

EARTHCUBE DATA FACILITIES END-USER & TEST GOVERNANCE ASSEMBLY GROUP WORKSHOP

January 15-17, 2014
Arlington, VA

Executive Summary



Convener

M. Lee Allison, University of Arizona/Arizona Geological Survey,
EarthCube Test Governance Project (Principal Investigator)

Facilitation/Evaluation Team

Joel Cutcher-Gershenfeld, University of Illinois
Jewlya Lynn, Spark Policy Institute

Steering Committee

Tim Ahern, Incorporated Research Institutions for Seismology
Jennifer Arrigo, Consortium of Universities for the Advancement of Hydrologic Science, Inc.
Sky Bristol, US Geological Survey
Cyndy Chandler, Woods Hole Oceanographic Institution
Stephen Diggs, Scripps Institution of Oceanography
Danie Kinkade, Woods Hole Oceanographic Institution
Kerstin Lehnert, Lamont-Doherty Earth Observatory
Don Middleton, National Center for Atmospheric Research
Mohan Ramamurthy, Unidata

EarthCube Test Governance Project Operations Team at the Arizona Geological Survey

Rachael Black
Anna Katz
Kim Patten
Genevieve Pearthree

This material is based upon work supported by the National Science Foundation under Grants No. 1340233 and 1417948. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.



EXECUTIVE SUMMARY

Introduction

EarthCube is a National Science Foundation (NSF) initiative for the development of a community-driven cyberinfrastructure framework to understand and predict responses of the Earth as a system—from the space-atmosphere boundary to the core, including the influences of humans and ecosystems. To fulfill this mission, EarthCube is facilitating the creation of a commons-like environment where stakeholders can bring together existing and new tools, models, databases, software, and collaboration spaces to facilitate the conduct of cross-disciplinary and interdisciplinary research to transform the way we do science.

Data facilities, a stakeholder group representing long-term federal and private investments in cyberinfrastructure, are a key resource in the pursuit of innovative scientific research by aggregating, preserving, and disseminating large quantities of data sets, ranging from highly complex petabyte-scale data to simple metadata catalogs. The visionary potential of EarthCube can only be met if data facilities bring their cyberinfrastructure capacities, researchers bring their data, and federal and private sector funders bring their resources.

One of the first and most fundamental challenges of EarthCube is developing a structure and governance system that will inspire and empower many diverse stakeholders to participate. To develop this governance structure, the EarthCube Test Enterprise Governance Project¹ (a virtual team composed of dozens of partners throughout the country) is implementing a two-year process to identify community-guided solutions for EarthCube governance and test them out—prototyping governance much in the same way as we prototype technology. Community workshops are a key conduit to identifying and testing these solutions.

Workshop Overview

More than 80 leaders from data facilities across the geosciences, regardless of scale, type, or format of data attended. This workshop served two purposes. First, it functioned as an EarthCube end-user workshop—one of a series of 25 NSF-funded workshops targeting a broad spectrum of Earth, ocean, and atmospheric scientists, with goals to generate a clear articulation of the challenges facing each domain and specific ideas on how to address them.

Second, this workshop was the first of four Stakeholder Assembly workshops² convened by the Test Governance project team, bringing together six Assembly groups.³ Assembly workshops will solicit clear guidance on the governance of EarthCube as part of the development of an overarching draft charter, by-laws, and terms of reference⁴ to be presented to the EarthCube community and the NSF for review in June 2014. These documents will guide an EarthCube Demonstration Governance Pilot from September 2014 – August 2015.

Invited participants consisted of key personnel and decision makers representing a range of data facilities. Participating facilities were selected for inclusion according to the following criteria:⁵

¹ Funded as part of the NSF EarthCube awards made in September, 2013. For more information, please see www.earthcube.org.

² Assembly workshops will be convened January – March 2014.

³ These groups include Data Facilities and Users, EarthCube Portfolio, EarthCube End-User Communities & Workshop Participants, Professional Societies, Information Technology and Computer Sciences, Industry & Free and Open Source Software (FOSS): Instrumentation, Software, and Technology Developers. For more information about the Assembly groups, please see: <http://www.earthcube.org/page/assembly-groups>

⁴ Hereafter referred to as a “draft governance framework.”

⁵ Developed by the workshop Steering Committee.

- 1) Facilities funded by the NSF, specifically the Geosciences Directorate (GEO)
- 2) Federally Funded Research and Development Centers (FFRDCs) related to geosciences
- 3) Facilities that house, maintain, archive, repurpose, and generally make data available for scientific consumption
- 4) Facilities can include both those in maintenance and in development mode
- 5) Facilities may be housed within a larger data facility that focuses on a specific constituent group

END-USER WORKSHOP OUTCOMES

Outcomes related to the end-user portion of this workshop include initial elements of a data facilities definition, the operational implications of near- and long-term challenges, problem-solving insights from the past that can lead to innovative solutions in the future, and consensus topics/visions of success.

Elements of a Data Facilities Definition

Participants identified several potential elements of a data facilities definition. A data facility...

- Archives and maintains data
- Makes data available to scientists
- Makes data available to other stakeholders
- Can follow and keep pace with evolving standards
- Curates data
- Ensures trustworthiness and transparency of data
- Provides free data to end users
- Scope broadly accepted by a defined community
- Has a defined and published governance for the facility
- Has a certain level of longevity (defined as ongoing rather than a shorter-term project)
- Provides a value added to the domain end users (i.e., not always the data stream)
- Facilitates the advancement of scientific progress
- Educates other scientific domains
- Is interoperable with other data facilities
- conforms to minimal standards
- Has a defined scope and constituent community
- May fit into a 3-part definition:
 - All data facilities must...
 - All data facilities should...
 - All data facilities may...

Challenges

Participants identified the following societal and operational challenges facing data facilities:

- Creating products that address issues of climate change, mitigating risks for energy development, and other science challenges
- Capacity versus capability challenges for data facilities, which are linked to sustainable funding

- Issues of data quality and transparency, including how people trust data, including quantifying errors, dealing with issues with continuous data, etc.
- The cyclical competitive funding model for data facilities does not result in real infrastructure
- Sustainability, including the need for a periodic technological refresh, workforce development and retention, and changing meanings of vocabularies
- Data preservation
- Capturing the teachable moment by having the right data at the right time AND in useful formats in response to peak demand, such as flood data when a flood occurs.
- Increasingly diverse stakeholders and diverse uses of data
- Deciding which data to archive, since not all can be archived

Problem-Solving Insights

Drawing from past experiences, participants brainstormed how data facilities can find innovative solutions to the challenges listed above.

- Incremental solutions can lead to breakthrough innovations
- External forcing, including funding, can act as a driver to move forward
- Pain points (such as extreme weather) can be motivators for data integration
- Use scenarios can describe the value-added to the private sector
- Bring together people with diverse expertise and use mediators who can bridge across groups
- Find others with expertise beyond your own
- Work across data facilities, including effective in-person collaborations
- Innovations require strong, visionary leadership (champions)

Consensus Topics/ Visions of Success

Participants identified the following consensus topics/visions of success, around which data facilities could potentially organize and make progress on these issues:

- Data will have a proper identifier and will be properly cited, which will lead to better documentation—based on actions by funding agencies and publishers
- Supporting prototype activities, generating exemplars
- Test-bed marketplace to try out new ideas
- Shared infrastructure (including web services and storage)
- A scientific workflow toolkit integrated across EarthCube facilities (drag and drop connecting data sets to algorithms, voice-driven)
- Culture change (developing the workforce, shift in the academic reward systems, data transparency)
- EarthCube data management plans (DMPs) with appropriate specificity
- Council of data centers (inclusive of considering an annual meeting with community-led training)

Next Steps

While there was consensus on issues that could be addressed if data facilities work together, the group was divided among three options moving forward:

- 1) Forming working groups around specific consensus issues;

- 2) Developing a more formalized governance framework/charter (including mission, roles and responsibilities, etc.) to codify themselves as a group;
- 3) Deferring a decision, either because they needed to check with their home institutions first, or because the options moving forward were not entirely clear.

Workshop organizers synthesized the consensus issues/visions of success into potential actionable next steps that each met one of the three options. With consent of the NSF Program Manager for the project, the Test Governance Team offered the participants staff and logistics support and modest funds⁶ for travel, to implement the priorities and goals of actionable next steps, as long as they were also supportive of Test Governance needs. This helped energize the group by providing resources immediately to pursue their goals, a situation many said was unprecedented. With these resources in mind, and with an emphasis on addressing gaps while leveraging existing initiatives, participants self-selected into three groups:⁷

- 1) **Council of Data Facilities:** A first step in developing a more formal data facilities governance framework to enable collective bargaining, particularly with respect to the NSF, and to facilitate greater collaboration among data facilities to achieve individual and collective goals. Participants identified leaders/champions who have already drafted an initial charter ready to be disseminated to the broader community.
- 2) **Rapid Prototyping Working Group:** Formed by participants who wanted to demonstrate tangible progress by leveraging EarthCube developments from the last two years as well as current EarthCube funded projects (Research Coordination Networks, Building Blocks, Conceptual Designs, and Test Governance).
- 3) **Data Citation and Management Working Group:** Formed by participants who want to compile and disseminate resources and best practices for data facilities, focusing on leveraging existing initiatives working on data citation (such as ESIP, RDA and others), and working with the NSF through the Council of Data Facilities to improve the implementation of data management plans across data facilities.

TEST GOVERNANCE ASSEMBLY GROUP OUTCOMES

This workshop produced a significant change from the Test Governance plan originally negotiated with NSF, though overarching workshop goals were still met. We pivoted during the workshop from how we anticipated working with this community and opened up the ways in which we may interact with other EarthCube community groups. Because of the great diversity among the upcoming Assembly Stakeholder groups, it is likely that this workshop represents the first in a series of pivots, and we will reassess our approach after each workshop.

Anticipated Outcomes

The EarthCube Test Governance proposal anticipated that workshop representatives would identify via a chartering process how they want to govern themselves as a group and in relation to EarthCube, and that this charter would be disseminated to EarthCube stakeholders via crowdsourcing mechanisms. This charter would ultimately, as part of an iterative process building on each Assembly workshop, form the basis of a draft governance framework for EarthCube, to be vetted and ideally adopted in complete or partial form before the start of the

⁶ Funding allocated to support two members of an Assembly Advisory Committee that were to be selected from this workshop.

⁷To accommodate option 3, participants could check back with their home institutions, if necessary, before committing to participating in next steps.

EarthCube All-Hands Meeting in June, 2014.

The proposal also anticipated that each group would appoint representatives to an Assembly Advisory Council (AAC), to provide input to the charter and act as a sounding board throughout the Test Governance process. It was later determined that developing a charter and selecting AAC representatives for the entire group were not appropriate due to the diversity of institutions represented, varying levels of prior engagement with EarthCube, limited cohesiveness of the group prior to this workshop, and variation in opinions on actionable next steps.

Actual Outcomes

Although the processes employed to achieve workshop goals were different than anticipated, workshop goals were still successfully met in surprising but positive ways. Workshop participants made significant strides toward:

- 1) **Generating a clear articulation of the challenges facing data facilities and specific ideas on how to address them, and develop a plan of action for how to collaborate moving forward.** The data facilities community identified consensus elements of a data facilities definition, recognized key challenges and operational implications across data facilities, and used insights into how they solved problems in the past to develop innovative solutions to these challenges. Finally, the bottom-up formation of the Council of Data Facilities (CDF) and the two working groups (Data Citation and Management, Rapid Prototyping), each with associated milestones, deliverables and champions, represent actionable next steps that can only be achieved if data facilities collaborate moving forward.
- 2) **Providing clear guidance on EarthCube governance as part of the development of an overarching draft governance framework for EarthCube.** As a result of this workshop, there is now a more codified 'community' that can continue to inform the Test Governance process. In forming the CDF, the data facilities community not only indicated that it wants to organize and govern itself but took preliminary steps to define how it will interact with EarthCube. The CDF included how it will organize within the EarthCube Test Governance framework and how it will act as a coordinating body within the data facilities community.

Modifications Moving Forward

In accordance with the agile development process, we will implement modifications to the process to develop the draft governance framework, each of which will be tested, evaluated, and re-assessed to meet future project needs.

- 1) **Re-envisioning the AAC.** Although the AAC may not exist in the form initially anticipated, each workshop will create emergent leadership opportunities and a community to feed into the Test Governance process.
- 2) **Re-envisioning the Project Timeline.** We now plan to bring together emergent leaders from this and upcoming workshops (as part of an evolving AAC) for a fifth workshop in April 2014 to synthesize outcomes from each of the workshops into the draft governance framework. These drafts will then be disseminated virtually via crowdsourcing mechanisms and strategic pathways exercises to the broad community of EarthCube stakeholders for input and review.
- 3) **Gathering input from the Test Governance Advisory Committee.** We reviewed workshop goals and outcomes with the Test Governance Advisory Committee

immediately following the data facilities workshop. The Committee was encouraged that the Test Governance Team was flexible in its approach, and that the Team was able to accommodate the data facilities' needs as a community. We will meet with the Advisory Committee between the third and fourth Assembly workshops⁸ to review outcomes of workshops thus far and to make final changes prior to the fourth.

- 4) **Clarifying differing interpretations of how workshops will contribute to Test Governance project goals.** We recognize several means for each workshop to generate valuable input into the Test Governance process, and intend to work with the facilitators/Evaluation Team, the extended Test Governance project team, and the Assembly participants to determine the best fit moving forward. These options include:
 - a. Encouraging each Assembly Group to begin to organize as a *community of practice* by self-identifying leadership and crafting a governance framework/rules of the road (chartering process), that then fits into the EarthCube draft charter as a whole.
 - b. Using each Assembly Group to gather as much information as possible to enable the Test Governance Team to design an effective governance structure for EarthCube.
- 5) **Reallocation of funding originally intended for the AAC.** Funding originally designated for two AAC members per Assembly Group will be reallocated to meet that community's governance needs, as long as they are also supportive of Test Governance needs.

Best Practices

Several practices used in the workshop permitted the organizers to achieve their goals, despite the fact that they were achieved in a different way than originally anticipated:

- Using the Evaluation Team to foster agile development
- Organizing an active steering committee
- Meeting people where they already are
- Balancing plenary sessions with small and large group discussions
- Developing clear goals for plenary sessions
- Conducting daily check-ins and recaps
- Identifying opportunities for ongoing leadership and clear, actionable next steps

Lessons Learned

Organizing and conducting this workshop highlighted several processes that were less successful or resulted in a different outcome than anticipated. Lessons related to agile agenda development, clear communication of workshop goals, and communication of the EarthCube and EarthCube Test Governance vision will be tested and improved upon as part of upcoming Stakeholder Assembly workshops.

⁸ This meeting will occur in mid-March between the IT/Computer Science/FOSS and the End-Users/Professional Societies workshops.