EarthCube Strategic Vision

SUBMITTED: MAY 22, 2015

Authors: Farzad Kamalabadi, Basil Gomez, Yolanda Gil, David Arctur, Steve Richard, Jay Pearlman

EarthCube’s overarching goal1 is to catalyze the scientific mission of the geosciences community by leveraging advances in information science and technology to propel scientific progress and discovery in ways unimagined before.

The scientific scope of EarthCube entails knowledge that supports improved understanding of the Earth’s environment. Advances in scientific understanding address the critical societal needs for sustainable resource utilization, water and energy availability, effective preparation and response to extreme events and geo-hazards (e.g., droughts, floods, earthquakes, volcanic eruptions, hurricanes, solar storms, tornadoes and other types of severe weather), and adaptation to long-term changes in the environment, weather patterns, sea level and climate. In addition to enabling unprecedented progress and discovery in geosciences, EarthCube aims to enable systematic knowledge that will inform practices and policies regarding complex environmental issues and decision making in geosciences.2

The vision of EarthCube is to realize these lofty scientific goals by fostering modern modes of scientific inquiry that infuse technological innovations in the emerging practices of digital scholarship, data science and analytics, data and software stewardship, and open science, into basic research in geosciences.

Science Imperatives and Frontiers
The primary scientific goal of EarthCube is to enable geoscientists to make significant progress in understanding and mitigating complex, large scale environmental problems, such as: geospace variability and extreme events; planetary-scale (global) changes in the Earth system; geo-hazards; and water and energy sustainability. Specific objectives are: 1) to improve the utilization of scientific information in decision making designed to mitigate the impact of or facilitate adaptation to disruptive natural events (including solar storms, floods, landslides, volcanic eruptions, earthquakes and climate change), that occur across all relevant temporal and spatial scales; 2) to understand the impacts of climate change and direct human perturbations, such as land use change; and 3) to

1Mission Statement: EarthCube’s mission is to enable geoscientists to address the challenges of understanding and predicting a complex and evolving Earth system by fostering a community-governed effort to develop a common cyberinfrastructure to collect, access, analyze, share and visualize all forms of data and resources, using advanced technological and computational capabilities. See http://earthcube.org/document/2015/earthcube-charter
determine the magnitude, trajectory and time response of disruptive natural events and human perturbations on key Sun–Earth, solid Earth, hydrosphere, and atmosphere systems.

Addressing these objectives is predicated on the need to understand the co-evolution, operation and resultant configuration of coupled Earth systems, such as the climate–carbon system and the helio-, geo- and bio-spheres, during periods of stasis or (rapid) change. For example, characterization of how physical processes originating on the Sun affect human activities on Earth; how bio– and geo–chemical fluxes from the land surface to the coastal ocean are affected by the magnitude, duration, sequencing and spatial extent of atmospheric events; and how complex emergent properties in ocean ecosystems are related to physical, chemical, and biological processes. A specific need is to advance capabilities for identifying the processes responsible for initiating feedback that either sustains equilibrium or moves systems towards thresholds and tipping points (for example, through the influence cloud cover exerts on climate and the biosphere). Furthermore, it is important to determine if governing processes are scale–dependent, to understand the temporal and spatial variability of those processes, and to improve predictions of the impacts on human society.

The motivation for EarthCube–enabled science in support of these fundamental goals can be distilled into three essential science frontiers:

1. **To quantify** limits of prediction and better understand the constraints on and limits of data and model accuracy and utility.
2. **To characterize** the key processes, interactions, causations, and feedbacks operating at and across different temporal and spatial scales within physical, chemical, and biological domains.
3. **To deliver** a holistic, quantitative representation of critical physical, chemical, and biological states and fluxes, in order to inform fundamental science and societal decisions.

The imperatives for advancing these geoscience frontiers include enhancing capabilities to make significant progress in understanding, communicating about, and mitigating complex, large scale environmental problems, such as:

1. Understanding the consequences, impacts, and effects of planetary–scale (global) variability and changes in Earth systems, including recognizing the geophysical signal within the natural variability.
2. Increasing understanding of geohazards and extreme events, through the effective characterization and communication of uncertainty and relative risk.
3. Providing sustainable solutions for water, energy, and mineral resource use by defining mass and energy balances associated with past and present conditions to accurately project future states.
Technological Imperatives and Frontiers

Rapid and meaningful progress along the scientific imperatives and frontiers articulated above necessitates concurrent advances in computer and information science and technology. Cyberinfrastructure should make disciplinary boundaries permeable, nurture and facilitate knowledge sharing, cultivate unanticipated uses of information, and enhance collaborative pursuit of cross-disciplinary research. The science vision of understanding and predicting a complex and evolving Earth system can be catapulted by advancing research in data science and related cyberinfrastructures to collect, manage (archive and access), analyze, share, visualize, interpret, and understand all forms of relevant geoscience data.

In order to foster integrative research that considers holistic models of geoscience processes at different scales, EarthCube must have an analogous integrative approach on the technology side. The challenge is to utilize the extensive existing infrastructure resources as the foundation for an architecture that enhances their interoperation, accelerates growth to incorporate new advanced capabilities, and facilitates sustainability.

EarthCube is occurring at a pivotal time in scientific history. Scientific research products, whether data or software or articles, are not only of great interest to scientists in other fields but have well recognized societal value³. Open science practices are crucial to the dissemination of scientific knowledge, including digital scholarship, data and software stewardship, and routine reproducibility.⁴

The advent of the digital era has opened phenomenal possibilities, bringing cyberinfrastructure resources to many scientific communities and bringing data-driven scientific research to new levels. However, many of these powerful technologies have had uneven adoption and are often not well integrated into scientific practice. A more fundamental change needs to occur, making scientists active participants driving the requirements and providing continuous feedback to engineers as technology is developed. Crucially, EarthCube cyberinfrastructure will be socially organic and driven by geoscientists’ desire to discover and utilize multidisciplinary and multiscale datasets from distributed repositories.

Therefore, the core tenets of the EarthCube technology strategy are:

1. **Open science**: EarthCube needs to encourage the publication of all science products so they are discoverable and accessible, to enable reproducibility, and to ensure that they can be adapted to solve new problems.

2. **Knowledge-rich components**: Scientific resources and products must be described with rich metadata to enable others to understand and reuse them. Advanced analysis techniques must be developed -- as well as automated learning techniques from data -- to take advantage of knowledge-rich

---
³ [https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)
components for the effective, mining, fusion, and assimilation of data into sophisticated inference models.

3. **Federated organization:** EarthCube participants, from organizations to individuals, will contribute resources designed to interoperate through agreed standards. Rather than centralized control, EarthCube will provide coordination by fostering standards and integration.

Recognizing that technology is the facilitator but not the agent for social change, EarthCube will endeavor to engage the community in adopting these principles. EarthCube will actively disseminate among scientists novel technologies designed to improve science practice. EarthCube will develop and communicate best practices for interoperability and standards.