Interviews with EarthCube Funded Projects: Perspectives on EarthCube's Impact and Sustainability Priorities

August 25, 2022

Karen Stocks and Helen Evans

EarthCube Office

EXECUTIVE SUMMARY

After nearly a decade, the NSF-funded EarthCube program is coming to a close. Aimed at furthering the geosciences with cyberinfrastructure, EarthCube encompassed a set of funded projects, as well as community activities that engaged the geosciences, cyberinfrastructure, computer science, and associated communities, overseen by the EarthCube Office (ECO) and EarthCube governance bodies.

To guide the EarthCube Office and governance on transition and sustainability planning, ECO staff conducted small group interviews with the leads or designated representatives of current and completed EarthCube projects to ask about the value and impact of EarthCube, and the priority components to sustain past the end of EarthCube. In addition, if time allowed, participants were asked about the largest unmet cyberinfrastructure / data / informatics-related challenge in their field.

Across almost every discussion group, a clear consensus emerged that the highest perceived value of EarthCube was around networking and community building. This brought concrete value to participants by fostering new collaborations, expanding professional networks, and creating a forum for projects and individuals to learn from each other. This was particularly important for students and early career researchers, becoming a formative or career-changing experience for some. The EarthCube community was seen as unique in the degree to which it includes both technologists and geoscientists, was multidisciplinary, and was "friendly" and "nurturing."

These results are consistent with a poll given at the 2022 EarthCube Annual Meeting, where 61% of respondents indicated that they have had or currently have a project with a collaborator they met through EarthCube activities, while 65% have an in-development or planned project with a collaborator they met through EarthCube activities. Furthermore, 54% responded that one or more of these collaborators is early career.

Overall, there was a strong call for the community to be sustained going forward, a recognition that building a community in this way takes substantial time, and concern that the current community will dissipate without EarthCube. With respect to specific EarthCube activities that participants felt should be persisted, the activities most often mentioned were: community meetings (an Annual Meeting but also potentially more frequent interactions), the Council of Data Facilities, GeoCODES/schema.org, and notebooks as scholarly objects.

INTRODUCTION

A partnership between the Directorate for Geosciences (GEO) and the Office of Advanced Cyberinfrastructure (OAC) at NSF, EarthCube consists of interconnected projects and activities that engage the geosciences, cyberinfrastructure, computer science, and associated communities. The goal of EarthCube has been to transform the conduct of geosciences

research by developing and maintaining a well-connected and facile environment that improves access, sharing, visualization, and analysis of data and related resources.

Outside of the individual funded projects, the EarthCube Office and governance has supported the following community-focused activities. The EarthCube website (www.earthcube.org) provides further detail on each.

- Governance committees: the Council of Data Facilities, Technical and Architecture Committee, Science and Engagement Team, Council of Funded Projects, and Leadership Council, as well as their working groups.
- An Annual meeting
- Community communications: monthly newsletter, central website, Slack workspace with multiple channels, Youtube channel, and social media presence.
- Technical infrastructure: the GeoCODES resource registry and supporting schema.org implementations, and the review and publication of notebooks as scholarly objects.
- Community education: webinars, data help desks at scientific conferences, and informational guides/presentations/blogs.

After nearly a decade, the EarthCube program is coming to a close. To guide the EarthCube Office and governance on transition and sustainability planning, the EarthCube office staff conducted group interviews with the leads or designated representatives of current and completed EarthCube projects to assess perspectives in two areas. First, each interview asked a set of questions on the value and impact of EarthCube and the priority components to sustain past the end of EarthCube. In addition, if time allowed, participants were asked about the largest unmet cyberinfrastructure / data / informatics-related challenge in their field. A second set of questions asked about projects' primary challenges and lessons learned was included, and is presented in a separate document.

METHODS

Discussions were done as small groups of 3-5 people or, in a small number of cases where scheduling was difficult, as individual interviews. In total, 49 people who participated in 60 EarthCube awards took part (some people were part of more than one EarthCube award). All current and past funded projects were invited, and the participants include those from the first round of EarthCube funding in 2013 to the last in 2021. Participants also represent a range of engagement in EarthCube outside of their funded project work, some having served on governance committees, been involved in the early EarthCube planning and workshops, etc. Each interview was asked the following specific questions

- Outside of your project funding are there other benefits that EarthCube has brought to you?

- Looking past the end of EarthCube, are there EarthCube elements or activities that you think should continue? Which activities would you want to participate in even if there was no funding?

To encourage free discussion, interviews were not recorded and two EarthCube office staff took simultaneous notes during each interview. *The indented statements in italics below express the opinions of individual interviewees, but while every effort was made to capture comments as precisely as possible, they are not always exact quotes.*

RESULTS

General Benefits of EarthCube

Networking and Community Building

Across almost every discussion group, a clear consensus emerged that the **highest perceived value EarthCube of was around networking and community building**. This brought concrete value to participants by **fostering new collaborations** and expanding professional networks. EarthCube participation was frequently mentioned as launching new collaborations that led to successful proposals, both within the EarthCube solicitation and elsewhere. Participants specifically mentioned that they met people through EarthCube that they would not have met otherwise, sometimes even from their own institution. This was **particularly important for students and early career researchers**, several of whom emphasized the impact EarthCube has had on their career, calling it "formative" and saying that it served as a "stepping stone" for moving into geoinformatics.

> *I would not be where I am now professionally without EarthCube*

The EarthCube community was broadly perceived as unique and particularly valuable because it **spanned information technology and geosciences** – no other venue was seen as bringing together those two communities as effectively, and it was seen as "one of the most valuable outcomes" of EarthCube.

> *Bringing together science and cyber is a unique EarthCube thing, and this is a hard thing to do*

In addition, it is **multidisciplinary**, allowing geoinformatics projects to break out of domain silos to learn from projects with similar goals across all the geosciences.

A Learning Community

Structured meetings and informal networking were also important for projects in different scientific application areas to **share approaches, ideas, and technologies**. Participants very frequently mentioned the value of learning from each other, and "seeing what other groups are doing"; the multidisciplinarity of EarthCube was key to this. Beyond technical sharing, the EarthCube community frequently noted as easy to join; it was "friendly" and "nurturing."

> *EarthCube has a low barrier to entry...it is a place where people will take the time to explain the technical details, a good place for people without deep technical expertise to learn.*

Geoinformatics Progress

NSF identifying geoinformatics as the focus of a coordinated effort was seen as having both symbolic and practical importance. Symbolically, the NSF investment in EarthCube signaled a priority on geoinformatics and its importance to the future. It "shone a spotlight" on data and cyberinfrastructure that motivated communities to move towards cyberinfrastructure and open science in general. In a practical sense, it provided support for work that was previously difficult to get funded and not seen as a priority. A sense that the field has moved forward in important ways over the past decade was commonly expressed.

Sustaining the EarthCube Community

There was a strong call for the community to be sustained going forward, and a recognition that building a community in this way takes substantial time. Participants frequently mentioned being "sad" and "disappointed" that there was no clear sustainability plan for EarthCube as a whole. There was also concern and confusion over whether this represented a shift in NSF away from geosciences cyberinfrastructure work, as no follow-on programs or major activities have been announced.

> *Having a scientist-centered technology group is critical for NSF to move into the new technology realm*

Several discussion groups, particularly those with the older, completed EarthCube projects, explicitly recognized that it took time for EarthCube to evolve, and important learning about how to effectively frame and manage the effort and support the community occured. It was seen as just hitting its stride now after initial evolution in focus, organization and governance, and there was concern that in ending EarthCube and starting one or more new initiatives in its place, NSF will lose the focus and cohesion that EarthCube has developed.

Many participants strongly called on NSF to continue EarthCube's successful trajectory. Outside of this, the pathway for preserving the community was not clear: it was noted that ESIP does not have the same degree of engagement with geoscientists, and the AGU Fall Meeting is so large that participants stay within their domain or informatics track, and has less participation from computer scientists/technologists. Several suggested creating a nonprofit "EarthCube, Inc" to continue the work (though funding for this option was not clear). A couple of participants noted that incentives to participate were important; they realistically would not have participated in governance and annual meetings without funding and a clear expectation. These were the same people who identified the community activities as high value, so this may be more a statement about overcommitment than perceived value. Two participants suggested that a similar "cohort" could be launched for NSF geoinformatics or CSSI award recipients.

> *Collaborative activities…need to be the seed of anything further, and definitely need to continue.*

> *EarthCube occupies an important niche [...] a place where exciting science enabled by geoinformatics is showcased*

Specific EarthCube Activities/Components

With respect to specific activities carried out by the EarthCube Office and Governance, a very broad array was perceived as valuable. The Annual Meeting; governance groups (Council of Data Facilities, Council of Funded Projects, TAC); technical infrastructure (GeoCODES and the related schema.org work, notebooks) and guidance (the EC Standards and Specifications document); and outreach activities (Data Help Desk, AGU booths, workshops, FAIR training materials, the Jargon page, the website) were all mentioned at least once as having value. This speaks to the diversity of the EarthCube community, with different communities entering at different levels and having different needs.

Within this diversity of responses, the **activities most often mentioned as valuable and worth persisting were: community meetings** (an Annual Meeting but also potentially more frequent interactions)**, the Council of Data Facilities, GeoCODES/schema.org,** and **notebooks as scholarly objects**.

Future meetings were highlighted as important for persisting the networking and community building aspects of EarthCube that were so highly valued. As mentioned above, no other venue was seen to have the same characteristics. Various models were suggested: an annual meeting similar to the current one held either in conjunction with a major meeting like AGU or independently; workshops that identify a particular challenge and bring together scientists and technologists to progress it; training/educational webinars focused on moving scientists towards cyberinfrastructure.

> *There is lots of novel thinking at the Annual Meeting. It was an exciting time when I was engaged in my EarthCube project and meeting other funded projects*

While TAC and CFP were also mentioned, the CDF was the governance group most often mentioned as important to maintain. It was seen as a venue for facilities to recognize common needs across data facilities, identify ways to address them, promote standardization, exchange information, coordinate activities, and speak with a cohesive voice to external groups, e.g. journal publishers, government bodies developing policy. It was noted that the CDF is not just for sharing information, but also supporting working groups that have impact across repositories.

> *CDF serves a function for disparate entities to speak with one voice*

The GeoCODES cross-domain search interface and supporting schema.org development and community education were seen as valuable both from the perspective of scientists finding data, and of repositories wanting their data to be findable. It was seen as having high impact for the investment, and as showing "uptake". One group suggested that NSF could encourage supplement grants for developing schema.org extensions within specific geosciences domains.

> *GeoCODES has a lot of potential; a way to maintain, grow and support it should be found*

Participants were enthusiastic that EarthCube recognized and promoted the value and dissemination of non-traditional outputs. This included code repositories, but most frequently the notebook work was mentioned. Several participants who were not aware of this technology before their EarthCube interactions are now planning their own calls for notebooks in their community, integrating them into their classroom teaching, or planning hackathons around notebooks. In addition to promoting and facilitating access to EarthCube project resources, the notebook collections gave newer users models for how this technology can be used. There was a strong call for this "important work" to continue, with future notebook calls, as well as work on notebook standards, how to version notebooks, etc.

> *[The notebooks activities are] the best thing from EarthCube!*

Both the **Data Help Desks** and the original **end-user workshops** were also mentioned as valuable multiple times, though less often than the above activities. The Data Help Desks were considered "a good activity that needs to continue" and "a smart community service". The end-user workshops were seen as getting some communities starting to talk about their data, prompting new work and attracting people who continued to stay engaged.

> *I have repeatedly referred to those white papers [through the years] and it may be valuable to close EarthCube by holding another set to inspire the next phase of geoinformatics.*