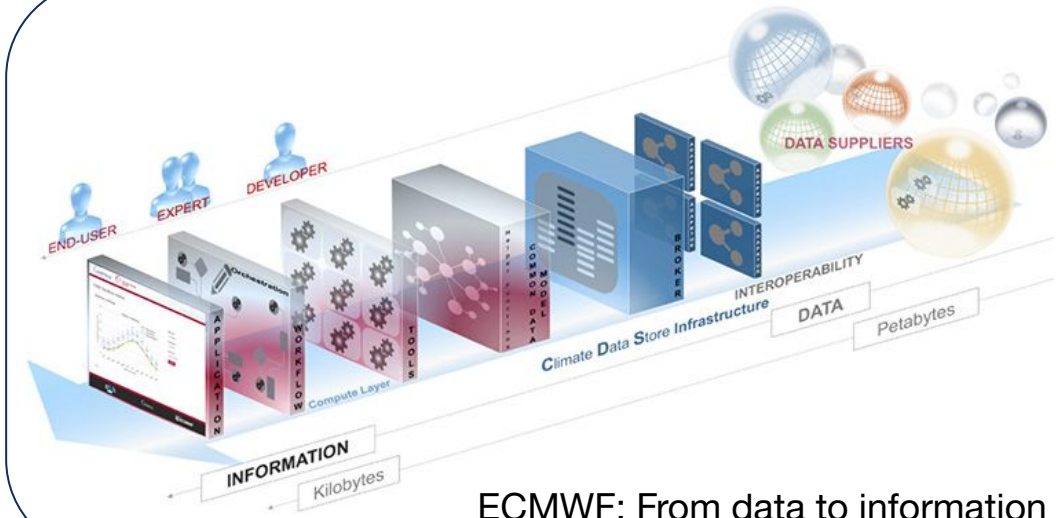
Modernization of modeling software

Co-design between scientists and computer scientists

Data Science, new workflows, AI/ML, data management

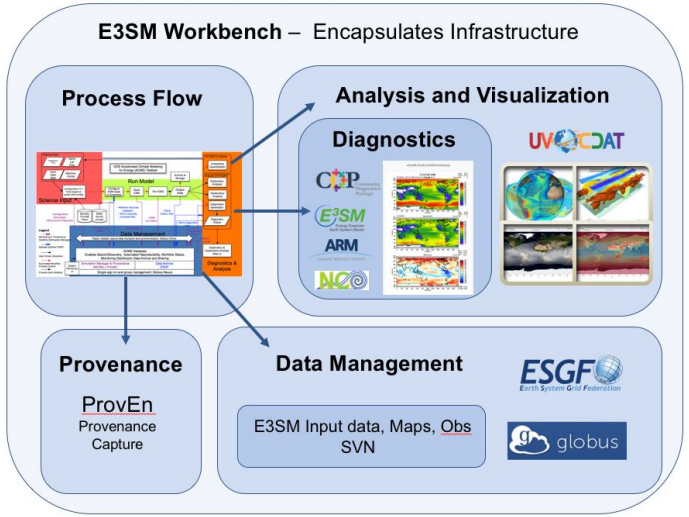
Investment in software engineering, scientific design

Adapted from UK Met Office & EPSRC: Harnessing Exascale Computing

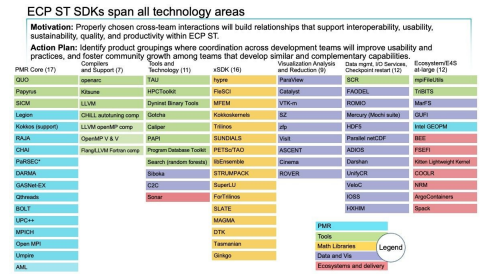


ECMWF: From data to information

DOE E3SM



DOE Exascale project



Many opportunities for cross-agency and international partnership on tools and methods

NCAR | UCAR

Is it time for a community of practice around code modernization?

- Interagency Council for Advancing Meteorological Services (ICAMS) High-Performance Computing Focus Team
  - Working on a report about exascale readiness of climate and weather models
  - Monthly presentations about different models, approaches

- Can we create a forum to share failures, success, and challenges on a regular basis?
  - Team Science
  - Software engineering practices
  - Partnerships
  - Project management
  - Workforce composition

Looking forward to continued discussions