EarthCube Sustainability Planning Report

FINAL PLEASE DO NOT EDIT

February 10, 2022

Report prepared by the EarthCube Sustainability Panel:

Mike Daniels (EarthCube Leadership Council), J-F de La Beaujardière (National Center for Atmospheric Research), Robert R. Downs (Columbia University), Dave Fulker (OPeNDAP), Denise J. Hills (Geological Survey of Alabama), Gwen Jacobs (University of Hawaii), Shantenu Jha (Rutgers University), Matt Jones (UC Santa Barbara), Danie Kinkade (Woods Hole Oceanographic Institution), Steve Kuehn (Concord University), Lynne Schreiber (EarthCube Office), Doug Walker (University of Kansas), Catherine Cramer (EarthCube Office)

Introduction

As EarthCube enters its 10th and final year, the program is in transition as its official term with NSF concludes and actions are undertaken to sustain and evolve the valuable elements that the Geosciences community relies upon. The purpose of this report is to describe the sustainability actions taken to date and identify actionable pathways toward the sustainability of the most valued collective EarthCube community-driven initiatives and activities organized by the EarthCube Office (ECO) and Governance. While this report does not specifically address the sustainability of the individual EarthCube funded projects, given the positive momentum of the initiatives identified below, ECO and Governance will take steps to advance these pathways before ECO activities end.

Sustainability actions to date

Gathering Community Input

In early 2021 the EarthCube Governance and Office began a series of discussions and information gathering activities with various EarthCube stakeholders on sustaining the most valued aspects of EarthCube. These activities included interviews with NSF Program Managers, community listening sessions that were held at the 2021 Annual Meeting, and Funded Project interviews. The EarthCube Sustainability Panel was then assembled to assess this feedback and make recommendations that define actionable pathways. The EarthCube community was then invited to comment on this report at the EC Townhall held at AGU in December 2021.

Earthcube Office

ECO has begun taking actions toward the sustainability of selected services that it provides to the community. These actions include: EarthCube organizational documents being archived with enhanced metadata; FAIR activities being promoted through GO FAIR; Data Help Desk activities being adopted by other science organization, e.g. AMS, EGU; Funded Projects being encouraged to host their own websites; and the EC website domain being renewed for five additional years. Plans are currently underway for writing up lessons learned, as well as producing a retrospective document for the EC community and stakeholders.

High priority sustainability items
The following items have been identified through the above process as the most valued collective aspects of EarthCube and those most worth sustaining. (*Note: these topics are in no particular order*.)

Council of Data Facilities (CDF) - Provided by the CDF Executive Committee
CDF members agreed unanimously at their General Assembly Meeting in January 2021 that a body such as the CDF should persist, but, at that time, there was not a clear agreement on how the CDF should operate in the future nor what its main activities should be. Therefore, in 2021, the CDF organized two strategic planning sessions (consisting of an XLEAP session at the EarthCube Annual Meeting and a breakout group discussion at its Summer 2021 General Assembly) to identify what CDF members consider to be the most important CDF functions that could advance the impact of data facilities on the Earth, environmental, and space sciences over the next 5 years and rank them according to priority. The highest ranked value proposition for the CDF was identified to be its role in facilitating convergence and agreement on standards, leading practices, and protocols to move the community toward more FAIR and open data, and, in particular, interoperability standards including vocabularies, APIs, and metadata schemas. The second highest priority was placed on communication and community building, encompassing both the engagement of the research community to increase awareness and use of available data resources as well as knowledge exchange among data facilities about technical and operational approaches, experiences, and leading practices. Shared infrastructure, and technology approaches and the development of strategies for long-term funding sustainability for data repositories, were also prioritized.

A second major topic in CDF's strategic planning discussion has been the future structure of the organization. CDF members weighed different options, such as a "Cluster" within the Earth Science Information Partners (ESIP) organization, a membership model similar to NASA's ESDSWG (Earth Science Data System Working Groups), and a facilities consortium. One possible option appears to be partnering with ESIP, but CDF members favor a more substantial structure than that presently afforded by ESIP Clusters, though this would require augmentation of CDF and ESIP resources.

Actionable pathway:
1. A CDF Task Force was formed in Fall 2021 and it is charged with identifying pathways, both for the CDF to define its future and to implement the necessary changes. The Task Force will deliver a report to the CDF Executive Committee by April 30, 2022.
2. CDF will continue to explore collaboration or partnership opportunities with ESIP, developing a structure that is workable and mutually beneficial.

Council of Funded Projects sustainability initiative
In January 2021 the Council of Funded Projects, as a group, shared presentations and discussed project-level sustainability models. A working group was later formed to address core

issues of project sustainability relevant to EarthCube by examining similar projects which have demonstrated success by outlasting their initial funding and period of inception. The ultimate goal of this process is producing a peer-reviewed publication focused on different models for project sustainability. The initial findings suggest that projects providing infrastructure faced periods of major uncertainty, often associated with funding, because there is no clear path for continued funding for digital infrastructure in the venues provided to scientists through governmental organizations. Each organization spent significant effort finding creative ways to fund their digital initiatives, and each organization arrived at a different business model. Talented leadership and skilled governance were key for successfully navigating crises and changes in these projects.

Actionable pathway:

3. Propose a Research Collaboration Network (RCN) to address and share the characteristics of sustainable and resilient projects, NSF's role in the formation and ongoing nature of these projects, and the role of academic communities and publishers in addressing the digital data requirements that are becoming increasingly common.

GeoCODES

ECO is continuing to support the development of GeoCODES by onboarding CDF members with the adoption of schema.org and developing additional features based on community feedback, including a data spatial search. ECO is working to implement a communications plan to support this ongoing work, which includes: 1) raising awareness of GeoCODES as a platform and a demonstration of geosciences and data standards in action; 2) demonstrating NSF investment in EarthCube as 'the sum greater than its parts' and contributing to interoperability and standards in the geosciences and beyond; 3) attracting builders and data scientists to geosciences data challenges; and 4) raising the visibility of the effort to support notebooks as scholarly objects as being part of the GeoCODES brand, as well as notebooks being a way of using GeoCODES functionality. GeoCODES has an excellent opportunity to expand to other fields and embrace and enhance cross-disciplinary work.

Actionable pathway:

4. ECO will explore avenues for future GeoCODES funding.

Notebooks as scholarly objects

EarthCube has capitalized on the potential of *executable* notebooks (limited primarily to Jupyter Notebooks but implicitly including similar entities in R-Studio and Matlab) to support learning and scholarship. Of particular note is the growing interest in notebook presentations at the EC Annual Meetings. These have become a locus of fruitful interaction between technological and geoscientific specialists in the EarthCube community. The educational potential of notebooks has been applied to learning about EarthCube capabilities per se, as well as their application to real (and often transdisciplinary) geoscience problems.

Actionable pathways:

5. Charter a committee or interest group to focus on notebooks as scholarly objects and champion items b) through f).
6. Champion effective notebook and code preservation. Exploit and advance related progress made by JetStreamII and ECO, extending it to embrace notebooks in the RStudio and Matlab formats. Exploit standard versioning platforms such as GitHub to preserve code bases beyond those already mentioned.
7. Lead and promote conference sessions at ESIP, AGU and similar meetings, in which peer-reviewed, open-source notebooks are presented. The sessions should emulate and advance those held at recent EarthCube Annual Meetings, extending them to fully embrace notebooks in the Jupyter, Matlab, and RStudio formats.
8. Extend collaborations with the Journal of Open-Source Software and AGU and forge new ones to increase opportunities for notebook publication, recognizing how long-term trust is key to an object's becoming truly scholarly. (Walker, J.D., 2021, Doing Geology in an online world: GSA Today, v. 31, no. 2, p. 4-7.)
9. Secure funding for workshops that support scholarly, standards-compliant notebook construction, informed by efforts such as The Geoscience Paper of the Future Initiative.
10. Foster research on making executable notebooks persistent (i.e., reusable) over the long term and on linking notebooks, software, data, and publications (as in OntoSoft, GeoCODES and Throughput, e.g.) to help scientists find all relevant resources.

Earth Science community to build CI-science collaborations and capacity

One of EarthCube's most notable achievements is its sustained engagement of research scientists in the development of novel cyberinfrastructure. EarthCube's community is unique in that it: 1) is composed of a mix of science and technical experts who may have otherwise not interacted, 2) has multiple funding opportunities shaped over the course of ten years to serve the needs of earth scientists, and 3) has a governance structure and central office which brings cohesion, engages early career researchers, develops broad technical solutions and best practices, promotes common infrastructure, supports individual projects, provides lightweight funding mechanisms that advance new ideas and leads and implements longer-range plans. Over the course of the program EarthCube has cultivated a community of nearly 2000 cross-domain professionals, enabling close and fruitful collaborations between researchers and data and technical professionals. The benefits have included a corpus of resources and tools that directly address geoscience research needs and infrastructure gaps and have increased capacity within the community, yielding a more data-savvy workforce.

Actionable pathways:

11. Support annual meetings in 2023 and 2024 for those Funded Projects whose end dates extend beyond the end of the EarthCube Office. These could be co-located with other meetings to help mitigate the need for funding and organizer support. Such gatherings would require a small team to coordinate and to interface between the funded projects and meeting host(s).
12. Leverage existing communities of practice (e.g., ESIP and RDA) and scientific professional societies (e.g., AGU, EGU, GSA) to continue science engagement, foster collaborations, and further build human capacity, thereby promoting cohesion and

support for common goals. These communities may be under-resourced to engage the sizable EarthCube community. Appropriate support could facilitate a smoother transition and avoid increasing demands on volunteer labor and the associated potential for burnout. Example activities which could engage with these other communities include:

- Partnering with existing interest groups, sections, and/or divisions within the above-mentioned organizations or others, such as an EarthCube Science Collaborations Cluster within ESIP or a similar Interest Group within RDA.
- Facilitating dedicated sessions and/or programming which highlight technical solutions with strong science applications; demonstrate the value proposition of technical-science collaborations; and promote the adoption of community-recommended practices (see subsequent section), such as a recurring collaborative AGU Domain Section-ESSI Town Hall or oral session.
- Continue branded sponsorship of and coordinated participation in Data Help Desks to showcase reusable EarthCube resources and to provide opportunities for data training and problem solving. This effort would depend on continued volunteer efforts and require coordination among community members.
- Facilitate invited speaking opportunities on existing technical needs at forums such as the Council of Data Facilities, SciPy meetings, etc.

13. Create mechanisms that enable community feedback to inform NSF of geoscience research CI needs and coverage gaps and to ensure that researchers are engaged when developing CI. This could be accomplished via an RCN to inform NSF on the needs of the GEO community prior to the release of new solicitations or programs or via an RFI initiated on behalf of NSF.
14. Have key EC members be available for discussions about the unique strengths of the EarthCube community with the intent of serving as a catalyst and guide for similar research efforts.

Community recommended practices, standards, data management and interoperability

EarthCube has contributed to various community practices and standards that facilitate the creation, dissemination, and use of Earth science data, including practices for improving data management and interoperability among data products and services.

Actionable pathways:

15. Continue work on the "Standards and Specifications" document summarizing EarthCube recommended practices regarding data management, formats, metadata, and interoperability in a manner that supports its community-informed evolution in a rapidly advancing cyberinfrastructure environment. First step: to ensure ongoing interest, revise the document to make it more generally applicable to Earth science data rather than to EarthCube specifically.
16. In conjunction with events, such as those mentioned in the previous section, that bring together domain scientists, data stewards, and technical experts, provide opportunities for members of the Earth science community to identify requirements for improving data management practices, standards, and interoperability.

17. Establish an online forum to support coordination among Earth science community members to collaboratively pursue the identification, development, testing, and adoption of best practices and standards that will contribute to the creation, stewardship, interoperability, and sharing of data and products, and for members to demonstrate the value of new practices, technologies, and standards for Earth science. For example: https://www.oceanbestpractices.org/.
18. Engage early career data professionals to infuse awareness of new technologies that may supplant older, previously-recommended approaches.
19. Provide guidance on authorship and proper citation of data, software, and other research artifacts, including the use of Digital Object identifiers (DOIs), and on the preparation of such artifacts for reuse by others. This will help support the adoption of community recommended practices and standards for creating, documenting, and sharing data, software, and other research outputs in a responsible manner.
20. Encourage community education surrounding the adoption of better practices in implementing Data Management Plans.

Advancing EarthCube's achievements in FAIR infrastructure

Noteworthy progress has been made through EarthCube on realizing the FAIR (Findable, Accessible, Interoperable and Reusable) principles for scientific data, conceptually extended to include software, notebooks as scholarship, and other geoscientific artifacts. Three FAIR-related EarthCube efforts merit particular note: 1) hosting the Council of Data Facilities, a potential forum for coordinating persistence and FAIRness across repositories; 2) adopting crawler-aggregated JSON-LD metadata as the foundation for GeoCODES and a basis for federating metadata from multiple repositories; and 3) elevating the potential of executable notebooks to serve as persistent scholarly objects that, in particular, enhance Reusability. To develop the essential scientific resources that researchers rely upon and that adhere to all of the FAIR principles, restructuring suitable, long-term infrastructure for basic persistence will be required.

Actionable pathway:

21. Initiate a high-level study or RCN on the need for FAIR-critical infrastructure (including responsible funding organizations as well as suitable technologies) across the entirety of the research and education enterprise.
22. Work with the existing work within ESIP (e.g. data citation guidelines, catalog of outputs, etc.) to continue discussions of this topic.

Climate-change research as an opportunity for sustained CI advancement

Future federal funding opportunities in the near term will focus on climate change research, a very fertile ground for geoscience research and cyberinfrastructure innovation for predictive modeling, forecasting, public awareness and disaster preparedness. EarthCube's efforts to improve discovery and access to geoscience data, combined with continued and expanded investment in geoscience cyberinfrastructure and geoinformatics research, will be an important part of building community resilience and mitigating climate change. While advances in CI research cannot occur without an important domain science application area, there are ongoing

funding opportunities in OAC that can be leveraged in support of these kinds of goals. Exemplars of NSF-funded software and CI frameworks that have evolved over time to support multiple communities include: Iplant, CyVerse, AGAVE , Tapis, and DesignSafe. These software frameworks are very amenable to multiple domain science applications and are ripe for collaboration with the geoscience community. Exemplar projects funded through OAC that focus on geoscience data and tools are: Hydroshare, GeoEDF and CyberGIS**.**

Services and facilities enabled by EarthCube-like research may be beyond the scope and ability of any individual institution that currently exists. There are existing entities in place that could serve as models or guidelines for supporting a sustainable geoscience cyberinfrastructure and geoinformatics research center or network. A federally-funded research and development center or distributed network (e.g., National Center for Atmospheric Research (NCAR), Long Term Ecological Research (LTER) Network) could provide the coordination necessary to continue the success of EarthCube-like efforts.

*Actionable pathways***:**

23. Form cross-agency working groups or workshops to assess common needs and resources to enable ongoing geoscience cyberinfrastructure
24. Work with ESIP, AGU and/or RDA to convene workshops to explore the role of geoscience/CI in climate change research
25. Explore NSF CI Center of Excellence program to continue EarthCube-aligned CI activities
26. Coordinate with other federal government agencies that are supporting (geo)science cyberinfrastructure efforts, such as DOE's Energy Data eXchange (EDX), NOAA's Joint Technology Transfer Initiative (JTTI), USGS's Pacific Island Ecosystems Research Center and Pacific Islands Water Science Center, NASA DAACs.
27. Explore and promote support opportunities that may be available through foundations, nonprofits, and NGOs, e.g. Simons Foundation, Nature Conservancy, Sloan Foundation, Laura and John Arnold Foundation, and the Gordon and Betty Moore Foundation. (The Enabling FAIR Data Project was supported by a grant from the Laura and John Arnold Foundation, demonstrating that funding is available from these entities for geoinformatics work.)