EARTHCUBE END-USER PRINCIPAL INVESTIGATOR WORKSHOP

Executive Summary

August 14-15, 2013
Tucson, AZ

Workshop Convener:
M. Lee Allison, Arizona Geological Survey

Workshop Organizing Committee:
Joel Cutcher-Gershenfeld, University of Illinois
Kim Patten, Arizona Geological Survey
Genevieve Pearthree, Arizona Geological Survey
Erin Robinson, Foundation for Earth Science
Steve Richard, Arizona Geological Survey

With Support From:
Rahul Ramachandran
The University of Alabama in Huntsville
National Science Foundation, Award EAR-124039

INTRODUCTION

A two-day 'End User Principal Investigator' workshop was held August 14-15 in Tucson, Arizona, bringing together geoscience domain and cyberinfrastructure scientists who are organizers or participants in the two-dozen EarthCube End-User Workshops. They were joined by a small number of cyberinfrastructure specialists and social scientists, whose purpose was primarily to listen to the presentations and discussions, and provide input on ways to fulfill identified needs in the short and long-term.

The goals of this workshop were to synthesize outcomes from the completed end-user workshops, increase communication among the scientific domains represented, assist with planning for upcoming workshops, and to lay the groundwork for producing documents to inform and guide the EarthCube community, including upcoming End-User Workshop organizers, and NSF EarthCube awards (Test Enterprise Governance, Building Blocks, Conceptual Designs, and Research Coordination Networks).

The 32 in-person and 8 virtual workshop participants represented a wide variety of scientific and technological fields, EarthCube End-User Workshops and EarthCube Working Groups (groups funded to write roadmaps prior to the June 2012 EarthCube charrette):

EarthCube End-User Workshops

1. MYRES (Meeting of Young Researchers in Earth Sciences) V: The Sedimentary Record of Landscape Dynamics
2. Envisioning Success: A Workshop for Next Generation EarthCube Scholars and Scientists
3. Structural Geology and Tectonics
4. EarthScope
5. Calling All Experimentalists- Experimental Stratigraphy
6. Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction
7. Engaging the Critical Zone Community to Bridge Long Tail Science with Big Data
8. Envisioning a Digital Crust for Simulating Continental Scale Subsurface Fluid Flow in Earth System Models (Hydrology/Dark Data)
9. Modeling
10. Cyberinfrastructure for Paleogeoscience
11. Education
12. Community-based Cyberinfrastructure for Petrology, Geochemistry, and Volcanology
13. Sedimentary Geology
14. Integrating the Inland-Waters Geochemistry, Biogeochemistry and Fluvial Sedimentology Communities
15. Deep Seafloor Processes & Dynamics
16. Integrating Real-Time Data Into the EarthCube Framework
17. Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community
18. Increasing Access to and Relevance of Marine Seismic Data
19. Bringing Geochronology into the EarthCube Framework
20. Rock Deformation and Mineral Physics Research
21. Community-Based Cyberinfrastructure for Polar Science Instrumentation, Technology, and Environmental Monitors

EarthCube Working Groups

1. Cross-Domain Interoperability
2. Data Discovery, Mining and Access
3. Earth System Model
4. EarthCube Stakeholder Alignment Survey
5. Governance Framework
6. Physical Samples as Part of Cyberinfrastructure
7. Workflow

A representative from each of the previous workshops summarized the science drivers, challenges, needs, and action items identified in previous workshops. We have synthesized these results by categorizing the various factors identified, empirically defining facets and categories in those facets, and then grouping in the categories. The following section is a discussion of this synthesis.

PREVIOUS WORKSHOPS: SUMMARY OF RESULTS

Science Drivers

Science drivers identified by the user workshops have been categorized to synthesize the results. Dynamic Earth system models are central to science drivers identified by 11 workshops, science methodology related drivers were identified by 6 workshops, and development of shared 3-D Earth models was identified by 5 workshops. Other science drivers mentioned are related to education, biology, new discoveries, planetary models and space weather. Science drivers for each category are presented in detail below.

EARTH SYSTEM MODELS

Many of the issues in critical zone science, climate modeling, hazard assessment, and anthropogenic impacts require more accurate dynamic Earth system models that have greater spatial and temporal resolution, and account for more of the non-linear feedback processes in the Earth system. These models require improvements in modeling algorithms, better understanding of the dynamics of coupled solid-earth, hydrosphere, atmosphere systems, increased computational capabilities, and more accurate and complete spatial and temporal observation data for the state variables that characterize the system.

SCIENCE METHOD

Drivers related to scientific methodology were focused on computational approaches to improved modeling capabilities, optimizing sampling and observation strategy, knowledge representation, and linking scientific results to decision-making processes. The most challenging modeling issues appear to be related to integrating coupled models at different temporal and spatial scales for systems with varying degrees of coupling, and integrating results from ensembles of models (particularly for weather). The limits of model accuracy, both from a computational point of view, and in light of the natural variability of the Earth system need to be better understood. The high cost of obtaining observational data from samples and sensors to understand processes and to enhance dynamic Earth models is motivating investigations of optimized sampling techniques, particularly for field experiments that rely on real-time data. Analysis of events captured by real-time sensor networks requires data and derived products to be presented in a timely fashion and in a form that is useful to decision-makers. More sophisticated approaches to knowledge representation in computational systems is needed to facilitate integration of multi-scale, multi-domain data, and to make results more readily accessible for real-time analysis and understanding, both as part of the experimental process and to improve utilization of scientific information in decision making.

Solid Earth Models

The need for high-resolution three-dimensional models of the present state of the solid Earth are implicit in science drivers related to tectonics and a better understanding of Earth history. We cannot experimentally test tectonic hypotheses for Earth dynamics because the time scales, temperatures, pressures, and chemical environments of the Earth's interior cannot (for the most part) be reproduced in the lab. Much of our understanding of the Earth is based on abductive reasoning supported by geologic mapping on the surface, borehole data in the near surface crust, and by probing the Earth's interior with various geophysical techniques. An Earth Model resource that supports registration and integration of 3-D models from the community in a dynamic information system enabling feedback, updates, multiple interpretations, and provenance tracking is envisioned as a foundation resource for all kinds of geoscience research. Such a resource would improve reproducibility, reduce duplication of effort, and streamline project development through more straightforward access to available legacy research.

Miscellaneous

Not surprisingly, there were other science drivers identified by various workshop participants that reflect the widely varying interests of the community. Education was brought up as an issue at several workshops, highlighting the importance of cultivating the next generation of cyber-savvy Earth Science researchers. The issues identified centered on how to teach skill sets and habits of mind necessary to use data and models, address novel and ill-structured problems, work in collaborative teams, adopt appropriate data and analysis strategies, and effectively communicate. Some other miscellaneous drivers included understanding the physical limitations for evolution of life, understanding planetary formation, prediction of large solar flare events, and discovery of new, unexpected phenomenon.

Challenges

The top challenges identified by previous workshops related to data (15 workshops), standards (11 workshops), and incentives (7 workshops). Challenges related to data can be categorized into curation, integration, access and documentation.

Data curation challenges (9 workshops) are associated with the process of preparing data for archive and reuse, including transformation into archive formats, documentation, placing it into a secure yet accessible repository, and curation of physical resources such as samples. Many of the 'dark data' issues stem from the difficulty that individual researchers, operating on very limited budgets, experience in trying to curate data produced by their research. Specific challenges include lack of tools, conversion to digital formats, and lack of clearly defined best practices.

Data integration (7 workshops) challenges address processes and capabilities to transform multiple datasets into formats and schema so they can be used together. Issues include data heterogeneity, lack of documentation (provenance, quality, semantics...), different spatial reference systems, and integrating data collected at various spatial and temporal scales or with different sampling strategies.

Data access (9 workshops) challenges relate to processes and capabilities required to allow data consumers to get and use data, including repository architecture, data services, real-time access, definition of information exchanges, and registration of resources in catalogs. Some specific challenges mentioned include tool complexity, access restrictions, bandwidth limitations and difficulty exporting data from databases.

Data documentation (7 workshops) challenges are related to specifications and practices for documenting the content, format (schema, encoding), semantics, provenance, quality, and stewardship of resources. Such documentation is essential to enable cross domain use of data, or reuse/repurposing of data obtained from repositories. Specific issues mentioned include reproducibility of model results, ability to assess uncertainty, difficulty discovering and evaluating data for research purpose, and citing the origin of data.

Standards issues relate to development and adoption of community conventions for data management, documentation, exchange, and analysis. To utilize non-standard, heterogeneous data from different sources requires significant effort to analyze each dataset to understand its content and determine how to transform it to integrate with other data. Lack of standard vocabularies for specifying data schema and property values complicates the problem because the meaning of the data may be unclear. Inconsistent practices for data sharing make each new data acquisition a time consuming learning experience.

Incentives challenges relate to social and financial factors that motivate good data management practices and the sharing of data or models. Lack of credit for data publication, either in the form of citation or career advancement is a major factor. Cost, effort, technical barriers, and concern about misuse outweigh the tangible benefits, especially for tenure track and project-funded researchers.

Some other challenges mentioned include misuse of data, communication between researchers in different domains, cost, absence of data, connecting target users with relevant data, lack of training, and insufficient computational capabilities.

Needs

The most widely identified needs are related to data (12 workshops). As with the challenges, the needs are grouped in to access, discovery, curation, and integration categories. A common thread was the need to make these operations easier and less time consuming, and the inclusion of a broad spectrum of resources included not only data, but models, workflows, samples, and tools. Data access needs (10 workshops) also mentioned the ability to subset large datasets, cross-domain data access, and integrating data access seamlessly into processing workflows. Discovery needs (6 workshops) identified included a single catalog for searching across the gamut of geoscience domains, map-based, space-time ('4-D') searching, and 'smart' searching technology to improve search efficiency. Related to data curation (8 workshops), the need for incentives or change in culture to promote better data management and data sharing was highlighted. Data integration needs (6 workshops) mentioned the need to integrate data in space or time, and to develop better approaches to integrating noisy or sparse data.

A common thread through many of the identified needs was the need for better software tools (10 workshops) to facilitate data documentation, archiving, and publication, data discovery and evaluation, data integration, and data analysis. Data visualization tools in particular were mentioned in issues raised by 7 workshops. In the data integration arena, improved capabilities for subsetting large datasets, transforming between schema, up-scaling and down-scaling, and resampling were mentioned.

The needs for standards (7 workshops) included mentions of standards for web interfaces to data, metadata and data formats, and standardized software libraries. Education needs were prominent (6 workshops), both to develop the workforce to support cyberinfrastructure, and to inform scientists about available technology and how to use it. Also mentioned were the needs to link data with other resources (datasets, samples, projects, researchers, and tools), shared domain databases, improved communication and social networking, governance to coordinate programs and pilot projects, develop metrics, and identify gaps, and the need for funding.

Action Items

The suggested action items from the workshops focused on community building (10 workshops) and communication (6 workshops). Community building activities mentioned include increasing participation in EarthCube, providing online social networks and community collaboration tools, outreach events, codathons, workshops, developing data systems for communities that currently lack such systems, and promoting the emergence of small, targeted workgroups. Communication enhancing activities included a possible journal of geoscience data, tracking of NSF solicitations, listings of scientists interested in sharing data, and facilitation of social networking.

Education-related action items (6 workshops) include developing and testing learning activities for utilizing data and models and promoting research on how humans learn using data and models, and organizing training workshops and webinars on data management, documentation, and other data management and utilization topics.

Data related action items (5 workshops) include implementing linkages from publications to data and models, forming workgroups to develop standards and recommend best practices, compiling inventories of existing resources, and constructing infrastructure to support access to sensor networks and real-time data streams. Specific software processing capability action items were mentioned in 4 workshops, and include data analysis software seamlessly linked to data centers, tools for community annotation of resources, and leveraging existing library and visualization software. Other action items mentioned include implementing a governance mechanism to prioritize resource allocation, and developing a community database for 3-D Earth structure.

Previous Workshops: Best Practices And Lessons Learned

Representatives from five upcoming EarthCube End-User Workshops (Geochronology, Ocean Ecosystem Dynamics, Marine Seismic Data, Rock Deformation, and Polar Science Instrumentation) gave a short presentation on the goals and target communities of their workshops, and participated in a 'reverse panel,' in which they asked organizers of previous end-user workshops about best practices and lessons learned. Here is a summary of the observations from previous workshops.

1. **Invited Speakers:** Get motivated speakers with a clear, strong vision to open up the discussions, and have them stay throughout the workshop so that they can contribute to workshop dynamics.
2. **Communication:** Ask other Workshop PIs for copies of their proposals, and/or agenda materials to use as examples. Attend another end-user workshop, if possible.
3. **Engaging your respective community:** Workshop steering committee members should reach out personally to potential workshop participants. Personal contact is much more effective than e-mail to get people excited about the workshop and to commit to participating. Potential

personal outreach avenues include meetings, department hallways, and a personal phone call. If someone is unable to attend, ask if they can recommend a colleague.

4. **Workshop Registration:** Develop a Google Form for workshop registration that you can monitor to get a sense of the demographics of people who have registered. For example, if you ask participants for their primary and secondary scientific interest, you can monitor how well your current participant list is covering the necessary disciplines and target invitations to representatives of communities that are not well represented. An example signup form can be found here.
5. **Workshop Scheduling:** In order to maximize in-person participation, do not schedule your workshop for a Monday or for the day after a holiday.
6. **Inviting NSF Program Officers:** Invite multiple NSF Program Officers (POs) to attend your workshop**.** The POs will be there as observers and will send a signal to the domain scientists that many programs within the NSF are supporting EarthCube.
7. **Workshop Action Items:** Identify use cases and specific, actionable next steps, such as a RCN (Research Coordination Network) proposal, workshop, white paper, or other concrete action items.
8. **Discussion & Breakout sessions:**
    a. *Minimize the talks/presentations and devote more time for discussion.* A good example is the CZO (Critical Zone Observatories) workshop. With a big chunk of time – 1.5 hours – the CZO workshop organizers and participants came together on a proposal that was submitted only four weeks after the workshop and was just recommended for funding.
    b. *Highlight the breakout discussions.* For many workshops, the breakout discussions in small groups were the most productive aspect of the workshop*.* Make sure a scribe and a facilitator are identified for each break out to capture discussion and conclusions.
    c. *Debriefing:* Make sure you schedule enough time for the breakout sessions to debrief with the whole group, in order synthesize and integrate outcomes of each of the breakout sessions.
    d. *It is important to get people thinking big.* Allow people to air their immediate frustrations and workflows, but also encourage them to think bigger and look beyond their immediate issues. A good starting question is *what would you do if CI was not a barrier?*
9. **Virtual Participants:** A robust virtual participation component should be enabled. Components include good microphones, sound system and visual display, so that virtual participants can see the presentations and participate in group discussions and breakouts.
    a. *Sound Quality:* Place microphones around the room that connect to the virtual participants. If necessary, have microphone stewards who move mobile microphones around so that the speaker is always audible for virtual participants. In one of the workshops Polycom systems were purchased on Ebay and there was a projector in every room. Virtual participants won't be able to participate if they can't hear the discussion.
    b. *Breakout Sessions*: Integrate virtual participants into in-person breakout groups instead of assigning them to their own virtual breakout session.
    c. *Webex Capabilities*: EarthCube has a Webex account with three host licenses that can be used to provide a virtual participation component for the End-User Workshops. Up to the three meetings can occur simultaneously on the EarthCube Webex (one per host license). Each host license can accommodate up to 100 people. If you'd like to use the Webex account, please contact genevieve.pearthree@azgs.az.gov.

10. **Note-Taking**
    a. *Use Two Screens*: Have two screens and two projectors set-up. One screen will display the presentation material; the other will display notes taken in real-time.
    b. *Flip charts:* For breakout sessions, have participants take notes on a Google doc (for the record) and on a flip chart (for people in the room). Make sure you have one flip chart per breakout session.
    c. *Use Google Docs to take notes.* Google docs worked very well for many of the workshops. Take a look at some examples from other workshops: Critical Zone, Real-Time Data, Paleogeoscience, Inland Waters Geochemistry.
    d. *Training:* All workshop participants should have access to and know how to use Google docs prior to the workshop. You might need to train them prior to the workshop (i.e. a webinar or something similar).
    e. *Accessibility:* Make sure all the Google docs that will be used for note-taking are created prior to the workshop and that they are easy to find. On the agenda, share the link to the Google doc that corresponds to each activity so that it is easy to find. Make sure the Google docs are editable by all workshop participants so that many people can contribute and correct notes in real-time.
    f. *Note-taking Protocol:* Establish a note-taking protocol and identify a principal note-taker for each breakout session and group discussion. For example, in the CZO and Real-time data workshops, the note-taker's cursor was on top. Other people could add entries down below that the note-taker could then integrate into the main notes.
    g. *Designate two note-takers for each breakout session*: One person will take notes on a flip chart, the other can record the notes on Google docs.
    h. *Other Technology.* Tablet computers also work well for taking notes during breakout sessions.

11. **Special Events:** Arrange a tour, use live displays, or arrange another special event during the workshop to spark discussion and encourage attendance.

12. **Workshop synthesis & final report**: The leaders and scribes of the break out groups and the organizing team should get together for a half day after the workshop to synthesize workshop results before workshop organizers leave. This will prevent a long delay in finishing the workshop executive summary. In advance of the workshop, conveners should get commitments from other workshop organizers to stay late to synthesize workshop results.

13. **Workshop Survey:** Encourage workshop participants to take a survey on the workshop, noting what went well, what could be improved, and what next steps they'd like to take. For an example, please see the EarthCube End-User Principal Investigator Workshop Follow-Up Survey.

THE FUTURE OF EARTHCUBE

Potential Paths Forward

Workshop participants identified and discussed five potential paths forward for EarthCube. For each of these strategies, breakout groups analyzed the current state, desired/future state and delta state (action items and milestones needed to reach the desired state).

1. **Low Hanging Fruit:** *How do we leverage and expand upon existing communication and collaboration?* Focus on items for which there is already some convergence on requirements, ideally with existing software systems in place. The target communities are indicated by high percentages of researchers reporting interactions between them (Figure 1).

2. **Undervalued Opportunities:** *How do we close the big communication/collaboration gaps?* Focus on communities that do not currently report much interaction. Rapid progress may be possible by engaging previously separate groups (Figure 1).

3. **Institutional Stakeholders:** *How do the big players (data centers, academic centers, government agencies, etc.) work together?* Work with existing data centers that already have signification data holdings and experience, integrate those practices, and build from there.

Percent Reporting Daily, Weekly or Monthly Communications

	Atm.	Ocn.	Geol.	Geo phys.	Hydr.	Crit. Zone	Clim. Sci.	Bio. Eco.	Geog.	Comp. Cyber.	Data Mgr.	Soft. Eng.
Atmospheric/Space Weather Scientist n=151	98.7	34.6	14.9	23.3	33.0	7.8	63.1	23.7	18.7	52.3	60.4	54.8
Oceanographer n=114	47.3	98.4	52.8	36.3	30.6	18.2	59.4	78.6	12.9	37.9	50.5	33.7
Geologist n=273	23.1	34.3	96.9	74.5	44.7	34.2	42.5	37.1	26.1	35.7	30.7	20.0
Geophysicist n=112	16.1	35.1	80.8	96.8	28.8	27.0	32.4	22.8	11.2	51.2	49.2	44.4
Hydrologist n=72	41.6	22.9	61.1	42.7	93.9	63.1	59.2	69.3	37.8	52.5	52.6	37.9
Critical Zone Scientist n=28	28.6	25.9	89.6	60.0	76.7	100	64.2	71.5	50.0	36.6	48.2	28.6
Climate Scientist n=70	92.8	66.7	20.8	20.9	34.7	11.3	99.1	50.0	33.4	59.7	58.8	47.0
Biologist/Eco Systems Scientist n=67	28.4	50.6	49.3	26.8	51.4	33.2	52.1	100	43.7	47.9	62.9	28.3
Geographer n=29	51.6	37.8	48.5	32.3	48.4	33.3	55.1	72.4	93.6	51.6	62.0	41.4
Computer/Cyber Scientist n=60	40.0	40.3	29.7	21.8	50.0	38.2	43.2	51.7	32.8	95.3	82.3	91.7
Data Manager n=37	48.6	47.4	22.5	32.5	35.0	11.1	56.7	39.6	32.5	85.0	92.3	82.1
Software Engineer n=22	54.6	42.8	26.0	30.4	34.7	17.7	52.4	26.3	21.7	82.5	78.2	100

4. **Making Dark Data Light:** *Can the transformational goals of EarthCube be realized if we don't bring dark data to light?* Capture legacy (dark) data (in publications, not accessible in databases) and grey data (data in researchers' files, never edited and documented for sharing). Emphasis on 'long tail' ('mainstream') community.

5. **Next Generation of Geoscientists:** *Emphasize education and workforce training*. Train the next generation in the use of cyberinfrastructure; they will be more active users and are better suited to determine priorities.

Vision of Success

From these five potential paths forward, workshop participants elucidated a vision of EarthCube success. In this vision, EarthCube is a cyberinfrastructure for geoscience research that fosters new, transformational science using formerly unrelated resources by enabling simple discovery, evaluation and access to all data. EarthCube will be an easy-to-use system, similar to the World Wide Web, and provide useful capabilities to a broad spectrum of geoscience users. EarthCube will align scientific needs and technology development, building on existing cyberinfrastructure and geoinformatics while embracing open source culture and methodologies. The architecture required to meet science requirements is envisioned as emerging from convergence on conventions for services, interchange formats, and protocols. Geosciences communities without a mature cyberinfrastructure framework will develop a more mature framework. A successful cyberinfrastructure will enable broad participation in science, engaging the academic community as well as public citizen science.

Specific short and long-term components of an EarthCube vision of success are outlined below, organized according to their association with governance, science practice, software development, community building, and education.

1. **Governance: Co mmunity adoption of policy and specifications**
    a. Short-Term. Achieve consensus on requirements to increase the efficiency (cost/benefit ratio) and effectiveness of workflows for finding, sharing, and reusing data, tools and models. Use these requirements to establish infrastructure development priorities. Governance processes foster best practices for collecting, monitoring, and utilizing community feedback on the quality of content, reuse of resources, including practices and policies for access and authentication for this information, as well as privacy concerns
    b. Long-Term. A portfolio of community specifications is in place, under stewardship of EarthCube governance, and in use by the community.
2. **Science Practice: Changing the culture of how scientists work**
    a. Short-Term. Researchers frequently find, share, and use data, tools, and models from the EarthCube system in ways that would not have happened otherwise. Data, tools, and models are being 'mashed up' in new combinations to support new scientific discovery.
    b. Long-Term (Data). EarthCube supports everyday data discovery, access, reuse and stewardship with tools integrated into scientist's normal workflow. Scientists use and reuse data that is easily accessible through EarthCube repositories. Intelligent tools allow people to find, understand, and use data from disciplines outside their area of expertise. A culture of data sharing and availability based on EarthCube technology promotes sharing of published modeling tools, data, results, and big data.
    c. Long-Term (Scientific Discovery). EarthCube facilitates scientific discovery and cyberinfrastructure evolution for interdisciplinary science, particularly among Early Career scientists.
    d. Long-Term (Standards Process). EarthCube fosters standardized data collection procedures and descriptions across communities with requirements for formats and conventions. It is easy to create and publish useful metadata to enable data reuse.
3. **Software Development: New software capabilities and their adoption for daily use**
    a. Short-Term. Prototype systems are operational, based on currently functional cyberinfrastructure components, and provide access to essential variables and operations via web APIs (Application Programming Interfaces). These prototypes demonstrate integrated, interoperable geospatially searchable data with mapping capabilities for visualization, integrative tools for synthesis and analysis, such as a “digital crust” or “critical zone” prototype demonstration and tools for curating 'dark' (legacy) data, 'grey' data, and continuously created data.
    b. Long-Term (Entry point). An online, 3-D virtual globe provides an entry point to explore Earth Science data at different resolutions and from different perspectives.

c. Long-Term (Platform). This EarthCube platform provides community cloud cyberinfrastructure with APIs for data storage, resource cataloging and discovery, and high performance computing that Earth scientists use as a foundation for application-specific user interfaces and tools both within and across communities. Platform components support curating resources in ways that are immediately useful for others to reuse (understand and access), data management practices that maximize resource documentation and minimize the overhead for Earth scientists collecting data or generating new resources, and support integrating data with spatial and temporal scaling of essential variables for input to large, complex models or real-time use.

4. **Community: Development of an EarthCube community of practice**

   a. Short-Term. Alliances among institutional data centers implement pilot interoperability and integration experiments. Communication between geo scientists and cyber scientists exploits and promotes synergies between individual domains. Metrics are in place that demonstrate progress towards realizing the EarthCube vision.

   a. Long-Term. The Earth Science community identifies itself through participation in EarthCube, as data providers, data consumers, system developers, maintainers, and managers. EarthCube participants move from a set of individuals working on a global platform to a set of individuals who are the global platform.

5. **Education: Training and education to foster cyberinfrastructure use and adoption**

   a. Short-term. EarthCube provides intuitive, modular learning objects and self-directed lessons that are used by teachers from K-12 through the graduate level.

   b. Long-Term. An online, EarthCube 3-D virtual globe is the entry point for students to explore Earth Science data at different resolutions and from different perspectives. EarthCube helps everyone (including individuals outside of the geosciences) better understand how to use and interpret data, because EarthCube helps improve computational literacy at all levels.

EarthCube Fears

The following fears for the future of EarthCube emerged during discussion at the workshop.

1. **Adaptability**. Workshop participants are concerned that the data management backlog grows faster than our ability address cultural obstacles to data sharing.
2. **Funding**. Workshop participants are concerned that NSF funding won't materialize to help scientific domains mature their cyberinfrastructure framework, that NSF funding and support will not continue into the future, and that there is a lack of coordination between EarthCube Program Officers and other Program Officers within NSF.
3. **Governance and Prioritization**. Workshop participants are concerned that the EarthCube community will be unable to agree on priorities and demonstrate concrete progress. The community will have to establish relative importance of building innovative cyberinfrastructure (software, tools, etc.), bringing 'dark data' to light, and standards development, and use these priorities to guide resource allocation.
4. **Community Engagement**. Workshop participants are concerned that EarthCube will not engage a broad community across the geosciences.

5. **Data practices.** Workshop participants are concerned that data quality control will not be built into the system. Thus, processes must be developed and implemented to minimize data misinterpretation or misuse; access controls and respect for data ownership are necessary to assure that data are not shared too soon or too widely; adequate credit must be given for contributing data and models.
6. **Utility of EarthCube.** Workshop participants are concerned that EarthCube will not produce something useful to wide segments of the community. If we attempt to build cyberinfrastructure that meets the needs of all, we might end up with cyberinfrastructure that is ineffective at meeting anyone's specific needs. There must not be so many data and tool choices that data consumers are overwhelmed. EarthCube should enable global science, not solely focused on the U.S.A.
7. **Education and Workforce Training.** Workshop participants are concerned that education and training will be left out. Thus, EarthCube should foster training in data, modeling, and information science and technology to meet the needs of geoscientists.

MOVING FORWARD: ACTION ITEMS

A variety of near-term action items were established for workshop participants.

1. **EarthCube White Paper:** Participants will collaborate on an EarthCube white paper to serve as the voice of the end users at this key juncture for EarthCube (to be submitted to *Eos* or another appropriate forum, target date is October 2013).
2. **Workshop Meeting Report:** Participants will collaborate on a 500-word *Eos* Meeting Report, to be submitted to *Eos* as soon as possible.
3. **Data Facilities End-User Workshop:** An End-User Workshop for Data Facilities will be organized in January 2014 in conjunction with the Facilities Assembly Group as part of the upcoming EarthCube Test Enterprise Governance process.
4. **Low-Hanging Fruit:** Workshop organizers were tasked with identifying one or two 'low hanging fruit' projects for their scientific community, which will be compiled and presented as recommendations to NSF.
5. **NSF and Governance Webex Presentations:** A series of Webex presentations to the EarthCube community will be held in the fall of 2013. This first of these presentations will be an NSF announcement of funded EarthCube components, followed by an introduction of the EarthCube Test Governance Process and announcement of opportunities for continued participation in EarthCube, such as the EarthCube Test Enterprise Governance Assembly and other avenues.
6. **EarthCube Events at Professional Conferences:** Participants will collaborate with NSF to hold an EarthCube event, such as a Town Hall, at the Geological Society of America (GSA) 2013 Annual Meeting and the American Geophysical Union (AGU) 2013 Annual Meeting, in addition to organizing EarthCube events at other professional conferences in the geosciences.