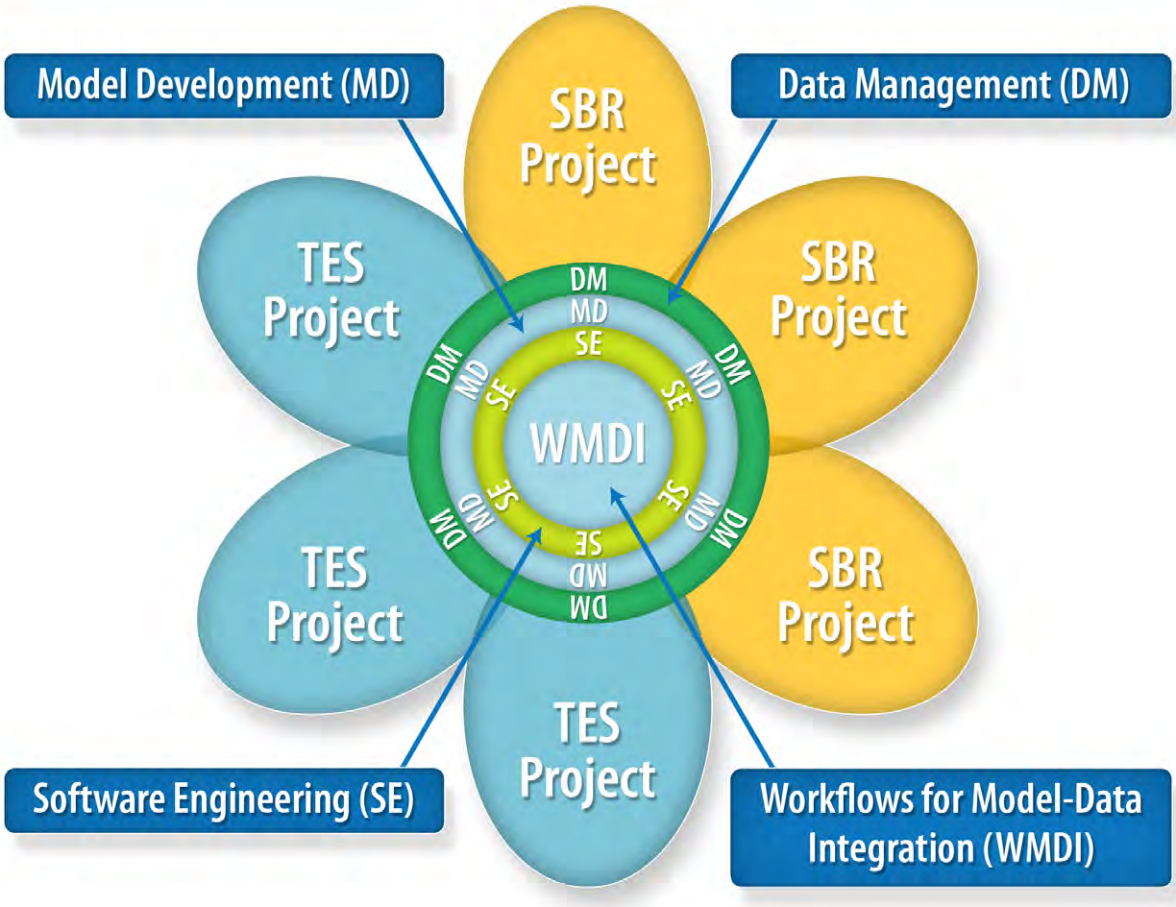
EXECUTIVE SUMMARY

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

EXECUTIVE SUMMARY

The Environmental System Science (ESS) activity within the Department of Energy's (DOE) Office of Biological and Environmental Research (BER) seeks to advance a robust, predictive understanding of terrestrial environments, extending from bedrock to the top of the vegetative canopy and from molecular to global scales, through an iterative cycle of model-driven experimentation and observation dubbed MODEX. Considerable progress has been made toward achieving this overarching goal, but widely recognized is the fragmentation across projects and disciplines of the relevant modeling and simulation capabilities, observational and experimental data, analysis algorithms, and workflow tools. This fragmentation creates significant challenges for studying impacts and feedbacks in these complex multiscale systems. These challenges are further exacerbated by ongoing disruptive changes in high-performance computational architectures and the exponential growth in the types and volume of data that render obsolete the conventional approaches to software development and data management. Overcoming these challenges will require the development of a BER Climate and Environmental Sciences Division (CESD)–wide enabling cyberinfrastructure to support data management, cross-domain modeling, data analysis, and collaborative research.

To explore the potential for working groups to initiate and guide a more integrated and community-based cyberinfrastructure, BER held the ESS Workshop on Model-Data Integration: Modeling Frameworks, Data Management, and Scientific Workflows on April 30–May 1, 2015, following the ESS Principal Investigator Meeting in Potomac, Maryland. Participants included model developers, software engineers, and data management specialists from eight national laboratories, which represented a wide range of projects and programs from CESD as well as cross-cutting projects from DOE's Office of Advanced Scientific Computing Research (ASCR; both BER and ASCR are operated from within DOE's Office of Science). A series of plenary talks provided background information and clarified the workshop's three objectives: (1) develop requirements for this community-based cyberinfrastructure to ensure enhanced scientific productivity of the community as a whole; (2) identify challenges associated with developing this new cyberinfrastructure using a phased approach guided by project-driven use cases; and, given these requirements and challenges, (3) chart a path forward for ESS working groups to lead the phased development of the new cyberinfrastructure.

To address these objectives, three breakout sessions were organized and intermixed with lightning talks that provided additional information. The first breakout session discussed requirements and challenges for near-term development of the community-based cyberinfrastructure (Phase 1: 0 to 2 years), and was tasked with identifying initial capabilities that could be developed under ongoing BER-funded projects. Capabilities and use cases were identified in areas of data management, model interoperability and coupling, complex model and data workflow, and provenance capture.

The second breakout session focused on requirements and challenges for longer-term development (Phase 2: 2 to 5 years; Phase 3: 5 to 10 years) and split discussions into two subtopics. Subtopic one targeted multiphysics-multiscale process coupling. A key finding was the

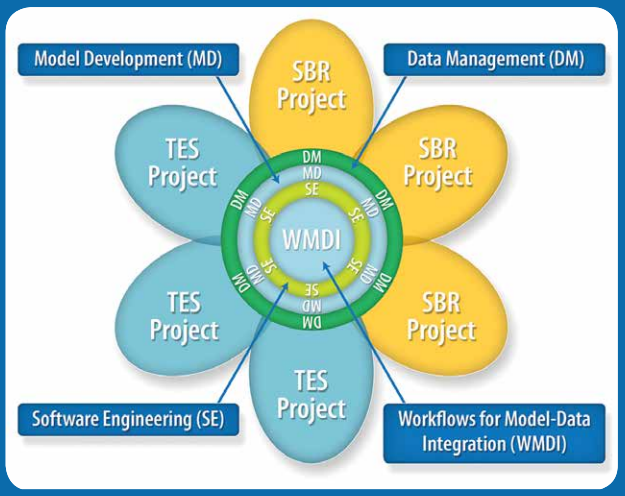
potential for community-based, flexible multiphysics and multiscale frameworks to enable sharing of capabilities across projects and scales to significantly enhance predictive understanding. Subtopic two examined model-data integration workflows and touched on issues surrounding the collection of model input data, model parameterization, initialization, and uncertainty quantification. This discussion identified several high-priority capabilities that would naturally be supported in a community-based cyberinfrastructure and significantly enhance scientific productivity, including metadata archiving, code sharing for parameterization, and modular parameter estimation and uncertainty quantification.

The third breakout session centered on operational issues associated with ESS working groups meeting the requirements and addressing the challenges identified in the first two breakout sessions. Three topics were identified for discussion: (1) working group governance and management, (2) setting of working group priorities and their relationships to existing ESS projects, and (3) licensing and intellectual property issues associated with the code and tools comprising the proposed community-based cyberinfrastructure. The first discussion identified governance as a critical factor in ensuring that working groups support the overall community and add value to existing ESS projects. The second discussion concluded that working group priorities should be set through input from the broader ESS community. Finally, the licensing and intellectual property discussion raised several important points relating to the established funding model, in which modeling capabilities are often considered a critical part of a team's competitive advantage. Specifically, shifting to a community-based cyberinfrastructure that enhances sharing of capabilities and accelerates the development of predictive understanding requires a research business model that acknowledges and even rewards these contributions. A complete solution to this complex and critical piece of community-based cyberinfrastructure does not yet exist, but the incredible growth of open-source scientific software provides a solid foundation upon which to build. In addition, other DOE offices (e.g., ASCR) and federal agencies (e.g., National Science Foundation) are facing these same challenges, creating collaboration and leveraging opportunities through the adoption of common policies and building of consensus across the broader scientific community on the approach and implementation.

Based on these workshop discussions, this report proposes a two-level structure for the CESD-ESS cyberinfrastructure working groups: an overarching executive committee and dynamically formed working groups to address specific topics and scope. Key topics identified for the initial set of working groups include data management, model-data integration, software engineering and interoperability, and community governance. The associated activities are well aligned with existing ESS projects and will benefit the community as a whole. This report recommends the launch of the ESS executive committee and formation of these four working groups.

Building a Cyberinfrastructure for Environmental System Science: Modeling Frameworks, Data Management, and Scientific Workflows

About the Cover: Envisioned is a community-driven cyberinfrastructure supporting Environmental System Science (ESS) activities within the Department of Energy's Office of Biological and Environmental Research (BER) to facilitate the iterative cycle of model-driven experimentation and observation and accelerate scientific discovery. This cyberinfrastructure will be developed in phases by a dynamic and coordinated set of working groups with expertise in model development, data management, software engineering, and workflows for model-data integration. Working groups will be formed by an overarching executive committee as high-priority needs are identified and will be dissolved when the specific tasks are completed. The executive committee will consist of representatives from major projects funded by the Terrestrial Ecosystem Science (TES) and Subsurface Biogeochemical Research (SBR) programs, which constitute BER's ESS activity. The committee will help to identify and prioritize the topics that the working groups address. Pursuit of these topics by the working groups will be based on specific use cases selected from existing ESS projects and designed to be of general utility to the broader ESS community.



This report can be downloaded at doesbr.org/ESS-WorkingGroups/.

BER Climate and Environmental Sciences Division

Contacts

David Lesmes
U.S. Department of Energy Office of Science
Office of Biological and Environmental Research
David.Lesmes@science.doe.gov, 301.903.2977

Justin Hnilo
U.S. Department of Energy Office of Science
Office of Biological and Environmental Research
Justin.Hnilo@science.doe.gov, 301.903.1399