EarthCube Strategic Science Plan

Emma L. Aronson*, *University of California, Riverside*
Ann Bryant Burgess, *Computer Science Department, University of Southern California*
Hilary Close, *Department of Geology & Geophysics, University of Hawai'i, Mānoa*
Tony van Eyken, *Center for Geospace Studies, SRI International*
Basil Gomez, *Department of Geography, University of Hawai'i, Mānoa*
Anna Kelbert, *College of Earth, Ocean & Atmospheric Sciences, Oregon State University*
Raleigh L. Martin, *Department of Atmospheric & Oceanic Sciences, University of California, Los Angeles*
Ken Rubin*, *Department of Geology & Geophysics, University of Hawai'i, Mānoa*
Andreas Schmittner, *College of Earth, Ocean & Atmospheric Sciences, Oregon State University*
Stephen Slota, *Informatics Department, University of California, Irvine*
D. Sarah Stamps, *Department of Earth, Planetary & Space Sciences, University of California, Los Angeles*
Karen Stocks, *Geological Data Center, Scripps Institution of Oceanography*
Mimi W. Tzeng*, *Dauphin Island Sea Lab*

*Working Group Chair

Draft: January 26, 2015

EarthCube Working Paper ECWP-2015-1

Table of Contents

Executive Summary

This working paper is a step towards the first iteration of a long-term strategy for EarthCube that will articulate what EarthCube-enabled science looks like. It was compiled, at the request of EarthCube's Science Committee, by three Working Groups formed at the end of 2014. The Working Groups respectively were tasked with the responsibilities of: characterizing the science drivers identified by participants of the 24 EarthCube End-User Workshops; detecting patterns in the objectives currently funded EarthCube projects seek to fulfill and determining if these objectives meet end-user requirements; and identifying 'grand challenges' for EarthCube that have emerged as common themes from the end-user workshops, in discussions among EarthCube's Science Committee members, and from the geosciences community as a whole.

Community discussion of the science drivers commonly revolved around the topics of 'change' and 'processes'. Those science drivers end-user workshop participants specifically identified reflect the collective need for a better understanding of the dynamics of coupled solid-earth, hydrosphere, atmosphere systems; the constituent physical, chemical and biological processes; and their interactions at all temporal and spatial scales. Particular objectives were to improve the utilization of scientific information in decision making designed to mitigate the impact of or facilitate adaption to disruptive natural events, climate change impacts and direct human perturbations; and to determine the magnitude, trajectory and time response of disruptive natural events and human perturbations on key solid-earth, hydrosphere, and atmosphere systems.

Only science-oriented funded projects could be easily aligned with any science driver. Moreover, although they are implicitly meant to complement each other, and there is considerable potential for promoting synergistic opportunities, interactions among funded EarthCube projects have been relatively limited thus far. An obvious impediment to progress is the lack of defined pathways and bridges along and across which such interactions can occur.

Grand Challenges that provide the broadest community motivation to pursue EarthCube include complex, interdisciplinary topics such as planetary-scale (global) changes in the Earth system; geohazards; and water and energy sustainability. Targeted, focused and immediate action also is required, to overcome potential barriers created by the conditions under which EarthCube is currently operating – specifically, the decision to actively fund cyberinfrastructure development while potential community-driven uses for that infrastructure are still being defined.

The successful implementation of Earth System Models may be key to making significant progress in understanding, communicating about, and mitigating many complex, large scale environmental problems. However, it remains to determine which primary motivator of the EarthCube project (*data* or *cyberinfrastructure* availability) will have a more immediate impact on any overarching scientific theme.

1. Background[1]

EarthCube began in 2011 as a joint initiative between the National Science Foundation (NSF) Directorate for Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI). Envisioned as an evolving, dynamic community effort, EarthCube is not only a new way for the NSF to partner with the scientific community, but also a challenge for the many academic, agency and industry stakeholders in the geo-, cyberinfrastructure, computer and social sciences to create new capabilities for sharing data and knowledge and conducting research.

EarthCube's goal 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. EarthCube's vision is to create a dynamic, community-driven cyberinfrastructure that will support standards for interoperability, infuse advanced technologies to improve and facilitate interdisciplinary research, and help educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science.

EarthCube will be supported by the existing foundation of cyberinfrastructure investments, including databases, software services and community facilities that have been created by the geosciences and cyberinfrastructure communities over the past two decades. Achieving the aforementioned objectives also requires a long-term effort, which the NSF anticipates supporting until at least 2022. It also requires that the constituent geoscience communities articulate their science goals and cyberinfrastructure needs, so that common themes, challenges and synergies can be identified and merged into a communal roadmap.

2. End-User Workshops

Beginning in summer 2012, NSF funded a series of 24 EarthCube domain end-user workshops[2]. These workshops targeted a broad spectrum of Earth, atmosphere, ocean, and allied senior, mid- and early-career scientists, introduced the ~1500 participants to EarthCube, and encouraged them to think about how data-enabled science could help them achieve their scientific goals. An overarching goal of the workshops was to gather information about the science drivers and data utilities, and the requirements for user-interfaces, models, software, tools, *etc*. with the objective of ensuring that EarthCube is designed to help geoscientists more easily do the science they want and would like to accomplish. That is, EarthCube should help foster a sustainable future through a better understanding of our complex and changing planet, and enable the geosciences community

[1] see: *EarthCube: Past, Present, and Future.* Yolanda Gil, Marjorie Chan, Basil Gomez and Bruce Caron (Eds). EarthCube Project Report EC-2014-3, December 2014 *http://www.earthcube.org/documents/2014/3/EarthCube_Past_Present_and_Future*

[2] *http://www.earthcube.org/page/end-user-workshops*

to develop a framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core[3].

To help better define the geoscience community's needs, an 'End-User Principal Investigator' workshop was held August 14-15, 2013, in Tucson, AZ, with the objective of synthesizing outcomes from 16 completed end-user workshops[4]. The most widely identified needs were related to data accessibility, discovery, curation, and integration. A common thread was the need to make these operations easier and less time consuming. Ideally this would occur by: (1) enhancing software tools and processing capability; and (2) supporting the development and adoption of community conventions and standards for metadata, data, and software that would facilitate data management, documentation, exchange, and analysis. From a practical point of view these challenges with data arise because, at present, utilizing non-standard, heterogeneous data from different sources requires significant effort to analyze each dataset for content and determine how to integrate it with other data. The lack of standard vocabularies for specifying data schema and property values complicates the problem because the meaning, quality and uncertainty of the data are often unclear and inconsistent practices for data sharing make each new data acquisition a time consuming learning experience. Best measurement practices and standards to facilitate knowledge sharing are also required. One of EarthCube's overarching goals is to streamline these processes and allow scientists to locate, access, store and share data in ways that facilitate and streamline their research.

2.1 Science drivers

Although a majority of the domain workshops involved participants drawn from the Earth Sciences, the science drivers they identified reflect the collective need geoscientists recognize exists for a better understanding of the dynamics of coupled solid-earth, hydrosphere, atmosphere systems; the constituent physical, chemical and biological processes; and their interactions at all temporal and spatial scales. This includes, for example, understanding the role of fluids in seismicity and tectonics, and how the structure of the upper mantle in a given location is related to surficial geological processes and mantle convection; as well as the processes and interactions that create geological structures, shape landscapes and govern mass fluxes of water, carbon, nutrients and erosion products.

Participants also highlighted the need to understand the co-evolution, operation and resultant configuration of coupled Earth systems, such as the climate-carbon system and the geo- and bio-spheres, during periods of stasis or (rapid) change. A specific objective is to advance capabilities for identifying the processes responsible for: generating heterogeneity (in, for example, the stratigraphic record); and initiating feedback that either

[3] *EarthCube Guidance for the Community,* NSF11085.

[4] See: *EarthCube End-User Principal Investigator Workshop, Executive Summary.*

sustains equilibrium or moves systems towards thresholds and tipping points (through, for example, the influence cloud cover exerts on climate and the biosphere). They also emphasized it is important to know if the governing processes are scale-dependent and to understand the transformations to those processes that occur through time and space. For example, how bio- and geo-chemical fluxes from the land surface to the coastal ocean are affected by event magnitude, duration, sequencing and spatial extent; and how complex emergent properties in ocean ecosystems are created by different physical, chemical, and biological processes.

Information derived from the end-user workshops shows that individual geoscience domains routinely utilize about 30 distinct data sources[5], and all the workshops recognized the need: to obtain more accurate and complete spatial and temporal observation data for the state variables that characterize solid-earth, hydrosphere, and atmosphere systems; and for EarthCube to facilitate integration of multi-scale, multi-domain data. Other methodological issues that were identified included the need to better integrate: coupled models at different temporal and spatial scales; and the results derived from ensembles of models. Two objectives being: (1) to improve the utilization of scientific information in decision making designed to mitigate the impact of or facilitate adaption to disruptive natural events (including floods, earthquakes and climate change), that occur across all temporal and spatial scales; and climate change impacts and direct human perturbations, such as land use change; and (2) to determine the magnitude, trajectory and time response of disruptive natural events and human perturbations on key solid-earth, hydrosphere, and atmosphere systems.

2.2 Common or overarching themes

Discussion of the science drivers identified by workshop participants commonly revolved around the topics of ‘change’ and ‘processes’; **five overarching themes** emerged. In one way or another, all of these themes call for the integration or synthesis of data and information across different scales and domains. Their direct relevance to EarthCube is that, in all cases, this is presently only possible to a limited extent.

Figure 1: *Word cloud derived from the Science Drivers section of the 24 end-user workshop executive summaries (the size of a word is proportional the frequency of its occurrence).*

[5] Gomez, B., Pearthree, G. and Cutcher-Gershenfeld, J., 2013, Cyber-infrastructure and synergistic opportunities across the Geosciences, Eos, Transactions American Geophysical Union, 94, 473.

1) There is a need to identify and characterize the key processes, interactions and feedbacks operating at and across different temporal and spatial scales and physical domains; identify the primary drivers of change; and integrate all these factors into models, in order to better account for spatial and temporal variability seen in solid-earth, hydrosphere and atmosphere systems.

2) There is a need to identify and discriminate between the effects disruptive natural events and human perturbations have on solid-earth, hydrosphere and atmosphere systems; identify their respective scales of influence, and the magnitudes of the effects they exert. Better understanding and prediction of these events and perturbations is essential if society is to adapt to and mitigate the impacts of environmental hazards and global change.

3) There is a need to improve predictability and better understand the constraints on and limits of model accuracy. In some cases, integration of real-time data holds promise for improved predictability of disruptive natural events, such as hurricanes and earthquakes. Whereas in other cases more static data are required to understand, for example, how climate change alters ocean chemistry, and impacts corals and the organisms that use coral reefs as habitat.

4) There is a need to obtain improved estimates of the flux / migration of energy, mass, fluids, sediments, nutrients, and carbon within and between the different solid-earth, hydrosphere and atmosphere systems. Such information is essential if the state and functioning of the complex Earth systems and their components are to be fully understood.

5) There is a need to better document the current state and past evolution of the different solid-earth, hydrosphere and atmosphere systems and their component parts. The rationale is that if we cannot comprehend how the past evolved into the present, we are unlikely to be able to make accurate projections and predictions for the future.

Critically, in order to cultivate future generations of researchers and also for EarthCube to be of use to policy and decision makers, geoscientists must be able to retain access to and communicate the results and uncertainties of information and research that advances knowledge to society.

3. Funded Projects

Following from the 2009 Advisory Committee for GEO GEOVision report[6], which identified the challenges and opportunities facing the geosciences in the next decade, and the 2011 Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21) initiative, which emphasized the importance of enabling computational and data-rich science, the National Science Foundation (NSF) began evaluating proposals[7], in Spring 2013, for: an EarthCube Test Enterprise Governance; EarthCube Research Coordination Networks (RCNs); EarthCube Building Blocks; and EarthCube Conceptual Designs.

Three strategic goals were at the core of the first solicitations[8]:

1. **engage** all stakeholders, including geoscientists, computer science and cyberinfrastructure specialists, and data managers and facilities, to create structure and begin closer collaboration and coordination with one another;
2. **build** on existing resources, with the recognition that not all research communities are equally well-served;
3. **begin** an iterative process over a ten year period that provides opportunities, then collects community input and assessment on an annual basis in order to accommodate needs, change and new developments[9].

Test Governance was expected to take documents released in August and September 2012 (a roadmap[10] and a framework proposal[11]), that were the end result of a year's initial research and community outreach, use them to create a working governance model, and put the model into practice to test its effectiveness.

Research Coordination Networks (RCNs[12]) are intended to create networking opportunities and multidisciplinary partnerships between geoscientists, cyberinfrastructure specialists, and data managers, and encourage closer cross-collaboration and coordination among these disparate groups.

Building Blocks (BBs) are intended to construct a cyberinfrastructure to better connect existing resources, integrate and develop resources that would serve broader communities, and begin the initial work of EarthCube.

---

[6] http://www.nsf.gov/geo/acgeo/geovision/geo_strategic_plans_2012.pdf

[7] **N.B.** The first end-user workshops were being held *at the same time* as the NSF was evaluating the first proposals.

[8] The fourth Amendment to the EarthCube solicitation also contains a Request for Proposals (deadline March 19, 2015) for EarthCube Integrative Activities designed to enable geoscientists to participate in EarthCube, which are not considered here.

[9] This report represents the first iteration of the community input and assessment process.

[10] http://www.nsf.gov/geo/earthcube/docs/EarthCubeGovernanceRoadmap.pdf

[11] http://www.nsf.gov/geo/earthcube/docs/EarthCubeGovernanceFramework.pdf

[12] Which continue under the fourth Amendment to the EarthCube solicitation http://www.nsf.gov/pubs/2013/nsf13529/nsf13529.htm

Conceptual Designs (CDs) represent the initial planning stage for the EarthCube architecture, with the goals of better understanding the landscape of existing resources and promoting innovative designs for the evolving system.

Based on the stated activities, objectives, and themes of each funded project, it is apparent that only science-oriented funded projects (*i.e.*, the RCNs and the BiG CZ SSI, funded through the Software Infrastructure for Sustained Innovation (SI^2) program[13]), could be easily aligned with the science drivers identified earlier in this report. In contrast, the more technically-oriented BBs and CDs cannot easily be related to specific science themes. This is because their intent is to provide the basic tools for scientists to facilitate and advance the work they do. However, by focusing on the broader technical themes related to different scientific approaches and the technical tools required to enable these approaches, it is possible to gain a perspective on the gaps that remain if the synergies between these projects were exploited.

To date, the NSF has funded a total of 25 EarthCube projects: a test governance project; 6 RCNs; 15 BBs; 3 CDs; and the BiG CZ SSI project which is intended to integrate the work of the Critical Zone Observatories (CZOs)[14].

The EarthCube Test Enterprise Governance (ECTEG) funded project has a charge of producing a charter/framework for EarthCube Governance (which includes the Leadership Council, the Science and Technology and Architecture Committees, the Council for Data Facilities, the Engagement and Liaison Teams, and the EarthCube Office). This document is the direct result of the charge given to the ECTEG to assess user requirements and gaps in those requirement as defined by the outcomes of the end-user workshops.

The other 24 funded projects deal with more specific science goals and technical needs. For example, the RCNs are intended to provide opportunities for academic geosciences communities to organize, seek input, come to consensus and prioritize data, modeling, and technology needs, as well as standards and interoperability within and across domains. BiG CZ SSI likewise addresses the science needs of the constituent CZO projects.

Though they are implicitly meant to complement each other, and there is considerable potential for promoting synergistic opportunities, interactions among EarthCube funded projects have been relatively limited thus far. These interactions are expected to develop. However, an obvious impediment to progress is the lack of defined pathways and bridges along and across which such interactions can occur. Potential pathways might, for example, be defined on the basis of commonly expressed technical needs and the solutions to these needs. On this basis, four opportunities for promoting interactions between different funded projects emerge: *1)* integration of data, methods, or models; *2)* linking observations across scales; *3)* integrating unique sample types and observational techniques; and *4)* building of earth-system models. These opportunities are not seen as

[13] http://www.nsf.gov/si2/

[14] A Critical Zone Observatory is an environmental laboratory, focused on the interconnected chemical, physical and biological processes shaping Earth's surface http://criticalzone.org/national/

being mutually exclusive, but rather they provide a means for identifying ways in which different funded projects can work together and begin to use new cyberinfrastructure tools to address specific scientific needs. Their perceived composition is outlined below:

Integrating data, methods, and models

- RCN SEN: A Sediment Experimentalist Network
- BB GeoWS: Geoscience Web Services
- BB GeoLink: Semantics and Linked Data for Data Sharing and Discovery
- BB BCube: Brokering Technologies to Discover, Share, and Access Data
- BB Geosemantics: Integrating Long-Tail Data and Models
- BB CINERGI: Inventory of Resources for Interoperability
- BB Geosoft: Software Stewardship and Open Source Software Sharing

These projects are linked via their goals: to better share and integrate scientific data, tools, models, methods, and other products; and their common desire to integrate long-tail data (*i.e.*, the products of individual investigators) into publically available searchable data resources.

Linking observations across scales

- RCN ECOGEO: Oceanography and Geobiology Environment Omics
- BB ODSIP: Open Data Services Invocation Protocol
- BB DisConBB: Integrating Discrete and Continuous Data
- BB Earth Systems Bridge: Interoperable Modeling Frameworks
- BB Collaboration and Discovery through Semantic Connections
- CD A Scalable Community Driven Architecture

The focus of these projects is on the integration, and best practices of collection and curation of data of multiple types and at multiple scales. Their shared intent is also to increase the interoperability, and reduce semantic discrepancy, of the architecture used to visualize and model data.

Integrating unique sample types and observational techniques

- RCN EC3: Challenges of Field Data Collection, Management, and Integration
- RCN iSamplES: The Internet of Samples in the Earth Sciences
- BB GeoDataspace: Data Management for Geoscience Models
- BB Digital Crust: Exploratory Environment for Research and Learning
- BB GeoDeepDive: Cognitive Computer Infrastructure
- CD A Data-Oriented Human-Centric Enterprise Architecture

Primary concerns of these projects are: the integration of different types of data or models into a digital cyberinfrastructure; and the development of analytical tools that will be of use across all geoscience domains.

Modeling Earth Systems

- RCN C4P: Collaboration and Cyberinfrastructure for Paleogeosciences
- RCN CReSCyNT: Coral Reef Science and Cyberinfrastructure Network
- BB CHORDS: Cloud-hosted Real-Time Data Services
- BB CyberConnector: Model Validation, Verification, and Intercomparison
- CD A Cross-Domain Integrative Information System
- Other BiG CZ SSI

By reinforcing and developing new capabilities for existing modes of work, rather than proposing entirely new workflows, these projects are more closely aligned with the science drivers identified earlier in this report. To facilitate the development, interconnection and comparison of Earth Systems Models and enhance their utility, they have a common interest in improving the quantity and quality of data available, and in the management of both real-time and historic data across heterogeneous temporal scales.

As these projects mature, it will become possible to evaluate whether the funded projects diverge or converge with the scientific drivers and themes identified herein and, crucially, with EarthCube's Vision[15] and Mission[16]. It will also be possible to determine how these activities have evolved with respect to the landscape envisaged in the EarthCube Community Groups' and Concept Teams' roadmaps[17]. Ultimately, it will be imperative to attempt to integrate these projects and select for interoperability. Future funding decisions made with the scientific Grand Challenges, identified below, in mind may be directly applicable to the needs of the geoscience community at large.

4. Grand Challenges

Grand Challenges for EarthCube that have emerged as common themes from the end-user workshops, in discussions among EarthCube's Science Committee members, and from the geosciences community as a whole fall into two categories: *1)* aspirations or expectations; and *2)* barriers that need to be overcome.

A particularly poignant theme is the expectation that EarthCube has the potential to galvanize the broader geoscience community and facilitate better relations with society as a whole through its problem-solving capabilities. Specifically, EarthCube should enable

[15] *EarthCube's long-term vision is a community-driven, dynamic cyberinfrastructure that will support standards for interoperability, infuse advanced technologies to improve and facilitate interdisciplinary research, and help educate scientists in the emerging practices of digital scholarship, data and software stewardship, and open science.*

[16] *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.*

[17] In March 2012, the NSF formed and funded a number of EarthCube Community Groups and Concept Award Teams, each of which was tasked with producing a roadmap to help move their area of EarthCube forward http://workspace.earthcube.org/type-doc/roadmaps

geoscientists to make significant progress in understanding, communicating about, and mitigating complex, large scale (wicked) environmental problems. Examples include: planetary-scale (global) changes in the Earth system; geohazards; and water and energy sustainability. Geoscientists' ability to make major advances on these complex, interdisciplinary topics often is constrained by their ability to access and analyze large and diverse data sets. That is, in order to tame these wicked problems geoscientists require access to a dynamic and sustainable cyberinfrastructure that will allow them to compile, analyze, visualize and share a wide array of data types and resources. The availability of such resources would also allow geoscientists to deconstruct the workings of complex natural systems and better understand how they are being perturbed by human activities. Geoscientists have repeatedly stressed and reemphasized these intellectual aspirations. For example in her 2000 Presidential Address to the Geological Society of America, Mary Lou Zobak[18] articulated that the big environmental problems, the grand challenges of the coming decade were:

- Recognizing the signal within the natural variability
- Defining mass flux and energy balance in natural systems
- Identifying feedback between natural and perturbed systems
- Determining proxies for biodiversity and ecosystem health
- Quantifying consequences, impacts, and effects
- Effectively communicating uncertainty and relative risk

A decade and a half later many of the end-user workshop participants spoke to these same grand challenges, albeit in more technical and community-oriented ways, and they provide the broadest community motivation to pursue EarthCube. Moreover, beyond this grandest overarching aspiration, the individual domain science communities also recognize that important advances could also be made on pressing and long-recognized complex research problems in their own disciplines should EarthCube's vision be realized, further strengthening support for its development

For these things to happen, targeted, focused and immediate action is required, to overcome potential barriers created by the conditions under which EarthCube is currently operating – specifically, the decision to actively fund cyberinfrastructure development while potential community-driven uses for that infrastructure are still being defined. The EarthCube community also had to overcome the social challenge of rapidly developing successful working relationships among the members of its primary constituencies (domain geoscientists and technologists), each of which has a different understanding of

---

[18] Zoback, M.L., 2001, Grand Challenges in Earth and Environmental Sciences: Science, Stewardship, and Service for the Twenty-First Century, *GSA Today*, 11 (December 2001), 41 – 47.

the key issues involved[19]; utilizes different modes of communication; and collectively have limited previous experience of working together.

EarthCube explicitly was conceptualized in a very general way so as not to predetermine for the community what it would include (or exclude). However, we must now address both the intellectual aspirations of the constituent end-user communities, which may have little to no understanding of the technological requirements that need to be fulfilled for them to realize them; and the drive to produce the cyberinfrastructure by technologists who know what can and can't be done, but may not fully grasp how the infrastructure will be used, nor have received clear instructions about priority use-cases.

Progress is being made on these all fronts, but the need to rapidly address these operational barriers in a way that engenders community trust is paramount for achieving the goal of successfully building an exceptional and unequaled sustainable infrastructure to enable truly transformative research advances and our ability to communicate them.

5. Scope and Vision of EarthCube-Enabled Science

This working paper is a step towards the first iteration of a long-term strategy for EarthCube that will articulate what EarthCube-enabled science looks like and how it matches the NSF's science vision. This, in turn, will help add value to the activities of the majority of domain geoscientists, by enabling them to better collect, access, analyze, share and visualize all forms of data and resources, using advanced technological and computational capabilities. Key to this is the requirement that the long-term sustainability of all funded projects be ensured, by enabling opportunities for them to work together and develop interoperability in the short-term.

Impediments to achieving the aforementioned objective include: a dearth of workflow systems[20] that can help geoscientists select models appropriate for their data, configure them with appropriate parameters, and execute them efficiently[21]; incongruities in model conceptualizations and data structures between different scientific and user communities; and technological divides between communities that rely on different standards or conventions for model and dataset/resource development[22]. Indeed, it was recognized almost from the outset of EarthCube's development, that the combination of the variety and complexity of models and the usability of existing diverse data and resources presents

[19] **N.B.** Currently, social scientists are not formally involved in EarthCube Test Governance.

[20] In their broadest sense, all scientific activities can be envisaged as collections of interdependent steps, such as gathering and analyzing data, represented as workflows.

[21] A Workflows Roadmap for the Geosciences http://workspace.earthcube.org/content/workflows-roadmap-geosciences

[22] EarthCube Earth System Model Coupling Roadmap http://workspace.earthcube.org/content/earthcube-earth-system-model-coupling-roadmap

a key challenge to the geosciences[23]. Nonetheless, Earth System Models which attempt to represent the major components and drivers of the atmosphere, biosphere, geosphere, hydrosphere, and anthroposphere in an integrated manner, have the potential to put powerful new hypothesis-testing capabilities into the hands of geoscientists that should enable new discoveries and pathways to understanding our earth system[24]. For this reason their successful implementation is key to making significant progress in understanding, communicating about, and mitigating the complex, large scale environmental problems that provide the broadest community motivation to pursue EarthCube.

Significant questions also remain to be addressed at the level of the constituent EarthCube communities, where better knowledge of the processes involved (such as river flow, atmospheric general circulation, and ocean biogeochemistry) is predicated on the integration or synthesis of data and information across different scales and domains. For example, is meaningful short-term progress on any or all of the five overarching themes to emerge from the end-user workshops more dependent on all data being universally accessible, or on data that currently are available and the development of innovative techniques for analyzing them, visualizing them, *etc.*? In other words, which primary motivator of the EarthCube project (*data* or *cyberinfrastructure* availability) will likely have a more immediate impact on these overarching scientific themes?

[23] Open Hydrospheric Modeling Framework Roadmap http://workspace.earthcube.org/content/open-hydrospheric-modeling-framework-roadmap
[24] *ibid* EarthCube Earth System Model Coupling Roadmap