EarthCube Project Challenges and Lessons Learned:

Perspectives from Current and Past EarthCube Projects

August 25, 2022

Karen Stocks and Helen Evans
(EarthCube Office)

EXECUTIVE SUMMARY

The projects funded during the almost 10-year NSF EarthCube program represent a substantial component of the recent geoscience cyberinfrastructure (CI) development in the US. Collectively, the set of EarthCube funded projects have gained deep and broad knowledge about effective strategies and critical considerations for geosciences CI from which future projects can benefit. To capture this knowledge, participants from 60 funded projects were interviewed about what their greatest challenges were and what lessons they learned during their work. They were also specifically asked about their strategies for sustaining their resource(s) past the end of EarthCube funding, and for increasing Justice, Equity, Diversity and Inclusion (J+EDI).

With respect to sustainability, there was almost complete consensus that a three-year award is not long enough to complete the process of technical development, hardening into an operational resource, and user engagement. Furthermore, these are distinct activities that often require different expertise, have different metrics of success, and should be explicitly planned. With respect to J+EDI, projects identified the low level of diversity within both geosciences and computer sciences as being challenges to cultivating diverse project teams. Those that had robust J+EDI initiatives generally partnered with an existing organization, such as a program to support first-generation college students, or an established STEM education effort in underserved communities. These results supplement the findings of the EarthCube Report on *Sustainability Models for Integrated Digital Earth Science*, which interviewed non-Earthcube earth sciences infrastructure projects that had successfully sustained for over 10 years[1].

While responses varied, certain additional challenges emerged as common: cultivating uptake in the target user community; attracting and retaining skilled students and staff, particularly those with technical skills; managing diverse and distributed teams composed of members with different motivations and priorities, especially technology and geosciences domains; connecting to and leveraging from external infrastructure that could change technical requirements for use, shift timelines, or completely close down. Finally, the COVID pandemic impacted EarthCube projects, just as it impacted science, and society, as a whole.

While the set of challenges was daunting, the funded projects provided extensive input on successful strategies, key considerations, and pitfalls to avoid, for the benefit of future projects.

INTRODUCTION

EarthCube was an NSF-funded program with the goal of transforming geosciences research by developing and maintaining a well-connected and facile environment that improved access, sharing, visualization, and analysis of data and related resources. It consisted of interconnected funded projects, supported by community-building as well as educational, infrastructure, and

[1] https://doi.org/10.6075/J0JH3MBN

governance activities. A collaboration between the Directorate for Geosciences (GEO) and the Office of Advanced Cyberinfrastructure (OAC) at NSF, the EarthCube community spanned geosciences, cyberinfrastructure, computer science, and associated communities.

After nearly a decade, the EarthCube program is ending. Recognizing that developing effective cyberinfrastructure is not simple, straightforward, or inexpensive, the set of EarthCube funded projects collectively represent deep and broad wisdom about effective strategies and critical considerations from which future efforts can benefit. To capture this, the EarthCube office staff conducted group interviews with the leads or designated representatives of EarthCube projects, both current and past. Each interview posed a set of questions about projects' primary challenges and lessons learned, including questions on sustainability and Justice, Equity, Diversity, and Inclusion (J+EDI). The goal was to solicit and consolidate the community's reflections into a set of informal guidance to aid current and future geoinformatics projects. The participants were also asked about their perceptions of the value of the EarthCube program, and whether components of EarthCube Office activities should persist; feedback on these latter questions is reported on in a separate document.

Methods

Discussions were conducted via videoconference in small groups of 3-5 people or, in a small number of cases for which scheduling was difficult, as individual interviews. In total, 49 people who participated in 60 EarthCube awards took part. (Some participants were part of more than one EarthCube award). All current and past funded projects were invited to participate, and the participants include those from the first round of EarthCube funding in 2013 to the last in 2021. Each interview asked the following questions:

> What was your primary challenge? What lessons did you learn from your work (i.e., what do you know now that you wish you knew then)?
>
> Did you have any effective strategies or lessons learned around promoting Diversity Equity and Inclusion? Around project sustainability?

To encourage free discussion, interviews were not recorded. Instead, two EarthCube office staff took extensive 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 may not be exact quotes*

RESULTS

Sustainability

Challenges

Projects were specifically asked about their sustainability challenges and lessons learned. Sustainability past the end of EarthCube funding is a pervasive concern for many projects.

There was almost complete consensus that a three-year award is not long enough for the necessary process of technical development, hardening into an operational resource, and user engagement. Furthermore, these are distinct activities that often require different expertise, have different metrics of success, and should be explicitly planned.

By and large, a clear pathway to cyberinfrastructure sustainability was not seen. While there was a recognition that not all projects should be sustained, there was a common concern that NSF has not developed a clear model for cyberinfrastructure sustainability, and this hinders both the persistence of valuable individual resources and the development of a more interconnected cyberinfrastructure landscape.

> *[Projects need] to cross the Death Valley of going from research software to operational software*

Lessons Learned

- The *most common* sustainability approach used by projects was seeking sequential funding, though this was usually opportunistic and could be inefficient, as goals and focus shifted across grants, and staff could be lost due to funding uncertainty.
- The *most recommended* approach was embedding the resource in an existing institution with long-term support. While unsupported community sustainability of open-source code is possible, it is difficult to move a project to that level, and requires planning and engineering to persist in the long term; begin cultivating your community of sustainers early.
- Creating a new organization, such as a not-for-profit or Limited Liability Company (LLC), can allow more diverse forms of funding, such as either mandatory fee-for-use or voluntary contributions (“NPR” model), but is also difficult to establish.
- Building a user base is critical for sustainability: if the resource is not used, it will not be sustained.
- Begin thinking and communicating about sustainability from the start.
- Consider development targets that have high value but low overhead, and can persist with minimal support, hosted and run with very little cost.
- Explore institutional infrastructure that may be available for sustainability; funded projects found a diversity of organizations to connect to, from university libraries to successful open-source programs and international entities.

Justice, Equity, Diversity, and Inclusion (J+EDI)

Challenges

As with sustainability, projects were specifically asked to comment on J+EDI challenges and approaches. Overall, the earlier projects had less activity around J+EDI and were more likely to respond with information about the diversity of their project team, and explain that J+EDI work was not an explicit project expectation. Some expressed that they were limited by the diversity of the pool of students and professionals in their fields. Current projects were markedly more likely to discuss pro-active education, training and mentoring efforts to increase J+EDI. This may indicate a successful shift in expectations within the EarthCube community towards more active responsibility for increasing J+EDI within geoinformatics. Many made the point that open data and open software, and resources with very low barriers to entry, all work to increase equity.

Lessons Learned

- Discuss, define, and track specific J+EDI targets.
- For greatest impact, collaborate with an existing organization with a J+EDI mission. Your institution may have a program to support first-generation college students, or STEM education in underserved communities; you can seek project partners at Minority Serving Institution (MSI), or non-governmental organizations active in capacity building in developing nations.
- Free virtual workshops, webinars, and hackathons (though often originally driven by COVID), engage a much broader and more international community than similar in-person events. Not all activities are as effective virtually, but it is worth considering whether some can be included.
- Physically go to the communities you are trying to engage. For in-person events, if your lead organizations are not MSIs, consider reaching out to a nearby MSI to discuss hosting your event, or a satellite event, on that campus.
- Consider students who do not have a science background, but are interested. Find a “hook” for advertising student positions - flying drones, interesting field work, high-tech visualizations - to attract a wide pool of students.
- Create opportunities for diverse people to work together. Even a one-time experience can lead to longer term collaborations.
- Twitter was a good way to reach a broader audience.
- Design your resources to be immediately usable, accessible with minimal expertise, and as well documented as possible. (And ensure your documents and presentations are themselves accessible). Any effort to lower the bar for uptake of CI resources expands the potential user community and broadens accessibility. Consider the point of view of

an audience who may have less technical expertise when making it your resource welcoming (e.g., classic GitHub pages can be unfriendly).

- Welcome, engage, include, thank, and give visibility and recognition to all contributors, especially those who may have less prominence in the project.
- Create modular lesson units to help educators who may not be experts to teach your resources.
- Individual mentorship can be very powerful.

Uptake

Challenges

A strong user base was seen as critical for project success and sustainability, but creating a user base needs effort and planning. Groups generally underestimated the time and difficulty of engaging the user community and supporting uptake through training and outreach. This often requires different skills from initial technical development.

> *"Build it and they will come" doesn't work*

Lessons Learned

- Plan early and devote resources; uptake and buy-in was named as the largest challenge by several groups, and the most critical for future sustainability.
- Be steered by community needs - have ways to ask your community what they want, and the flexibility to adapt plans to give them that. Use cases / user scenarios can be very helpful tools.
- Make adoption as easy as possible: meet researchers where they are at (and finding this out can take effort), integrate into existing workflows where possible and/or be flexible enough not to require one specific pipeline, make well-documented examples so users can immediately start playing with something without installing a lot. Notebooks and hackathons are two approaches that projects have found helpful. Have a low bar for initial entry, and then show how they can do more.
- You will need a translator or bridge person who can pitch to geoscientists without jargon...this is often not a developer, as it needs different skills.
- Be ready to repeat; it can take several contacts before people understand what you did and why it is useful. The more times you can say it, the better.
- When you have a resource ready for beta use or full adoption, identify potential adopters, and ask for an hour to sit down with them and show how to apply your resource to their specific problems. When seminar speakers come to your institution, ask

for a meeting with them. One-on-one pitches, targeted to someone's specific work, can be effective.

- Involve early career people as they can be more willing to adopt new approaches, and then connect them back to the more senior researchers. Social media can be effective here.

Human Resources

Challenges

Attracting and retaining personnel was a primary challenge, particularly technical staff and students. The labor market is tight, with high demand for talent, and positions at universities and other non-commercial research institutions are seen as less stable and lower-paying than the commercial sector. Further, many projects need part-time expertise (e.g. a third of a programmer, a couple of months of a web developer) and this is difficult for projects to find. Multiple interviews suggested strengthening the workforce training pipeline, as well as creating a pool of shared technical expertise with a geoinformatics background that could split time across projects.

> *The problem was that it all relied on students to write code, but they kept getting hired off by commercial companies*

Lessons Learned

Unfortunately, few robust solutions were presented to the difficulty of attracting and retaining technical staff. Sharing staff with other projects can help meet the need for a short-term person and balance funding over time, but the practical difficulties of arranging and sustaining this were also clear. Two projects noted that they had better outcomes finding a geosciences student/postdoc with coding experience and building on those skills, vs hiring from computer science.

Project Management and Collaboration

Challenges

Keeping a diverse and distributed team aligned and productive is not easy and was one of the most frequent challenges mentioned. Ensuring alignment across distributed team members and productivity towards group goals with many part-time and often unfunded participants, each with different motivations, is difficult. For some communities, sharing and collaborating is still seen as a competitive disadvantage.

> *The real challenge here was that the incentives/reward structure for the different collaborators is different*

Lessons Learned

- The project leader is key; they need to see and communicate a broad shared vision, and understand the connections both within the project team and to outside groups. Identify critical partners and bring them in as Co-PIs. For more than one team, a change in leadership was needed when project management was not going well.
- Allocate funding carefully. On the one hand, distributing funding can provide incentives for collaborators to engage, and motivate them to meet expectations (unfunded collaborators can be difficult to hold to timelines and specific outcomes). On the other hand, a fully devoted person can be much more productive than several part-time people split across institutions.
- Establish good communication. Regular meetings (weekly or biweekly were common), a Slack channel, a newsletter, shared Google documents, etc., have all been used effectively. Groups varied on the right balance of all-team meetings vs smaller group discussions: frequent all-hands meetings can promote team cohesiveness, learning, and shared vision; but smaller groups can be more productive for accomplishing work.
- Be goal-oriented and metric-oriented.
- Create and maintain a shared document or project management tool that tracks the arc of the project from beginning to end, so each sub-team is managing their part, but has a common view of the tasks, blockers, etc.
- Be kind to each other and to yourself.

Bringing Computer Science (CS) and Geosciences Together

Challenges

Collaboration across technology and geosciences expertise was seen as necessary to create technical solutions that meet real science needs (and was frequently highlighted as a strength of EarthCube as a whole), but is not simple to establish. For Computer Science and Data Science researchers, geoinformatics needs are not seen as cutting-edge research problems, but rather as applications of (recent) advances, and do not lead to research publications on the CS side, nor to commercial products. At the same time, geoscientists similarly may see little professional benefit in infrastructure development.

> *What is cutting edge for one side does not engage the other side*

Lessons Learned

- Publications are important to many collaborators - plan papers and have broad author lists. Discuss planned products with the entire team, so all members can contribute and can see the importance of the work. More generally, explicitly recognize the different goals and needs of the science and technology members of the team, and plan ways to meet them.

- Find science success stories and elevate them, and make clear in publications and presentations that the science success could not have happened without the technology.
- Learning to effectively communicate across the science-technology divide takes time and is iterative. Frequent team meetings are important to learn the language and motivations of the different participants; build in time for feedback, questions, and discussion. Give external presentations and posters to the whole project team, ideally in advance of the external conference or event, even when focused on a narrow technical or geosciences aspect. This fosters group understanding, and is also surprisingly effective at improving the presentation.

Leveraging and Connecting Infrastructure

Challenges

One of the motivating visions of EarthCube was to nurture a well-connected ecosystem of interoperable data, tools, models, storage, and compute resources. While this remains the vision of the future for many project participants, it presented practical challenges. Projects relying on externally maintained components had to adapt when those components weren't finished on time, disappeared due to lack of funding, or changed their requirements for use. Furthermore, technology transfer is hard, and implementing or connecting to another's resources was sometimes more difficult than expected. Projects providing resources to others often had new expectations for sustainability, standardization and containerization that were not in the initial planning and budgeting, and that required new skills or new training. Overall, the pace of advance in the cyberinfrastructure landscape, including in commercial resources such as Amazon and Google, was "a struggle" for small projects with few personnel. Nevertheless, funded projects did have successful examples of adopting, sharing, and leveraging resources from each other. Some of these connections were not planned in the initial proposals but arose out of interactions with other projects at EarthCube meetings.

Lessons Learned

- Plan for sustainability, standardization, and containerization.
- Expect to replace your technology every few years.

COVID Impacts

For projects active during the pandemic, COVID intensified the challenges described above: engaging end users, retaining staff, communicating across disciplinary lines, coordinating a team, etc. were all made more difficult when teams could not meet in person, and many could not work as effectively from home or were struggling with additional stressors such as childcare. Projects commented that communications took more time virtually, and motivating teams was harder. This was particularly true for projects that launched during COVID, or that were heavily focused on community engagement, such as workshop-heavy Research Coordination Networks. Many projects spoke candidly about difficult struggles during this time.

> *[during COVID] everything has been slowed down so much*

As with the rest of the business world, projects pivoted to holding virtual meetings, and using other communication channels such as email and Slack more heavily. Projects learned strategies such as avoiding 8-hour virtual days, expecting more no-shows with free virtual events than in-person events, and building in more interactive components. An unexpected benefit for many was the increase in participation, and diversity of participation, for virtual vs in-person events. Several projects are now planning to keep some or most of their community activities virtual to enable participation by those who do not have the resources to travel.