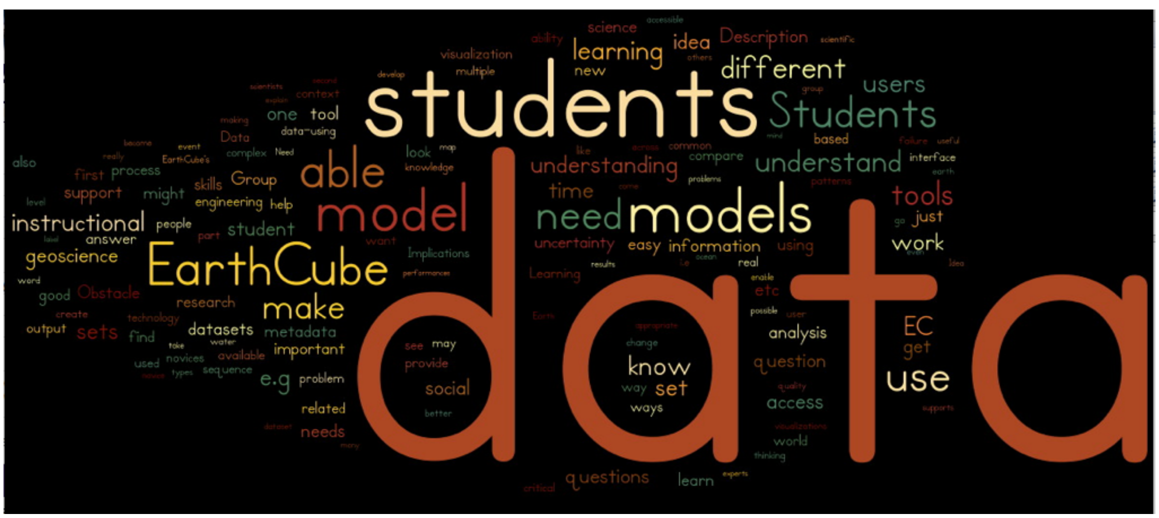
EarthCube Education End-User Workshop

Workshop: March 4–5, 2013
Scripps Institution of Oceanography, La Jolla, California



(image credit: Bill Finzer)

Convened and reported by
Kim Kastens & Ruth Krumhansl
Education Development Center, Inc.

and

Cheryl Peach
Scripps Institution of Oceanography

Report date: March 15, 2013; revised May 6, 2013

Summary

Purpose: Forty-six energized and motivated geoscientists, geoscience educators, data providers, employers, technologists, and curriculum developers met on March 4–5, 2013, at Scripps Institution of Oceanography to advise EarthCube's leaders and builders on the needs of end-users who will use EarthCube for education. The goals of the workshop were:

- to build EarthCube in such a way as to bring the power of learning through Earth data and models within reach of novices
- to use EarthCube to educate future geoscientists, who will be unprecedentedly adroit with data and models, and "native speakers" of interdisciplinary systems

Vision: Participants were motivated by a vision of a data-literate society, in which the entire populace makes better decisions in their personal and professional lives, decisions that are based on evidence—evidence grounded in data. Within the broad data-literate society would be a substantial and increasing fraction of "data-savvy" graduates, skilled at using data and models in answering difficult questions and solving knotty problems. We envisioned these graduates as "native systems thinkers," whose thinking ranges readily from discipline to discipline, from model to field to lab, across spatial and temporal scales, using digital cognitive tools as effortlessly as others ride a bicycle or drive a car, as seamlessly as a bilingual speaker switches languages.

Audience: The exemplar learners for our deliberations were undergraduate majors in earth and environmental sciences, because these are the pipelines for future geoscientists, the geoscience workforce, and K–12 earth science teachers. However, we believe these same recommendations will also make EarthCube useful to other important categories of motivated adult learners, including professionals who interact with the earth (e.g., environmental lawyers, architects, land-use planners), scientists from other disciplines collaborating with geoscientists on interdisciplinary problems, and even geoscientists working in parts of geoscience with which they are less familiar. Everyone becomes a novice in the other field when they begin to collaborate across disciplines. By making EarthCube powerful for undergraduates, we also make it powerful for interdisciplinary collaborators.

Challenges: Tools and interfaces optimized for use by scientists can pose a significant barrier to entry for both faculty and students, particularly for interdisciplinary topics. Although geoscience data and data-using instructional materials have been available on the Web for some years now, faculty need pedagogical content knowledge about how best to use them for teaching and learning, as well as a community of practice within which to share their ideas and seek answers to their conundrums. Many students come to college with a low level of experience working with data and lacking prior coursework in earth science or statistics. This may improve with the arrival of the *Next Generation Science Standards*; however, implementation of these ambitious standards presents its own daunting challenges. Making meaning from data involves difficult cognitive challenges, including dealing with uncertainty, spatial thinking, and metacognition. Cognitive-science and learning-science research on how people learn from data and models is fairly sparse, and the research findings that do exist have not yet been widely incorporated into instructional design.

Data sources in current use: The undergraduate geoscience education community has been proactive in making use of a very wide variety of geoscience data types, either by directing students to Web-accessible data sources or by extracting selected data and providing it in digital form as part of a data-using student investigation. To see the range and depth of data in current use in geoscience education, please browse the following collections:

- *Using Data in the Classroom: Data Sources and Tools:* http://serc.carleton.edu/usingdata/resources.html

- *Earth Exploration Toolbook Chapters:* http://serc.carleton.edu/eet/chapters.html

Cyberinfrastructure desired: With respect to EarthCube's cyberinfrastructure, the workshop participants prioritized data types that are germane to societal problems, data types (including eyewitness narratives and other qualitative data) that give students a "feel" for earth structures and processes, both real-time and historical data, student-collected data, and citizen-scientists' data. Especially important was the ability to input student-collected data in such a way that it could be analyzed using EarthCube's data analysis and visualization tools in the context of broader datasets, without diluting the quality of the research data collection. For user tools, the group prioritized collaboration tools, the ability to search effectively without prior knowledge of what group or instrument collected the data or made the model, and the ability to begin with a simple presentation and then gradually dig deeper to reveal complexity. In general, participants did not favor a dedicated education portal for undergraduate users; they would rather direct learners towards a slimmed-down novice view of the full-fledged EarthCube, in which less-commonly-used options and tools would be hidden until the learner developed a need for them. As for how data are displayed and analyzed, participants wanted EarthCube to "make "failure" really cheap," in other words, to make it easy to try things out, explore, and experiment with ideas and hypotheses. It will be important for students to be able to combine historical data and predictions, but equally important to draw a clear distinction between model output and measured data, perhaps telegraphing the difference graphically.

Social/community infrastructure desired: The education workshop considered the social engineering challenges in achieving the transformative impacts envisioned for EarthCube to be at least as critical as the technological problems. There are already geoscience data portals that are reasonably accessible to students, and some good curriculum materials based on these data; yet these have not come close to transforming geoscience education, let alone society. To achieve a data-literate society and a robust pipeline of data-savvy scientists, EarthCube will need to foster a community, or perhaps an ecosystem, in which educators, data providers, curriculum developers, assessment developers, credential providers, prospective employers, technologists, education researchers, scientists, and others find it mutually beneficial to work together towards spreading data-savviness. EarthCube should invest in understanding its users, what they know and how they think and learn with data and models. EarthCube should carry out user testing early and often and with diverse audiences, including professionals from outside geosciences, urban and low SES learners, and learners with disabilities. It's been said that you can't improve what you can't measure, and EarthCube-using educators will be looking for reliable, validated assessments of student proficiency with data and models—assessments that do not currently exist. EarthCube should engage with the learning-science and cognitive-science communities to encourage and then apply research into how humans think and learn with scientific data and models.

In summary, the workshop group was cautiously optimistic that EarthCube will contribute to a substantial step forward in students' ability to think and learn. By using EarthCube in appropriate learning environments, students will increase their understanding of both the Earth and of data, emerging as self-directed learners, who can find their way to answers to Earth questions and solutions to Earth problems. To ensure that EarthCube will evolve in a direction that serves novices of all ages and flavors, the voices of education end-users should continue to be part of the planning and testing process for EarthCube at all steps along the way. In the rest of the report, we elaborate on the education "driver" or vision for EarthCube empowered learners, the challenges to achieving such learning, and put forward specific recommendations for EarthCube's cyberinfrastructure and social/community infrastructure.

Drivers for EarthCube-enabled Improvement of Education

Here we describe the workshop participants' vision of a data-savvy college graduate, which we defined as:

> *Learners who have mastered the skill sets and habits of mind to use data and models to address novel and ill-structured problems, in collaboration with others, using appropriate data and analysis strategies, and who can effectively communicate their claims/evidence/ reasoning to others.*

Asked what data-savvy learners should know, understand, and be able to do with data and models, workshop participants came up with broadly similar criteria for what would constitute mastery of data-using skills. The skill set spanned the range from understanding the nature and characteristics of high-quality data to the ability to independently seek out and use data to address novel problems/questions and create new knowledge. Below is a synthesis of the participants' views of the range of skills, capabilities, and habits of mind that would constitute mastery. The ideas are divided into two categories: what a data-savvy learner should *know and understand* about data and models, and what a data-savvy learner should *be able to do* with data and models.

What data-savvy learners should know and understand about data and models

First and foremost, participants pointed out that data-savvy learners understand that data is an important category of evidence for answering questions and input for solving problems. One group of participants put it this way—*When confronted with a question (or an issue), we want students to have the habit of looking for data to answer the question or solve the problem.* Most groups pointed out that an important step to mastery was being able to recognize what *kinds* of questions can be answered with what *kinds* of data, and to formulate questions that can be solved with data.

Data-savvy learners were seen as critical consumers of data and models. They are able to evaluate data quality, understand "uncertainty" in data and how to report it, and can identify gaps in a dataset. Data-savvy learners are familiar with a wide variety of different types of data, understand the difference between data and model output, know how those data are collected (instrumentation) and processed (software), and are familiar with how final data products are generated. They are familiar with concepts of data taxonomy and provenance, and understand the importance of metadata in archiving and sharing data.

Data-savvy learners are able to accurately interpret and extract meaning from graphical and spatial representations of data, and understand linkages among related data (e.g., temperature and density). They can make connections from the abstract (e.g., false color) to what it means in the real world (e.g., symbology). They also have a clear understanding of the concept of scale, both spatial and temporal, and have developed the ability to place data in both spatial and temporal contexts.

They understand that models are useful but simplified representations of aspects or portions of the Earth system, and are conversant with a variety of expressed model types, including physical models, mathematical models, diagrammatic models, and computational models, as well as mental or conceptual models. They know that all models have limitations. They realize that models have several uses, including demonstrating, explaining, and predicting; but above all, they understand that scientists' models are hypotheses to be tested by comparison against data, and that modification, amplification, or even total rejection of a model is an inherent part of how science progresses.

What data-savvy learners should be able to do with data and models

Data-savvy learners are facile with determining the type of data they need, and either finding appropriate pre-existing datasets or collecting their own data. When collecting their own data, they are aware of and use best practices for gathering, recording, and sharing that data. If accessing pre-existing data, they are able to select appropriate data, assess its quality, and site the source of the data appropriately.

Being able to develop strategies for collecting/accessing, analyzing, representing, visualizing, and interpreting data to answer a question is a key attribute of a learner who has mastered data-using skills. These learners are able to enter and manipulate data easily using a variety of available technology, select and apply appropriate data analysis tools, and understand and use data manipulation software applications. They are also able to generate data-driven graphs and/or visualizations, and articulate clear reasons for choosing to illustrate particular relationships in the data.

Mastery also includes the ability to establish linkages between disparate types of data, often from different disciplines, and to combine data from different sources. Moreover, mastery of data-using skills includes the ability to collaborate on data-intensive projects across disciplines and engage in multidisciplinary teams tasked with solving complex problems that require diverse expertise. This includes being able to negotiate language and other barriers between cultures and disciplines.

Data-savvy learners are aware that data are often complex, especially when reflecting processes in a natural system, and that there are limits to what can be explained using data as a result of uncertainties inherent in any dataset. They should be able to apply their mathematical understanding to interpreting the data, and to develop a new mathematical schema to describe the data. Mastery in using data includes being able to effectively communicate findings to others, contextualizing results in ways that make them interesting, and developing a scientific story that anchors their findings in prior knowledge. Finally, data-savvy means being able to document their work and the methods used to obtain their results.

Current Challenges in Teaching and Learning with Data

The workshop addressed challenges and obstacles that stand in the way of achieving the envisioned data-savvy graduate. Challenges emerged concerning data tools and access, students' preparation, instructors' pedagogical content knowledge, instructional materials, and the nature of the cognition needed to make meaning from data.

Expert interfaces pose a significant barrier to student use.

Data-access portals and analysis tools designed for professional geoscientists typically include specialized terminology and data structures, and assume knowledge that only experts within a constrained discipline area possess. Because of this, the process of finding, downloading, and visualizing/analyzing data requires a significant investment of time on the parts of the instructor and students, and only a small proportion of this time is spent extracting meaning from the data. Beyond the difficulties non-experts typically encounter finding and downloading appropriate datasets, there is a lack of learner-friendly data analysis tools and understandable data visualizations—and guidance about which ones to use when—which hampers learners' ability to think about and learn from the data.

For data portals to be usable by educators, they need to support diverse users and continue to support them as they gain sophistication in their use of EarthCube data. The data presentation needs to be flexible and customizable to students with varied interests and aptitudes. They need to grow with the learner as he or she gains experience and wants to pursue more sophisticated investigations with more powerful tools. Participants discussed the fact that there is likely an inverse relationship between the usability of tools and the scale of what they can do. This means that the tools optimized for novices won't be as useful to learners as they gain proficiency.

Barriers to working across disciplines are compounded for novice users.

Because of the interdisciplinary nature of problems, such as climate change, that are of current relevance and interest to students, it is important for instructors and students to have access to datasets from multiple sources. However, because data from multiple sources typically embody different formats, timescales, spatial resolution, and measurement techniques, there are significant barriers to interdisciplinary data investigations. Tools designed to work with a typical type of data format don't permit use with others, and data search protocols don't work with all data types. The redundant time and cognitive energy that students invest in learning to work with multiple types is not available for making meaning from the data.

Students enter college with very uneven knowledge and skills around data, models, and the Earth.

There was the general sense that many undergraduate students come to college with little experience in working with data beyond mechanical making of graphs and calculation of simple statistics. Their transferrable math skills may be weak, and few have studied statistics in pre-college classes. They have limited experience working with data beyond small, student-collected datasets or artificial "data" invented for student exercises. Many students don't know how to find appropriate data to address a particular question or how to select and apply appropriate data analysis tools. They have seldom been asked to develop inferences about the represented system from data. Perhaps because of their general lack of experience doing scientific work with data, they have misconceptions about data, problem-solving, and scientific practices that significantly hinder their ability to use professionally collected datasets. Weak pre-college preparation in data-using skills is particularly, but not exclusively, a problem in underserved communities and populations underrepresented in science.

Many students also come to college with limited knowledge of the earth, and this inhibits their ability to make meaning from complex earth systems datasets. Most did not complete coursework in the Earth or Environmental Sciences in high school, and many did not have the informal outdoor experiences immersed in nature that an older generation of geologists takes for granted. Their lack of content knowledge and unaddressed misconceptions can make it difficult for them to see connections between a specific data visualization and earth processes.

Workshop participants felt that this lack of prior knowledge on the part of students made it essential to fill knowledge gaps. However, most K–12 educators lack experience in either learning or teaching with large professionally collected datasets, and don't know how to guide appropriate inquiry with data. If and when the *Next Generation Science Standards* (Achieve, 2013; National Research Council [NRC], 2011) are broadly implemented in K–12 classrooms, this should significantly improve students' pre-college preparation around data, models, and the Earth. However, it will take considerable effort and time to fully integrate the standards into schools, and in the interim, it will be necessary for undergraduate educators to fill the gaps.

Having painted a somewhat discouraging picture of the weak background of many students, we should also acknowledge that there are other subsets of undergraduates who have phenomenally strong data skills, computational skills, and/or understanding of Earth Systems. They may have participated in an authentic science research program in high school or in a summer program, or they may been part of a computer programming subculture where they brought themselves to a high level of proficiency in computing. They may be coming straight from a well-taught AP Environmental Science course with a substantial field research component. Another subset of students arrives at college deeply concerned about problems of environmental sustainability, and hungry to learn more about the Earth System to undergird their environmental actions. For self-motivated learners, the internet provides a level of access to knowledge that previous generations simply could not find in the typical school or town library, and some students have used this access to build deep knowledge of particular aspects of Earth Systems.

The bottom line is that college geoscience faculty face a very wide range of backgrounds in their students, sometimes even in the same class. EarthCube has the potential to help instructors differentiate their instruction so that each student moves forward substantially from his or her starting point, whatever that starting point may have been.

Appropriate instructional materials need to be developed and disseminated.

Participants cited a lack of knowledge of instructional materials that are currently available, a lack of knowledge about how to create appropriate instructional materials, and (given these difficulties) a lack of time as significant barriers to data use in the undergraduate classroom. Beyond that, they felt the need for more information about the context (e.g., scientific research purposes and discipline) in which the data were collected, the methods/ technologies used to collect the data, and data limitations. The following knowledge gaps among instructors were identified by participants:

- What data are available to use in classrooms, and what is the effective use of these data?
- What is the prior knowledge of students in their classes about data and models, and how can they assess this prior knowledge?
- What type of instruction will effectively build on different levels or patterns of prior knowledge?
- How do you adapt materials to different learners?
- What does research on learning say about the appropriate design of instructional materials to build students' data-using and modeling skills?
- Are there data ethics issues that might pose a barrier (e.g., when is it ok/not ok to use someone else's data in a lesson with students)?

In addition to the barriers listed above, faculty may not have (or feel they have) the expertise to cross disciplinary boundaries and use data from other fields. Participants cited the need for professional development for faculty, as well as the means (and motivation) to share data-using lessons and experiences. The type of knowledge most difficult for faculty to find is "pedagogical content knowledge": knowledge of how to teach a body of content as opposed to knowledge of the content itself.

Relevant cognitive/learning science is sparse and insufficiently incorporated into instructional design.

To maximize the effectiveness of EarthCube, more basic and applied knowledge is needed about the cognitive processes that humans use in making meaning from data and models, and the cognitive challenges that learners face as they try to employ these modes of learning. Examples of areas where the knowledge base is insufficient and/or not yet incorporated into curriculum materials and teaching practice include:

- Novice *misconceptions* about data and models, and how they relate to the nature of science. For example, non-scientists often view science as definitive (i.e., about what is known), whereas scientists view it as a process for understanding the unknown. In the former view, a scientific model is the correct answer; in the latter view, today's scientific model is a provisional stepping stone to a better future model.
- Students' difficulty dealing with *uncertainty, probabilities, and prediction*. Studies show that in statistics education, students have trouble understanding uncertainty even in one dimension. In the geosciences, they are asked to work with four dimensions as well as variations in the type of measurements.
- *Spatial visualization*. Work with geoscience data requires strong skills in spatial visualization, and learners vary significantly in the spatial abilities they bring to a learning task. We need to better understand the specific cognitive barriers faced by students as they approach a variety of data visualization tasks, and how to help them develop the skills to be successful.
- *Metacognition,* or thinking about one's thinking. The metacognition that supports formulating effective questions, solving problems, and drawing inferences is poorly understood,

Recommendations

Participants emphasized repeatedly that an effective EarthCube will need both effective cyberinfrastructure and a suite of effective social and community structures.

Recommendations for EarthCube technology/cyberinfrastructure

About types of data to be made available:

- *Data germane to humanity's pressing problems:* Students are motivated by seeing the relevance of their educational activities to issues in their own lives. This may be especially true of low SES urban students who have limited exposure to nature. Thus, participants prioritized data types that are relevant to environmental sustainability, resource limitations, and geological hazards.
- *Field data:* EarthCube needs a way to input, archive, and serve relatively "low-tech" field data types, such as beach profiles or dip & strike measurements. In this, education workshop participants echoed the recommendations of the field-oriented Structure & Tectonics Workshop. From an education perspective, these data types are important because they give students embodied experience with data acquisition directly from the Earth or environment.
- *Student-collected data:* EarthCube needs a way to input student-collected data in such a way that students can use EarthCube's visualization and analysis tools to work with their own data. In addition, students should be able to compare, contrast, and combine the archival data and their own student-collected data, for example, by seeing them displayed on the same graph, map, or other visualization.
- *Tiered approach to data quality:* Students who collect excellent data, with rigorous, well-documented protocols and stringent quality control, should have a way to get their data into EarthCube's main data archive. On the other hand, much student-collected data will be not be of research quality, and will need to be partitioned off in such a way that it can be used by those students but will not be seen by other EarthCube users.
- *Near-real-time data:* For fast-breaking earth events, such as earthquakes or storms, educators and motivated adult learners will want relevant data quickly. Also beneficial would be quick posting of contextual information about the event, along the lines of IRIS's Recent Earthquake Teachable Moments (http://www.iris.edu/hq/retm).
- *Historical archival data:* Participants want students to be able to look at long- and short-term trends and variability in earth phenomena across both the scale of human history and earth history.
- *Local informants' eyewitness accounts:* Participants recommended that EarthCube archive local informants' earth observations as a data type, for example, local informants' accounts of changes in an island's coastline over time or descriptions of an earthquake. Such accounts may help to fill in the historical record of earth's surficial processes, and, from an educational point of view, they provide students and curriculum designers with vivid, memorable narratives of dynamic earth processes, which can be both motivating and can help build students' "geoscience intuition" or "feel for the earth."

About user tools:

- *Interface design:* to make EarthCube accessible to users beyond domain experts, it is critical to follow good, basic interface design rules (Krumhansl et al, 2013; Johnson, 2010; Schneiderman & Plaisant, 2010). Appropriate data should be easy to identify and download; the tools provided to visualize and analyze data should be as intuitive as possible; data visualizations should be clearly presented and labeled, and readily customizable to the data and task.
- *Search:* Workshop participants want their students to be able to search for any type of earth data using common English vocabulary without knowing ahead of time the name of the instrument or the organization that collects and serves that data type. In cataloging datasets, metadata should include educationally-relevant information (for further detail, see Ledley, Prakash, Manduca, & Fox, 2008).

- *Entry-point via derived data products:* Raw data is often incomprehensible to novices. EarthCube should provide a level of derived data products that are relatively comprehensible to novices. But then "behind" these derived products, there should be the option for a motivated learner to dig deeper to learn how the data were acquired and processed, and to access more-nearly-raw data.
- *Slimmed down novice view:* In general, participants did not favor a dedicated education portal for undergraduate users, feeling that working in such an environment does not build adult competencies and that such secondary portals tend to be less well maintained than those used by professionals. Rather, most of the participants would favor a novice view of EarthCube's interface, in which less-commonly-used options and tools would be hidden until the learner developed a need for them.
- *Data exploration:* Participants want EarthCube to "make 'failure' really cheap," a "safe place to learn and experiment." By that they mean that multiple datasets can be easily manipulated and hypotheses can be explored; if one line of inquiry doesn't pan out, then the learner can move on to the next without having sunk vast effort into the first. This desire would suggest a "quick tool" for data exploration rather than (or in addition to) a more complex tool. The goal is for learners to spend a larger proportion of their time drawing meaning from data and exploring more data, and a smaller portion of their time on the data formatting and manipulation processes that absorbed so much time in the past.
- *Collaboration tools:* Participants envision small groups of students developing an expertise around one facet of a complex problem, and then working collaboratively to combine their newly acquired knowledge and skills to tackle the problem. We think that EarthCube should have digital tools to help interdisciplinary teams of scientists work collaboratively, and those tools should be simple enough for student use and made available for qualified student use (i.e., enough capacity that they don't get overscheduled).

About data display and analysis capabilities:

- *Spatial and temporal data displays:* EarthCube should have robust capabilities for analyzing and visualizing spatial and temporal data.
- *Adaptive data display options*. The most illuminating analyses and display options vary with the nature of the data and task. Experts can choose wisely from an exhaustive array of options, and can make sense even out of a display that uses sub-optimal default display parameters. However, for novices there is a delicate balance between giving them enough options that they gain expertise in design of data displays and yet they end up with displays that can support meaning-making. How to find this balance, across a range of data types and inquiry tasks, will require educators' intuition and experience, followed by usability testing.
- *Combining data and predictions:* Participants want learners to be able to combine and compare historical/archival data about a phenomenon with forecasts/predictions about the future. Tools could include data visualizations that put both on the same display, and analytical tools to quantify the similarities and differences.
- *Clear distinction between measured/empirical data and model output:* Participants want EarthCube to support students in learning to distinguish between empirical/measured data and model output. Perhaps EarthCube could develop or adopt an ontology of data/output types (e.g., raw measured values, data visualization interpolated between measured values, output of computational model) and use icons, color frames, or some other distinctive graphic signals to telegraph which of these types users are looking at.
- *Comprehensible metadata:* As other user groups will surely be recommending, the education workshop participants recommended that all data should be accompanied by metadata that documents how, when, where it was collected, etc. Education-user participants further requested that EarthCube have the capability to provide metadata in language accessible to novices (Ledley, et al, 2008), accessible through an intuitive pathway such as a "simple explanation" button, and with links to

explanations of tools and techniques used to acquire and process the data.

- *Supports for understanding uncertainty:* Workshop participants emphasized the importance and difficulty of helping students understand uncertainty about both empirical data and model output. EarthCube can help by providing clear and well-documented indicators of uncertainty (e.g., for maps, which often lack this information).

About models:

- *Student ability to build and contribute to building models:* Anticipating that EarthCube will support the aggregation and integration of models as well as data, workshop participants desire a robust but fair and transparent vetting procedure for user contributions to the community models, such that excellent contributions by students and novices are welcomed, and flawed contributions are sent back with useful feedback for further work.
- *Simpler versions:* Participants recommended providing simpler versions of models to help students "get their feet wet." From the simpler versions, there should be a well-documented growth path towards professional versions.
- *Clarity around epistemology of model-using science:* Most students and many instructors think that models are just a means of explaining or demonstrating that which one already knows. But for 21st century scientists, computational and physical models are a tool to create and test new claims. How scientists use models, combined with data, to create new knowledge is opaque to most of the public. Instructors need teaching materials built around simple models to break the idea that the model is "the truth," and establish the idea that a model is a hypothesis to be tested and improved, as well as an ontology for talking about different kinds of models.
- *Comprehensible metadata:* Metadata is as essential for models as it is for data (see above). Students should have easy access to what assumptions went into the model's construction, what parameters are selectable and what they influence, and similar information to help them become informed and critical users of models and model output.
- *A "file cabinet" of introductions to models:* Workshop participants thought the development of a "file cabinet" (i.e., a clearinghouse) of introductions to models—each similarly documented as to its evolution, purpose, limitations, etc.—would benefit novices.

Capabilities for teaching and learning:

- *Citizen science:* Participants envisioned that for certain data types that require intensive human-mediated interpretation (e.g., classifying organisms in seafloor videos from the ocean observatory), motivated learners should be able to participate in data interpretation on behalf of authentic research projects. A model is Galaxy Zoo (http://www.galaxyzoo.org) in which members of the public classify galaxies. This requires a technical infrastructure for partitioning and assigning tasks, and training and certifying participants, as well as an educational thread so that students can move beyond just coding data to interpreting data.
- *Mentoring:* Participants envisioned widespread, just-in-time virtual mentoring supported by artificial intelligence, plus high-level human mentoring. Mentoring is needed for instructors who are new to EarthCube, and then the instructors themselves need tools and techniques for mentoring students who are new to data and models.
- *Assessment techniques:* Instructors need ways to assess students' level of mastery of the practices of "analyzing and interpreting data" and "developing and using models." Part of the answer could be technological: EarthCube could provide online assessments of certain commonly used data-using skills and practices, and could provide certificates of mastery.
- *Tutorials:* The workshop group was split on the value of tutorials on how to manipulate EarthCube. Some thought that EarthCube should be so intuitively obvious that tutorials would be superfluous;

others thought that such simplicity was unachievable; others thought that regardless of how intuitive the tool, some learners' preference is to learn from tutorials and so EarthCube should provide them. However, the group definitely saw a value in some kind of online learning tools—maybe along the line of Kahn academy videos (http://www.khanacademy.org)—for data-interpreting skills such as recognizing and interpreting patterns on maps and graphs.

- *Venues in which to share:* This recommendation reappears below under "social structures," as participants want both virtual and face-to-face venues in which to share ideas, answers, etc. One model for virtual sharing of answers and solutions is StackExchange (http://stackexchange.com), and one for collaborating on projects is GitHub (github.com).

Recommendations for EarthCube community and social structures

- *Interpersonal connections:* Workshop participants hope that EarthCube will provide means to build links and bridges between people: between data-using instructors and students and scientists who are familiar with that data type; between faculty developing EarthCube-based teaching materials with other faculty with similar needs; or between experienced data users with novice data users.
- *Support for diverse populations:* Participants urge EarthCube to resist becoming an elitist organization, to understand the starting point of urban youth and low SES (socioeconomic status) populations with respect to both data and Earth, and then develop educational approaches that build from there.
- *Support for learners with disabilities:* Traditional geoscience education, with its emphasis on fieldwork, presents particular difficulties for students with mobility, vision, and other physical disabilities. Learning with data and models is a form of learning in which such students can participate more fully if their needs are taken into account in the design process.
- *Venues in which to share:* Participants desire venues and mechanisms through which to share lesson plans, pedagogical content knowledge (knowledge of how to teach a body of content as opposed to knowledge of the content itself), answers to problems, and ideas about data and models (e.g., how to cope with model runs that differ over time because the model has been changed). Participants see a role for both virtual venues (see above) and face-to-face professional development and workshop style events (along the lines of the Cutting Edge professional development workshops: http://serc.carleton.edu/NAGTWorkshops/index.html).
- *Assessment techniques:* Instructors need ways to assess students' level of mastery of the practices of "analyzing and interpreting data" and "developing and using models." Assessments of scientific practices are now relatively few and primitive. EarthCube, in collaboration with NSF/EHR, could support design research around how to assess students' abilities with data and models. Reliable, validated assessments will make it more plausible for colleges and universities to insist on a certain threshold level of mastery for either all students or for prospective science majors.
- *Training:* The workshop group was split on the value of training sessions and short courses for education users of EarthCube. Some felt that to be successful, EarthCube must be sufficiently intuitive: self-guided learning would be possible and formal user-training would not be necessary. Others felt that well-designed training sessions would be invaluable for bringing time-crunched instructors up to speed. On balance, probably both are needed for EarthCube to achieve widespread use among undergraduate faculty: an intuitive interface AND training opportunities for those who want them. The most effective training would include not just how to use the EarthCube toolkit, but also pedagogical content knowledge around teaching with data and models.
- *Support for entrepreneurial enterprises:* A segment of the populace who are not motivated by answering scientific questions or solving abstract problems may be motivated by making money. EarthCube should find ways to support learners who are dreaming up new ways to use data and models to advance business and industry.

- *Engage with the K–12 science education community:* The new K–12 science standards (Achieve, 2013; NRC, 2011) foreground the importance of science and engineering practices, including Develop & Use Models and Analyze & Interpret Models. However, most K–12 teachers have little experience teaching or learning with data or models. A higher percentage of undergraduates could reach data-savviness by graduation if they entered college with stronger backgrounds.
- *Engage with community college educators:* Many geoscience majors begin their course of study at community colleges. Many community college faculty have strong backgrounds in pedagogy and strong interest in curriculum development, which could be leveraged to help build out EarthCube's education offerings.
- *Engage with publishers:* Publishers have expertise in reaching large audiences, which most academics lack. They may be open to new ideas and new partnerships, as digital materials threaten their traditional markets.
- *Engage the learning and cognitive science research community:* Some very basic processes about how humans create understandings from data and models remain poorly understood (e.g., How do humans make inferences from observations? Under what circumstances does exposure to new data cause a person to change his/her mind?) EarthCube may be able to work with the NSF's EHR and SBE directorates to foster research on basic and applied research relevant to teaching and learning with data and models.

Acknowledgments

This work was supported by the National Science Foundation through grants GEO13-13866 and GEO13-13870 and through the encouragement and insights of program officers Barbara Ransom and Jill Karsten. The National Association of Geoscience Teachers was a co-sponsor. The workshop's cyberinfrastructure was provided by the Science Education Resource Center (SERC), with outstanding support by John McDaris and Sean Fox. Delia Constant of Scripps Institution and Silvia LaVita of EDC provided excellent administrative, logistical, and technical support. The conveners especially thank the workshop participants, who worked incredibly hard and gave generously of their insights and experiences.

Sources and references

This report draws on notes recorded in real time by participants at the EarthCube Education End-User Workshop, held at Scripps Institution of Oceanography, March 4–5, 2013, as well as on several previously available reports, workshops, and websites, plus the expertise and insights of the authors and conveners of those previous reports and workshops.

Prior work includes a National Science Digital Library workshop on *Using Data in the Undergraduate Classroom* (Manduca & Mogk, 2002), a series of workshops of the *DLESE Data Access Working Group* (Ledley et al., 2008; see also http://serc.carleton.edu/usingdata/dawg/index.html) and their follow-on Data Access workshops (Taber, Ledley, Lynds, Domenico, & Dahlman, 2013): a 2002 Cutting Edge workshop on *Using Global Datasets in Teaching Earth Sciences* (http://serc.carleton.edu/NAGTWorkshops/ globaldata02/index.html), a 2003 GSA discussion on *Using Data to Teach Earth Processes* (http://serc.carleton.edu/NAGTWorkshops/ gsa03/index.html), a 2008 Cutting Edge workshop on *Teaching with New Geoscience Tools: Visualizations, Models and Online Data* (http://serc.carleton.edu/NAGTWorkshops/tools08/index.html)., and the Oceans of Data DRK-12 project (Krumhansl, et al, 2013). Websites consulted include the NSDL Using Data in the Classroom site (http://serc.carleton.edu/ usingdata/index.html), the Earth Exploration Toolbook site (http://serc.carleton.edu/eet/index.html), and the Cutting Edge site on Teaching with Data, Simulations & Models (http://serc.carleton.edu/ NAGTWorkshops/ data_models/index.html).

Achieve. (2013, January). *Next generation science standards*.

Johnson, J. (2010). *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules.* Burlington, MA: Morgan Kaufmann

Krumhansl, RA, C Peach, J Foster, A Busey, I Baker, J Delisi (2013) *Visualizing Oceans of Data: Educational Interface Design*

Ledley, T. S., Prakash, A., Manduca, C., & Fox, S. (2008). Recommendations for making geoscience data accessible and usable in education. *Eos, 89*(32) . doi: 10.1029/2008EO2003.

Manduca, C., & Mogk, D. W. (2002). Using data in undergraduate science classrooms. Carleton College: National Science Digital Library. Retrieved from http://serc.carleton.edu/files/usingdata/UsingData.pdf.

National Research Council. (2011). *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Shneiderman, B. & Plaisant, C (2010). Designing the User Interface: Strategies for Effective Human-Computer Interaction (5^{th} ed.). Reading, MA: Addison-Wesley.

Taber, M. R., Ledley, T., Lynds, S., Domenico, B., & Dahlman, L. (2012). Geoscience data for educational use: Recommendations from scientific/technical and educational communities. *Journal of Geoscience Education, 60*(3), 249–256.

Appendix A: AGENDA

March 4–5, 2013

UCSD Scripps Institution of Oceanography | Robert Paine Scripps Forum | La Jolla, CA

MONDAY, MARCH 4

8:30 am	**Breakfast**
9:00	**Welcome & Introductions** • Welcome on behalf of SIO and NAGT • Introduction to EarthCube: Barbara Ransom, NSF
10:00 – 11:30	**Learning Goals & Learning Performances** *Learning goals:* What should an undergraduate geoscience major know, understand, and be able to do by graduation? *Learning performances:* How will we know a data-savvy graduate when we see one?
11:45 am – 12:45 pm	**Employers' Panel** What knowledge/skills/habits of mind do you want to see in your data-using employees?
12:45 – 1:30	**Lunch**
1:30 – 2:00	**Survey results from Joel Cutcher-Gershenfeld**
2:00 – 3:30	**Obstacles and Problems** What is getting in the way of being able to achieve such data-savvy graduates?
3:45 – 5:30	**Instructional Sequences** What will the data-using lesson of the future look like? And what are the technology/social engineering implications for each of our envisioned instructional sequences?
6:30	**Dinner, Hotel La Jolla** *Speaker:* Professor Edwin Hutchins, UCSD Cognitive Sciences

TUESDAY, MARCH 5

8:00 – 8:30 am	**Breakfast**
8:30 – 10:30	**Interface design** "Oceans of Data" recommendations for student-friendly data access How we would like to improve on existing data access interfaces
10:30 – 11:00	**Break**
11:00 am – 12:30 pm	**Geosciences models session** How will students compare empirical/observational data with model output? How could EarthCube support students' learning to become critical users of model? How could EarthCube support students' learning to be model-builders?
12:30 – 1:15	**Lunch**
1:15 – 3:00	**Blue skying the future of EarthCube-enabled education** Imagine trying to create a new kind of undergraduate, someone who is a "native systems thinker," someone whose thinking ranges readily from discipline to discipline, from model to field to lab, across spatial and temporal scales, using digital cognitive tools as effortlessly as I ride a bicycle or drive a car. Could we do this? And if so, how?

Appendix B: Participant List

March 4–5, 2013

UCSD Scripps Institution of Oceanography | Robert Paine Scripps Forum | La Jolla, CA

WORKSHOP CONVENERS

Kim Kastens (Education Development Center)
Ruth Krumhansl (Education Development Center)
Cheryl Peach (UCSD Scripps Institution of Oceanography)

WORKSHOP PARTICIPANTS

Lisa Adams (Kennesaw State University)

Leilani Arthurs (University of Nebraska-Lincoln)

Chris Atchison (Georgia State University)

Christine Bagwell (University of California, San Diego)

Tom Baker (ESRI)

Chaitanya Baru (University of California, San Diego)

Reginald Blake (New York City College of Technology, of the City University of New York)

Karin Block (City College of New York)

James Brey (American Meteorological Society)

Donna Charlevoix (UNAVCO)

Joel Cutcher-Gershenfeld (University of Illinois)

Annette DeCharon (University of Maine)

Steve Diggs (University of California, San Diego)

Bob Downs (Columbia University)

Bill Finzer (KCP Technologies)

Sean Fox (Science Education Resource Center, Carleton College)

Scott Glenn (Rutgers University)

Andrew Goodwillie (Lamont-Doherty Earth Observatory)

Margaret Hedstrom (University of Michigan)

Arjun Heimsath (Arizona State University)

Michael Hubenthal (Incorporated Research Institutions for Seismology)

Ed Hutchins (University of California San Diego)

Deborah Kilb (Scripps Institution of Oceanography)

Randy Kochevar (Hopkins Marine Station, Stanford University)

Bob Krantz (Cononco-Phillips)

Daphne LaDue (University of Oklahoma)

Tamara Ledley (TERC)

Cathy Manduca (Science Education Resource Center, Carleton College)

Dave Mogk (Montana State University)
Raj Pandya (University Corporation for Atmospheric Research)
Sian Proctor (South Mountain Community College)
Barbara Ransom (National Science Foundation)
Don Reed (San Jose State University)
Leslie Reynolds Sautter (College of Charleston)
Jeffrey Ryan (University of South Florida)
Russ Schumacher (Colorado State University)
Kurt Schwehr (Google Ocean)
Tim Shipley (Temple University)
Sandra Swenson (John Jay College)
Basil Tikoff (University of Wisconsin)
Laura Wetzel (Eckerd College)
David Wunsch (Delaware Geological Survey)
Ken Yanow (Southwestern College)