Moving EarthCube Forward through Science and Technology Collaboratories

Yolanda Gil, University of Southern California
David K. Arctur, University of Texas Austin
Anna Kelbert, Oregon State University
Jay Pearlman, University of Colorado Boulder
Ken Rubin, University of Hawai'i
Elisha M. Wood-Charlson, University of Hawai'i

May 13, 2015

This document proposes a new process that would *move EarthCube forward through the formation of "collaboratories" that are driven by science needs, aggregate elements resulting from the initial EarthCube funded projects, and build on the extended resources of the EarthCube community. This proposal was designed by participants of the EarthCube Sci-Tech Feasibility Workshop held on April 23-24, 2015, which was the culmination of activities by the community to converge towards a science and technology vision for EarthCube.*

1. Background

As of May 2015, EarthCube has four main types of funded projects for developing the cyberinfrastructure and practices: Research Coordination Networks (RCN), Building Blocks (BB), Conceptual Designs (CD), and Integrative Activities (IA). These all have different levels of expectations and corresponding funding levels, as described in the solicitation[1].

This document proposes an additional process that would move EarthCube forward through the formation of "collaboratories" that are driven by science needs, aggregate elements resulting from the initial EarthCube funded projects, and build on the extended resources of the EarthCube community. These collaboratories would result in an EarthCube Resource Portfolio (of developed project ideas and components) that would enable meaningful interactions between geoscientists and computer scientists, and foster development of EarthCube functionality in several ways, as described below.

The proposed process could be supported through a new funding mechanism for "collaboratories", which would fall between IA and BB in terms of expectation and funding level.

The rest of this document gives an overview of the proposed process, describes the role of collaboratories, and discusses how they build on earlier EarthCube efforts.

[1] NSF 13-529 solicitation, http://nsf.gov/funding/pgm_summ.jsp?pims_id=504780&org=NSF&from_org=NSF

2. Proposed Process: Science Driving EarthCube through Collaboratories

Overview

Figure 1 gives a diagrammatic overview of the proposed process, which is described in detail in the rest of this section. Use cases and scenario drivers would result in broad science scenarios and specific use cases that could then be mapped through a synthesis activity that would involve mapping use cases to broad scenarios, identifying relevant technology capabilities, and annexing or developing unique resources as appropriate. The organizational unit for this synthesis activity would be a "collaboratory", which would 1) encourage the adoption and development of advanced technologies and 2) promote testbed projects, enhancements to data facilities, and other community-driven integrative projects. Collaboratories would result in an EarthCube Resource Portfolio, which would embrace both germinal and more mature projects. The portfolio would consist of a dynamic, evolving set of resources at various stages of development that are visible to the community, promote direct engagement by domain scientists and technologists, and are accessible to other funding bodies and support mechanisms.

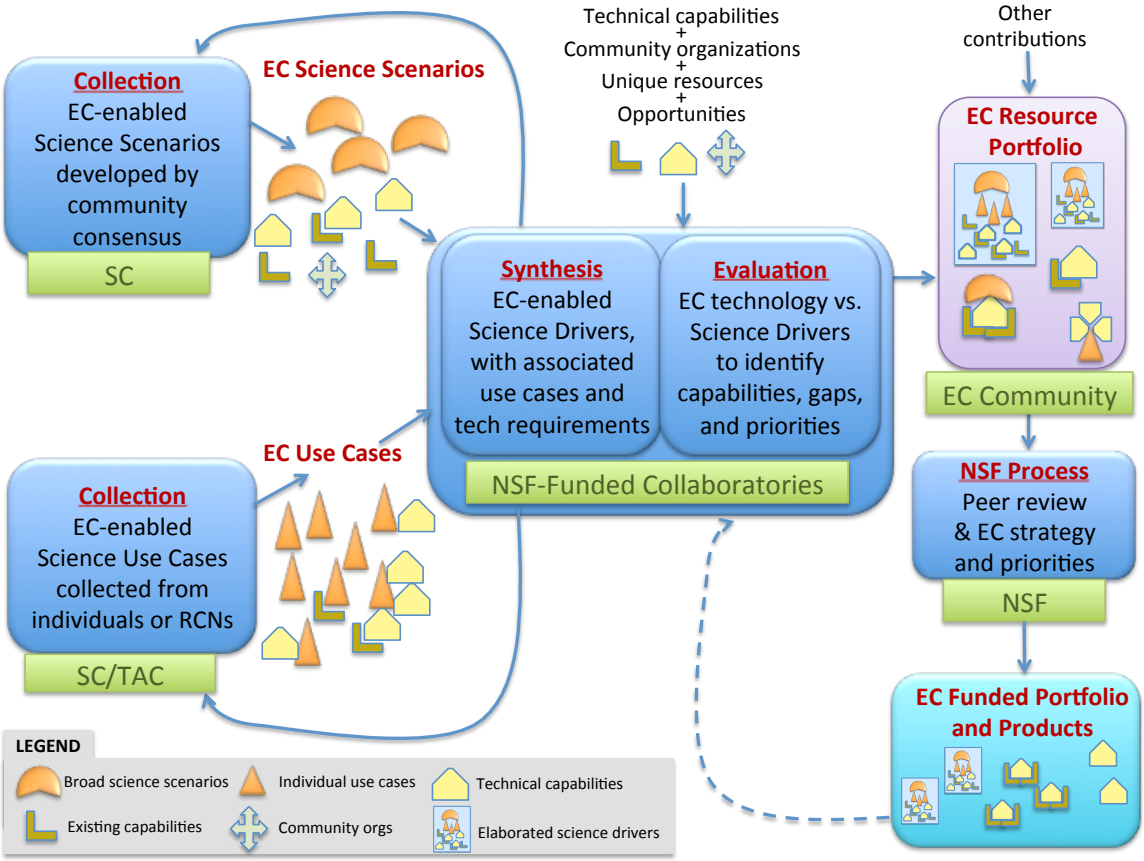

Figure 1. A two-prong strategy to collect general science scenarios as well as specific science use cases would then be mapped through synthesis and/or technology evaluation activities carried out by collaboratories that would result in the EarthCube Resource Portfolio.

Science Drivers

Science scenarios and use cases capture specific contexts where technology can be inserted in science practice, and can therefore play a crucial role in facilitating communication and fruitful interactions between computer scientists and geoscientists. A guided framework for such interactions, informed jointly by the science goals and by a thorough understanding of technological capabilities, is needed to direct EarthCube in its efforts to support original geoscience research.

Recognizing this, EarthCube should adopt a two-pronged strategy for stimulating and driving its cyberinfrastructure development:

(1) A domain-driven effort that would assemble science scenarios that have emerged as a community consensus from the End User Workshops, Research Coordination Networks (RCNs), and other geoscience community activities. These science scenarios or 'generalized' use cases would elaborate the vision of the EarthCube Strategic Science Plan Geoscience 2020[2] (§5). These scenarios would articulate science goals, highlight technology requirements, and point to existing resources (e.g., an existing data facility or an organized community) relevant to achieving the science goals.

(2) A user-driven/computer-scientist-aided effort that would focus on use cases contributed by individuals or a coherent group of scientists (e.g., RCNs), whose research agenda(s) would be advanced through specific integrative science and (existing or to yet be developed) technology and resources.

Both types of science drivers could represent either existing, near-term needs or a longer-term vision of areas of significant potential impact for research innovation.

Collaboratories

This dual approach would result in broad science scenarios and specific science use cases that could then be mapped, through a synthesis activity, to identify relevant technology capabilities, organize communities around synergistic topics, elaborate commonalities across scenarios and cases, and annex or develop unique resources as appropriate.

The challenges of communicating technology needs across the computer scientist-geoscientist divide, the person-hour intensive nature of this effort when scaled up to a large number of use cases, and the difficulties of doing this within the current EarthCube funding vehicles suggest that this new type of organizational unit and funding mechanism for this synthesis activity would be very useful. In fact without it the community is at risk of not having a sufficiently broad and representative group of multi- and cross-domain use cases and workflows to realize the full range of expectations for the eventual EarthCube infrastructure.

[2] http://earthcube.org/document/2015/earthcube-strategic-science-plan

We have called such organizational units *collaboratories*, and expect them to be formed by the community and supported through both volunteer work and funded activities. A collaboratory would 1) encourage the adoption and development of advanced technologies and 2) promote testbed projects, enhancements to data facilities, and other community-driven integrative projects that can be evaluated in the context of their use in original geoscience research. A collaboratory would also effectively connect with a broader range of users to better address community-wide opportunities and challenges. Much of the collaboratory support would be in the form of salary for participants.

The EarthCube Resource Portfolio

From the collaboratories and their activities, an *EarthCube Resource Portfolio* would emerge, which would embrace both germinal and more mature projects. The portfolio would consist of a dynamic, evolving set of viable projects and functional resources that are visible and would be made available to the community, and promote direct engagement by domain scientists and technologists. The projects in the portfolio would then be used to drive community-supported, science-driven cyberinfrastructure development efforts submitted as proposals to future EarthCube amendments, and reviewed through the NSF peer-review process. On the other end of the maturity spectrum, the functional portfolio resources that are already associated with science communities of users could be adopted for maintenance by the NSF or other funding sources, to ensure sustainability.

3. How Collaboratories Build on Prior EarthCube Activities

Science Scenarios: Drawing from EarthCube End User Workshops and RCNs

From mid-2012 through 2013, EarthCube sponsored 24 End User Workshops targeting a broad spectrum of Earth, atmosphere, ocean, and other geoscience-related research domains. These workshops spanned a wide range of geoscience domains with various levels of pre-existing coordination, community cohesion, and cyber-resources[3]. The broad goals and aspirations of these communities were distilled by the EarthCube Science Committee into the EarthCube Science Strategic Plan ("Geoscience2020"). This Plan, as well as the workshop reports themselves, lack the specificity required to develop actionable use cases for technology development.

One additional outcome of those workshops was the development of several domain specific Research Coordination Networks (RCNs) designed to engage a broader research community and explore their needs for technology and cyberinfrastructure, specifically highlighting gaps that hinder current research efforts. RCNs are funded to host two community workshops to crowdsource ideas and develop plans to address the identified technology and cyberinfrastructure challenges. Some of these challenges may already be

[3] The combined executive summaries of EarthCube End-user Workshops can be found at http://earthcube.org/sites/default/files/doc-repository/CombinedSummaries_12Dec2014.pdf

addressed through existing EarthCube cyberinfrastructure. However, several RCNs, such as ECOGEO (environmental 'omics), represent scientists, technologists, or research topics that do not fit into traditional geoscience frameworks. Therefore these RCNs have unique domains needs, specific use-cases, and science scenarios that require an additional CI resource or bridge be constructed before EarthCube can begin to enable their science.

A collaboratory would be an environment and vehicle to enable the end-user workshop communities, other science domains not represented by RCNs, and the aforementioned RCNs that are not traditionally "geoscience" research domains to build resources or designs that allow a community to overcome identified gaps in their science research and better leverage current/planned EarthCube resources.

Furthermore, collaboratories can provide a working environment for new opportunities in crosscutting research themes. The EarthCube End User Workshops were generally focused on specific disciplines/domains, yet many of the advances needed to address grand science challenges (see the 2015 EarthCube Strategic Science Plan), such as climate change, are highly cross-disciplinary. Looking to the future, facilitating more complex use cases across disciplines and having an environment for test and validation is a natural strength of Collaboratories for geoscience research.

Collaboratories: Drawing from EarthCube Technology and Architecture Foundations

As EarthCube's initial funded projects progress, they will result in technology components (e.g., the EarthCube Building Blocks), architecture concepts (e.g., the EarthCube Conceptual Designs), initial integrations of technical capabilities, and other technological advances. They will provide components that can be incorporated into the EarthCube Resource Portfolio.

Collaboratories will help facilitate the mission of the EarthCube Technology and Architecture Committee by providing a framework that will enable: 1) coordination between technology components and scientific requirements, 2) identification of technology gaps, 3) development of testbeds for evaluating technology components, 4) assessment of standards and other architecture concepts.

Collaboratories would also provide a framework where the geosciences and the cyberinfrastructure experts can work closely together on joint research initiatives, which would significantly extend the value of EarthCube to the broader community.

4. The EarthCube Resource Portfolio

Resources in the EarthCube Portfolio

The EarthCube Resource Portfolio resulting from the collaboratories would include a heterogeneous range of science-driven project ideas and functional technology with

different composition and degree of maturity, implementation, and interdisciplinary work. Portfolio projects could include:

1. Scenario-centered projects: These could include a science scenario with several use cases that can be mapped to that driving scenario. Initially, they may not include technology components, but the collaboratory participants and their science communities would converge on a specification of these components that would be included as part of the project as it enters the portfolio.
2. Implemented science use-cases: Projects resulting from direct interaction between computer and domain scientists, initially embedding full technological specifications and eventually including general reusable implementations of those specifications.
3. Technology integration opportunities: These could range from integrated feasibility demonstrations, to testbed experiments, to mature integrated systems.
4. Tech transitions to data facilities: Of particular interest would be integration opportunities that deploy new technologies to existing data facilities, as driven by science needs.
5. Exploration of architecture concepts: These projects could integrate selected technology capabilities and relevant standards to experiment with and test specific architecture concepts.
6. Innovative prototyping: an environment to explore new ideas, facilitating rapid prototyping and the collection of early feedback from scientists.
7. Science-driven building block integration: projects that require integration of existing EarthCube funded technologies to enable the science objectives

The EarthCube Resource Portfolio would evolve over time, reflecting the community's interests as they shift due to opportunities, science priorities, and technology evolution.

Benefits of Maintaining an EarthCube Resource Portfolio

The benefits of maintaining an evolving EarthCube Resource Portfolio include:

- Scientists can relate to the scenarios and use cases and seek technology partners to formulate projects
- A flexible and inclusive environment for exploring new research directions
- An open, collaborative setting that would encourage scientists in smaller facilities to reach out and get involved
- Technology developers can analyze the scenarios and use cases and propose collaborations with scientists that developed them
- NSF can use them as an indication of what the community wants to pursue
- Other agencies, foundations, and institutions can volunteer funding or resources to pursue particular projects in the portfolio
- Well-developed and used portfolio resources could be adopted for maintenance, promoting sustainability
- Outreach to other scientists who would be able to understand what EarthCube can do for them by looking at the science scenarios and use cases

- Inform government observers of what is happening in EarthCube in terms of advanced cyberinfrastructure capabilities and their science impact
- Leverage other funding as individual researchers can align their funded projects from other sources with EarthCube and explore synergies, pursue integration efforts, and make contributions to EarthCube
- Life-long learning by using the portfolio resources for hands-on demonstration of cutting-edge synergistic technical and scientific advances that could be made accessible to geoscientists, technology developers, students, decision makers, and the public

Acknowledgments

This document is the result of discussions at the EarthCube Sci-Tech Feasibility Workshop held on April 23-24, 2015. Workshop participants included Tim Ahern, David Arctur, Dave Fulker, Basil Gomez, Tom Guilderson, Anna Kelbert, Danie Kinkade, Kim Miller, Jay Pearlman, Steve Richard, Kathleen Ritterbush, Ken Rubin, Durelle Scott, Elisha Wood-Charlson, and Ilya Zaslavsky.