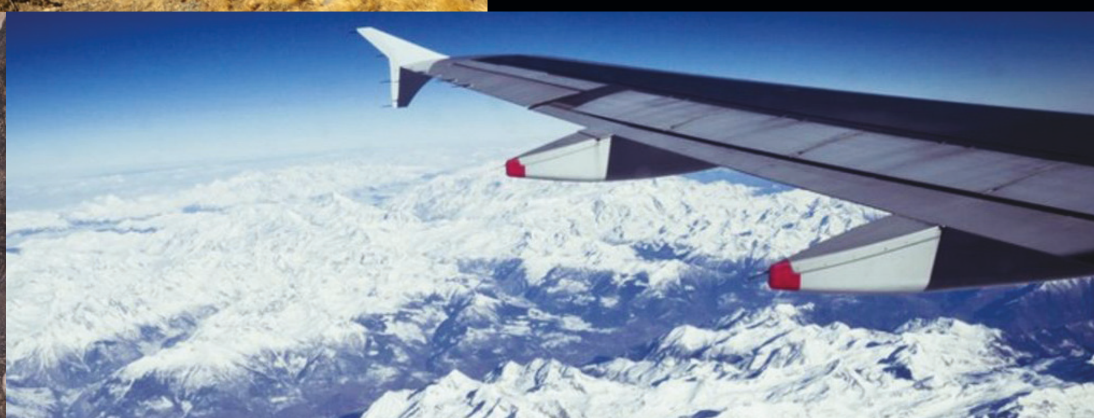
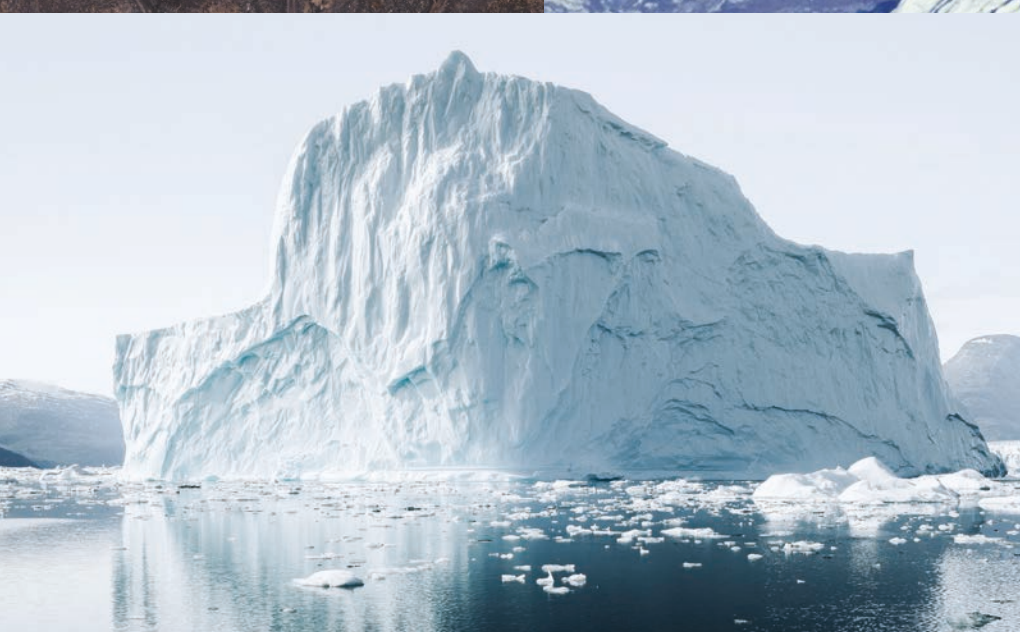
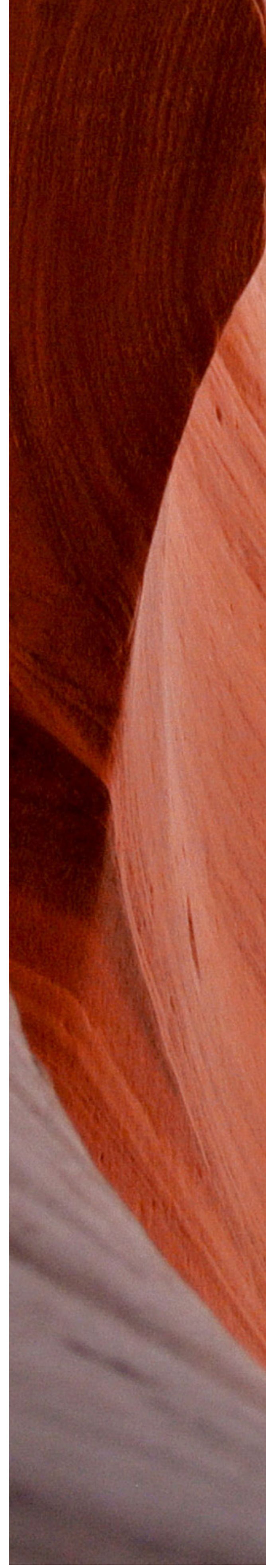
EarthCube

A RETROSPECTIVE 2011-2022







CONTENTS

EXECUTIVE SUMMARY

The National Science Foundation's Geosciences Directorate (GEO) and its Office of Cyberinfrastructure (OCI)—seeking to address the multifaceted challenges of modern data-intensive science—envisioned an enterprise, dubbed EarthCube, where low adoption thresholds and new (domain-neutral) capabilities could increase the productivity and capability of researchers working at the frontiers of Earth system science. This was to address NSF's goal to create integrated data management infrastructures across the geosciences through transformative concepts and approaches.

From 2011-2022, the EarthCube program fostered collaborations and built an agile, interdisciplinary community that spanned an array of geoscience domains. The program fostered findability, accessibility, interoperability and reusability (FAIR) principles for significant amounts of data and other resources. More than 2,000 community members shaped its design, direction and evolution, participating in governance and establishing priorities while actively building effective cyberinfrastructure (CI). These developments advanced the geoscience community's technological and computational capabilities, thus furthering EarthCube's goal of addressing complex and evolving Earth systems challenges.

Beyond its CI and geosciences advances, EarthCube reflects the successful creation, management and facilitation of a large, complex consortium. This consortium has supported and fostered innovative, and often convergent research by the members of a diverse, multi-disciplinary, creative, engaged, and productive community. Bringing together scientists and technologists to work at the intersection of geoscience and cyberinfrastructure has been a hallmark of EarthCube. While its community orientation has required time and effort, the result has been noteworthy EarthCube assets, such as GeoCODES, community standards, and notebooks as scholarly objects, that have significant value primarily due to community buy-in.

The impact of EarthCube is multifaceted, crosses domains, and serves as a model for building engaged, interdisciplinary communities. The 95 projects funded during EarthCube's decade-long existence produced demonstrable impacts too numerous to list in this summary. However, the true impact of NSF's investment is considerably greater than the sum of the project advances. A few notable program-level achievements are as follows:

The GeoCODES framework has helped integrate geoscience data into the advancing world of "linked data," setting the stage for easier discovery (e.g. via Google Search) of datasets and related tools and notebooks.

To our knowledge, the 2020 EarthCube Annual Meeting was the first scientific conference anywhere to invite submissions of executable notebooks for peer review and potential presentation. This reflects EarthCube's belief that executable notebooks (open source documents containing live code, equations, visualizations, and formatted text) can support learning and scholarship and that the authors of effective notebooks deserve recognition. Subsequent Annual Meeting notebook presentations garnered growing interest and became a locus of fruitful interaction between technology specialists, scientists, and engineers in the geoscience community and beyond. Furthermore, EarthCube's recognition of executable notebooks as scholarly artifacts has seen uptake in broader communities such as the American Geophysical Union.

EarthCube has helped to instill beneficial values across the geoscience community, especially in conjunction with cyberinfrastructure and advanced technological solutions. EarthCube has fostered a broad application of the FAIR principles for scientific data, conceptually extended to include software, notebooks as scholarship, and other geoscientific artifacts. A diverse cohort of early-career researchers have felt welcomed by the EarthCube community, and many have indicated that their involvement in EarthCube has been pivotal in fostering new collaborations and advancing their careers. The high value of community is seen across EarthCube, in fostering transdisciplinary resource sharing, interoperability, and collaboration, which in turn helps to build a data-savvy workforce across Earth system science.

While the EarthCube program has drawn to a close, components of the EarthCube Governance and Office have found pathways to continue serving the community. New NSF awards will allow GeoCODES to grow, community notebook development to continue, and FAIR education to expand in the important area of AI/Machine Learning. The vibrant Council of Data Facilities has found a new home as an Earth Science Information Partners (ESIP) Cluster. And community leaders are in discussion with NSF about continuing EarthCube Annual Meetings as a critical place for conversations across the geosciences and cyberinfrastructure domains.

The goal of this retrospective document is to tell the story of EarthCube —what inspired its creation, and its pathway to growth, the challenges encountered, and its lasting achievements.

> *"EarthCube is a niche community that is growing; without EarthCube there is no community that does this."*

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number (1928208). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

This retrospective has been written and compiled by members of the EarthCube Office and Leadership Council. November 2022

EARTHCUBE 101

The U. S. National Science Foundation's ten-year EarthCube program was a community-driven activity aimed at transforming the conduct of geosciences research and education by building a well-connected cyberinfrastructure. EarthCube created an inventive forum for its community to share and integrate data, tools, and knowledge across all geoscience disciplines in an open, transparent, and inclusive manner and to accelerate our ability to understand and predict the complex and evolving Earth system. The overarching goal of EarthCube encompassed the improvement of accessibility of geoscience data as well as the abilities to share, analyze, and visualize geoscience data and related resources.

As the funding source for EarthCube, the National Science Foundation (NSF) oversaw and provided feedback for all program activities. EarthCube Governance, comprising several operating committees, made decisions and set policies to lead EarthCube and worked closely with the EarthCube Office, which coordinated the overall EarthCube ecosystem. Research was conducted by the Funded Projects, which, together with other interested parties, formed the robust community of researchers and cyberinfrastructure builders.

EARTHCUBE COMMUNITY

The EarthCube community is a diverse, global and connected group of researchers and cyberinfrastructure and data science experts who work together to: improve how data are collected, accessed, analyzed, visualized, shared and archived; participate in interdisciplinary and collaborative research; and support open and inclusive science.

EARTHCUBE FUNDED PROJECTS

For eight years, from 2013-2021, NSF supported 95 EarthCube-funded projects[1] with awards across multiple solicitations that evolved over time. This evolution was based on the response to progress made via the EarthCube effort and the scientific and cultural needs of the geosciences community at the time of award. The solicitations were sponsored through a partnership between the NSF Directorate for Geosciences (GEO) and the Office of Advanced Cyberinfrastructure (OAC) within the Directorate for Computer and Information Science and Engineering (CISE).

EARTHCUBE GOVERNANCE

The EarthCube governing structure[2] evolved over the life of the program. The final structure consisted of six elected bodies:

- Leadership Council (LC) - the elected voice of the EarthCube community, establishing strategic direction for the program and making decisions critical to success.
- Science and Engagement Team (SET) - a connection between the academic geoscience and technology communities in EarthCube that linked EarthCube activities to relevant organizations and initiatives.
- Technology and Architecture Committee (TAC) - a forum for maintaining an architecturally-oriented overview of EarthCube's technological capabilities.
- Council of Funded Projects (CFP) - a forum for project personnel to interact, discover and work together on common needs.
- Council of Data Facilities (CDF) - a federation of existing and emerging geoscience data facilities exchanging experiences and promoting standards and best practices in the organization and operation of data facilities.
- Nominations Committee - a body that oversees the nomination of EarthCube community members to various Governance roles.

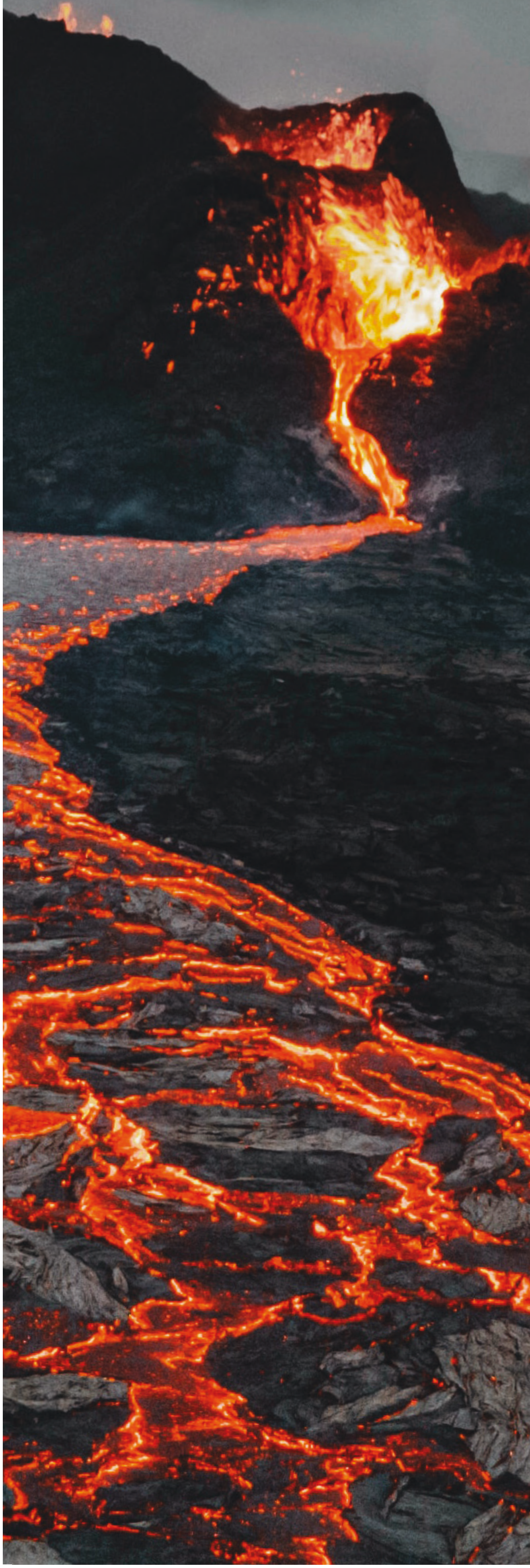
EARTHCUBE OFFICE

There have been three iterations of the EarthCube office, described below, all of which shared common attributes. The office supported the application of community standards for data sharing, publication, and reuse. In addition, the office facilitated the uptake of tools and data sources and assistance with the community's research, education and outreach practices to better enable the program's efforts. The office also translated governance and NSF directives into actions and outcomes, delivering a variety of high-value community engagement activities and events in support of research and education.

The heart of EarthCube is people. *We put people first and do our best to recognize, appreciate, and* ***respect the diversity*** *of our global contributors.* ***EarthCube welcomes*** *contributions from* ***everyone*** *who shares our interests and wants to contribute in a healthy and constructive manner within our community.*

"EarthCube was timely and played an important role in getting the right people interacting with each other."

NSF seeks transformative concepts and approaches to create integrated data management infrastructures across the Geosciences. In a new partnership, the Geosciences Directorate (GEO) and the Office of Cyberinfrastructure (OCI) recognize the multifaceted challenges of modern, data-intensive science and education and envision an environment where low adoption thresholds and new capabilities act together to greatly increase the productivity and capability of researchers and educators working at the frontiers of Earth system science. [from 2011 Dear Colleague Letter]



HISTORY

The NSF EarthCube program was created in 2011 to advance cyberinfrastructure in support of geoscience research. Leading up to the formation of EarthCube was the perceived need to engage the geosciences community with the complexities of data management, which prompted the 2011 NSF Geo-Data Informatics: Exploring the Life Cycle, Citation and Integration of Geo-Data Workshop[3]. Building on the outcomes of this workshop, EarthCube was established that same year via collaborative partnership between NSF's Directorate for Geosciences (GEO) and the Office of Advanced Cyberinfrastructure (ACI). An NSF Dear Colleague Letter[4] was released in June 2011 announcing the partnership and initial goals for EarthCube.

Several webinars followed, and an additional document, EarthCube Guidance for the Community[5], gave more detailed guidance. These announcements launched the first conversations about the future of EarthCube. Over the next two years, a series of webinars, community meetings (Charrettes) and White Paper and Roadmap solicitations provided a forum for potential participants to propose what EarthCube should look like in terms of science requirements, technology solutions, designs and governance. The roadmaps were the culmination of months of collaborative work and identified initial stakeholders and cyberinfrastructure components. They provided a cross spectrum of ideas and concepts from the Earth, computer and data science and other stakeholder communities on key elements to build EarthCube.

End user workshop: Envisioning Success- A Workshop for Next Generation EarthCube Scholars and Scientists; Oct 16-17, 2012; Washington, DC

Beginning in Summer 2012, NSF funded a series of 24 domain science end-user workshops[6] with a total of approximately 1500 participants, targeting a broad spectrum of Earth, atmosphere, ocean and allied senior, mid- and early career scientists. The purpose was to allow these communities to articulate their cyberinfrastructure needs and goals, particularly in relation to data accessibility, both within their disciplines and from other fields. These workshops offered a place for participants to network and start talking together about the uses of data, and offered pathways for engagement.

EarthCube Test Enterprise Governance (ECTEG), 2013-2016, Arizona Geological Survey

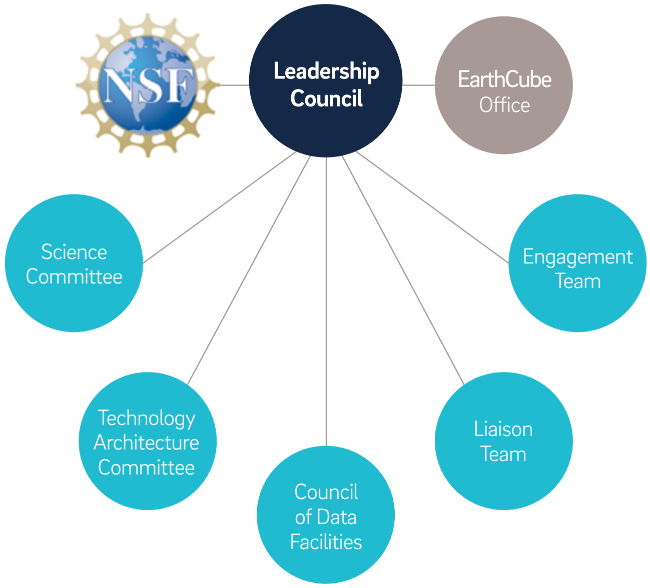
ECTEG established a voluntary, community-based organizational framework through which an EarthCube cyberinfrastructure could be developed.

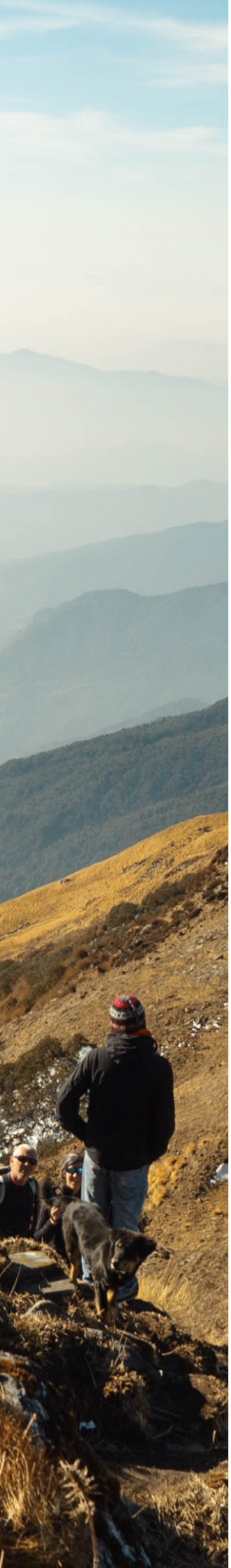
EarthCube Test Enterprise Governance (ECTEG), led by the Arizona Geological Survey, was the first EarthCube office, and was responsible for the development of an EarthCube Governance framework. This project supported community driven efforts to implement an agile model to identify, test and evaluate governance models to manage the development of an efficient, community-based geosciences cyberinfrastructure. In addition, the project sought broad engagement and participation of EarthCube stakeholders to define and assess governance models while seeking evaluation and cross-checks from advisory and management committees as well as two evaluation teams. (See EarthCube Past, Present and Future[7])

ECTEG OUTCOMES

EarthCube Governance

- The EarthCube charter and governance groups' charters were developed and approved by the community.
- The community governance nominating and election processes were designed, established and implemented.
- The governance bodies comprising the governance framework were launched, populated and implemented. They comprised: the Leadership Council, Science Standing Committee, Technology and Architecture Standing Committee, Liaison Team, Engagement Team, and Council of Data Facilities.





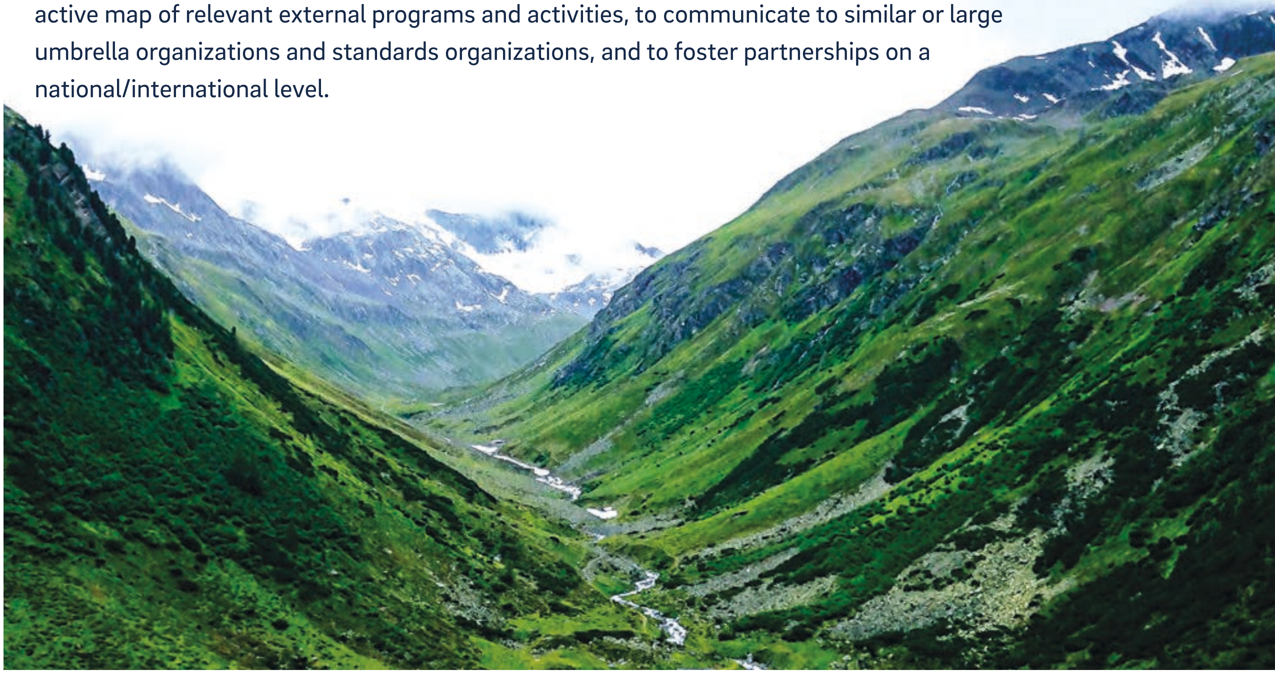
Cyberinfrastructure Building

- ECTEG helped establish and execute regular virtual meetings, workshops and face-to-face meetings for the EarthCube community to coordinate, organize, set priorities, and define goals, standards and boundaries for geosciences cyberinfrastructure development.
- Progress toward community convergence on a viable reference architecture was made by designing, establishing and undertaking processes and procedures for gathering, sharing and validating broad community input alongside points of disagreement and consensus.
- Cyberinfrastructure concept documents, including a Science Strategic Plan, a Technology Strategic Plan, a Strategic Roadmap and Implementation Plan, and an Architecture Development Plan were developed and approved by the community.

In March 2016, an EarthCube Architecture Workshop recommended the "creation of a Geoscientist's Knowledge Workbench, a flexible and collaborative environment where geoscientists and cyberinfrastructure experts discover, access and utilize an array of data and services that foster cross-domain advances in information and knowledge. The Workbench, as a gathering and collaboration space, will likely draw (early career) scientists facing unusually difficult, cross-domain research challenges."[8]

Community Building

- ECTEG facilitated and supported opportunities for EarthCube working groups, governance groups and funded project teams to engage with one another and with the broader community to better understand requirements for a geosciences cyberinfrastructure, define priorities for development and share milestones as well as achievements toward fulfillment of the program's mission and goals.
- ECTEG participated in conferences and exhibits of primary science domains and disseminated updates through newsletters, a website and social media.
- The EarthCube community began recognizing individuals in 2015 who contributed significantly and measurably to advancing the goals of EarthCube with the Community Service and Leadership and Legacy Awards.
- EarthCube's Liaison Team developed a 'map of the landscape' (*dusk.geo.orst.edu/ec-story*), an interactive map of relevant external programs and activities, to communicate to similar or large umbrella organizations and standards organizations, and to foster partnerships on a national/international level.



REVERSE SITE VISIT

EarthCube's progress over the first four years was reviewed in early 2016 by a Reverse Site Visit (RSV). The RSV Advisory Committee summarized its findings and provided key recommendations for improvement and the future evolution of the program.

> **The Advisory Committee notes that the Leadership Council has demonstrated an exceptional level of volunteer commitment to the success of NSF's EarthCube program. We recognize the cultural, social and scientific challenges faced by the Leadership Council in attempting to promote a grassroots movement enabling the geoscience community to move towards a new data-enabled cross-disciplinary research paradigm. Over the past several years, steps have been taken to establish a governance model, populate the EarthCube committees, and begin the work for EarthCube. The large number of community workshops and efforts to build consensus on community goals has clearly strained available intellectual and time resources. In weighing these struggles against accomplishments from the first five years, it is evident that significant structural changes are needed to create a thriving sustainable program within the next five years. *(from RSV Advisory Committee report[9])***

The RSV Advisory Committee identified three interrelated problems impeding progress in the Earth-Cube initiative, which were subsequently addressed:

- Persistent lack of a clear definition of "EarthCube."
- Reliance on a largely voluntary Leadership Council and a fully voluntary governance structure.
- Lack of interconnection between Leadership Council activities, proposal solicitations and assessment of outcomes from funded projects. In essence, the Leadership Council and the EarthCube principal investigators appear to be operating "open loop," except where interactions are mediated by the NSF.

The advice was well received, and these concerns were addressed in a formal response document[10] to the Advisory Committee. Over the next several years, the following structural changes were made to address these concerns, and enacted by the EarthCube office as well as governance:

- Expanded outreach and communication helped to define EarthCube goals and priorities. Examples include several updates to the website, increased social media presence, newsletters, Monday updates, blogs, Data Help Desk, presentations/discussions at governance meetings, communications with funded projects.
- Reliance on office staff to implement governance-approved decisions, policies, and programs helped to reduce governance volunteer fatigue.
- The Leadership Council gave yearly input to NSF regarding EarthCube solicitations and funded project requirements, expectations, standards and specifications.

EarthCube Science Support Office (ESSO), 2016-2019, UCAR

ESSO provided logistical support to the EarthCube community and its governance groups in order to stimulate and manage community dialog and led EarthCube to converge on a common cyberinfrastructure approach for data discovery.

Under the leadership of the University Corporation for Atmospheric Research (UCAR) and in partnership with the EarthCube community, its governance, and other stakeholders and collaborators, the EarthCube Science Support Office (ESSO) provided logistical support for EarthCube. ESSO provided necessary information, supported infrastructure, and led the development of services to ensure that EarthCube's portfolio of governance programs and projects were effectively and efficiently directed, managed, and delivered. ESSO's responsibilities included communication and outreach, community engagement, management of subawards, and the advancement of EarthCube's cyberinfrastructure.

ESSO OUTCOMES

- ESSO supported the vast array of EarthCube's permanent and temporary committees and teams. helping to strengthen and further develop the community's efforts while also providing administrative support to the EarthCube leadership.
- ESSO successfully delivered, deployed, and maintained the cloud-based software services, registries, and user interfaces for the core data and software resource registry, discovery, and access needs for the geoscience community utilizing web standards for interoperability. ESSO also supported the solicitation, award, and monitoring of many subawards.
- ESSO's efforts to enhance EarthCube's reputation in the community by expanding outreach content and highlighting funded projects were quite successful. ESSO worked to increase community engagement through improvements to the EarthCube presence on social media, dissemination of monthly newsletters, and distribution of the weekly Monday EarthCube Update.
- ESSO supported and developed three EarthCube Annual Meetings (formerly All Hands Meetings), each engaging over 150 attendees who spent three days discussing progress and next steps. Positive feedback was consistent and strongly supportive of the work carried out by the office at these events.
- ESSO planned, facilitated, and prepared exhibit spaces and presented talks at GSA, AGU, ASLO, AMS, ESIP, and EGU. EarthCube materials were circulated at all conferences, as well as informational literature and electronic presentations.

THE BEGINNINGS OF

GeoCODES

ESSO first deployed GeoCODES as a resource to highlight improvements for the geoscience research and education community through the adoption of standardizations around metadata. In 2016, the EarthCube Council of Data Facilities (CDF), the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS), and the Registry of Research Data Repositories (re3data) established the Registry Working Group (RWG) to recommend guidelines for a common and machine-readable method for sharing information about organizations and their data holdings. In 2017, the EarthCube Architecture Refinement Workshop (ARW)[11] was held at UCAR and its results identified three main areas for alignment with FAIR principles including resource registration, discovery, and access. In 2019, the EarthCube Leadership Council published a position paper in support of FAIR principles[12]. The ARW also generated input for technology inventories, interoperability specifications, and a project management plan for moving forward with the recommended guidelines from the CDF RWG.

During this same time period, the Technology and Architecture Committee (TAC) was created to oversee the technology and architecture development of EarthCube to assure that EarthCube infrastructure encompassed community-driven support for standards for interoperability, and incorporated advanced technologies to become a commonly used capability to support scientists on their research efforts. The TAC's efforts toward standards and best practices documentation eventually led to the Recommended Standards and Specifications for EarthCube Projects[13]. This document was developed to remain useful well beyond EarthCube funding. Further, the TAC was a leading advocate for the interoperability of GeoCODES as well as EarthCube's notebook-related efforts. These are just some examples of the many TAC-led activities that helped create the long-lasting positive impact of EarthCube.

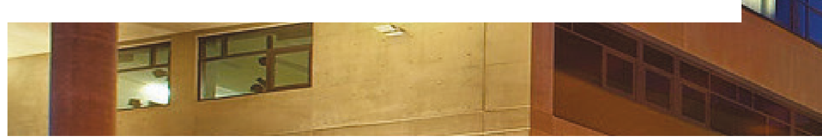
In late 2017 the seed project known as "Project 418" (P418) was launched to build on EarthCube's support for data discovery among NSF-funded data repository facilities.

Through P418, EarthCube directly engaged a set of 10 NSF data facilities, scoping nearly 48,000 datasets, to demonstrate and highlight to the repository community how they can employ and be part of this exciting development.

P418's key focus was to provide enhanced capabilities around domain-specific needs of the various Earth science communities by extending the vocabulary aspect of Schema.org and extensions to a set of NSF data facilities. Another focus was on describing dataset resources and evaluating the use of this structured metadata to address discovery.

P418's open and standards-based approach directly aligned with and enabled NSF data facilities to be indexed and discoverable by the Google Data Search Tool. Additionally, this same approach was used to enable commercial, academic, and non-profit groups to leverage the same resources to provide alternative, and even enhanced, search options for various communities. The P418 team was able to accomplish this by maintaining close communication with Google and the Earth science community during its development to ensure alignment of recommendations and to leverage existing work and experience. In 2018-2019, funding was approved to expand P418 and include the Resource Registry. These combined efforts to better enable cross-domain discovery of, and access to, geoscience data and research tools became GeoCODES.

Building on the ESSO effort, the next EarthCube Office (ECO) focused on creating a seamless data pipeline that allowed users to discover data across the geoscience community. The updated version of GeoCODES was developed with regular input from EarthCube governance, including feedback from user stories and the GeoCODES roadmap. The current version of GeoCODES (*geocodes.earthcube.org*) not only features a new user interface, but also brings together datasets and tools, identified from resources such as the resource registry, highlighting within query results the tools that can be used with returned datasets.



EarthCube Office (ECO), 2019-2022, San Diego Supercomputer Center UC San Diego

The EarthCube Office (ECO) refined the three major office functions: facilitating governance, driving data discovery and access—principally through GeoCODES—and enhancing science engagement.

The EarthCube Office (ECO) was proactive and supportive of the EarthCube community through a distributed organization that offered a wide variety of skills and expertise including: outreach, program management, metrics, FAIR data training, web and user interface design, science writing and scientific software development, as well as expertise in several geosciences domains.

RESTRUCTURED GOVERNANCE

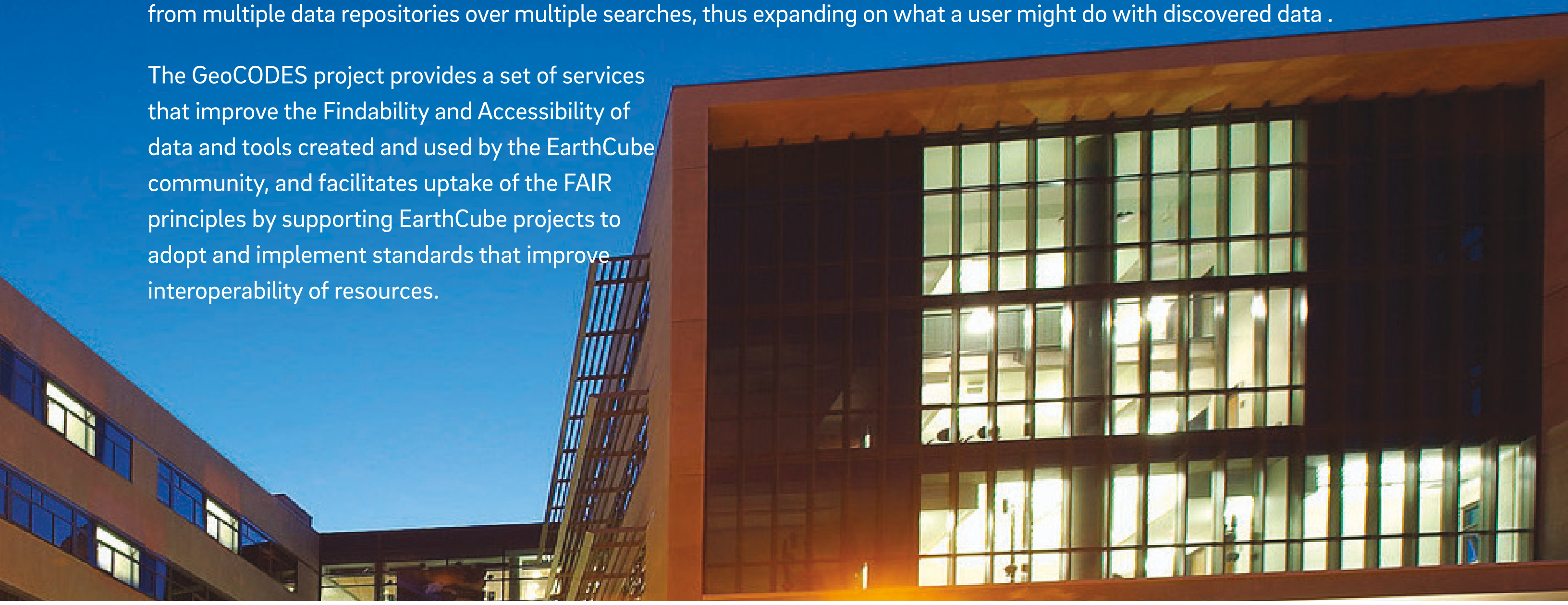
To align the evolving EarthCube needs and priorities, the Leadership Council and ECO restructured governance by fostering the recently formed Council of Funded Projects (CFP), merging the Science Committee, Liaison Team and Engagement Team into the Science and Engagement Team (SET), and designating one Leadership Council at-large position to be held by an early career researcher. ECO helped governance groups focus on practical goals and 'getting things done' over prolonged discussion.

EVOLVING GEOCODES

The ECO Tech Team revised the GeoCODES codebase and added an enhanced and updated user interface that allows for searching across CDF assets, and also produced a new set of use cases to ground development in desired functionality. The Team also engaged in numerous activities to support GeoCODES aims, including consulting with the CDF shared infrastructure initiative and other EarthCube members on the implementation of Science-on-Schema; testing GeoCODES extensibility through implementation of multiple data management frameworks; and participating in many outreach activities.

A clear priority of the Tech Team was ensuring the ability to record and disseminate developed tools and resources to be used beyond the EarthCube community. This mindset was the motivation for developing the resource registry. Aspects of the resource registry usability were addressed by incorporating tools with internal GeoCODES data, plus enabling the search interface to search the registered tools and, if possible, associate them with relevant datasets. The annual call for notebooks included a peer review process and incentivized the community to document and share their own tools in a manner that others could then use, in addition to providing scholarly output that could be cited. An "open notebook" was added to many queried results so that further work could more easily be done, and the capability was added to create collections of data made up from multiple data repositories over multiple searches, thus expanding on what a user might do with discovered data .

The GeoCODES project provides a set of services that improve the Findability and Accessibility of data and tools created and used by the EarthCube community, and facilitates uptake of the FAIR principles by supporting EarthCube projects to adopt and implement standards that improve interoperability of resources.



SCIENCE ENGAGEMENT ACTIVITIES

- ECO developed and disseminated EarthCube Community Best Practices including: Software Citation Guidelines (*kmaull-ucar.github.io/ec-sw-citation*) and Approachable FAIR Resources (*www.earthcube.org/fair-training-materials*).
- A novel element of the 2020, 2021, and 2022 EarthCube Annual Meetings was a call for digital notebooks, which led to 44 peer-reviewed Jupyter notebooks (*www.earthcube.org/notebooks*) that encompassed an array of geoscience data tools, software, services, and libraries. Each notebook was reviewed by scholars within the geoscience and cyberinfrastructure community.
- ECO created a Working Group on Diversity, Equity, and Inclusion in the Geosciences (DEI Working Group) in June 2020 to ensure that DEI concerns were fairly addressed, provided educational activities relating to DEI, ensured EC resources were accessible, and created safe and inclusive meetings. A primary outcome was the creation of Community Participation Guidelines[14] and a contract with a third party reporting service to ensure accountability of reported violations.
- To improve the overall data and informatics competency in the geosciences and to raise awareness of EarthCube project resources and opportunities for participation in the wider geosciences community, ECO supported in-person and virtual Data Help Desks, which were held at meetings of AGU, EGU, GSA, and AMS.
- The EarthCube website featured science "vignettes" (*www.earthcube.org/science*), which provided a glimpse of many research accomplishments achieved by the EarthCube community. ECO's science writer worked to increase the impact of the stories by sending select stories to NSF's Office of Legislative and Public Affairs, local institutions, and partner organizations such as the NSF Big Data Hubs.
- ECO gathered materials produced over the life of the program and created a collection[15] that will be stored at the UCSD Library. This includes organizational and governance documents, white papers, presentations, and reports produced by the Leadership Council and related Governance committee working groups and ECO staff.
- Interviews with current and completed funded projects assessed technical and outreach needs and gave projects an open-ended opportunity to raise topics. Later interviews asked community members about the value and impact of EarthCube and priorities for sustainability, primary challenges and lessons learned - specifically about DEI. In-depth results were included in two reports: Interviews with EarthCube Funded Projects: Perspectives on EarthCube's Impact and Sustainability Priorities[16] and EarthCube Project Challenges and Lessons Learned.[17]
- EarthCube hosted the first US Metadata for Machines (M4M) funded by NSF's Public Access program.

SUSTAINABILITY

In response to recommendations from the NSF-selected Midterm Review panel, ECO and Governance created a sustainability panel and developed key pathways[18] to sustain the most valued aspects of EarthCube.

- As a complementary effort to the EarthCube Sustainability Panel Report, the CDF Sustainability Task Force was constituted to explore options and possible pathways for sustaining the CDF once EarthCube support ceases, and produced an additional report[19]. The CDF has successfully transitioned to an ESIP Cluster.
- The EarthCube Governance sponsored a study of Earth science infrastructure project sustainability, led by the Council of Funded Projects, and produced a report titled Project Report: Sustainability models for integrated digital Earth Science[20].
- The Data Help Desk (DHD) connects researchers with data experts, with the ultimate goal of enhancing research and making data and software more open and FAIR. To help scale this effort to meet growing demands for more Data Help Desk events and to enable our partners to take the lead, two How-To documents were developed. One of these highlighted planning tips for in-person Data Help Desks[21], while the other focuses in great detail on planning virtual Data Help Desks[22]. AGU plans to include the DHD at their upcoming meeting, supported by guidance from ECO, as does the European Geosciences Union (EGU), with support from longtime partner ESIP.
- An additional ongoing collaboration with AGU included the development of pilot solutions to elevate notebooks as peer-reviewed publications. Documentation has been completed of work produced from the EarthCube effort, including the adoption of Science-on-Schema, enabling resources such as GeoCODES and the leveraging of notebooks as scholarly publications.
- The FAIR training materials developed by the EarthCube Office have been deposited into the EarthCube Materials Collection at the UCSD Library. The accompanying videos remain in the EarthCube video channel. EC FAIR materials and activities continue to be promoted by the GO FAIR US office across relevant events and in communication outlets such as the newsletter.

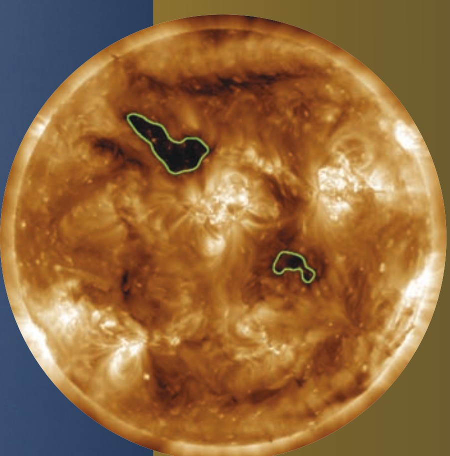
SUCCESS STORIES

EarthCube-Supported "ICEBERG" Team Paves Way for Novel Polar Geosciences Research Methods

Human-led surveys of polar animals such as penguins and seals are not only expensive, but often dangerous. Artificial intelligence techniques using satellite imagery were studied as an alternate survey option by a group of EarthCube-supported geoscientists at Stony Brook University in New York. The researchers found that computers can learn to survey polar animals with similar results as human image review, with a promising level of accuracy appropriate for many types of surveys.

MELODIES for MUSICA: A modular framework to compare model results and observations of atmospheric chemistry

Our ability to predict air quality and to understand chemistry-climate interactions depends on a comprehensive understanding of atmospheric composition, developed through the comparison of observations and models. The MELODIES (Model EvaLuation using Observations, DIagnostics and Experiments Software) framework is being developed as an extension of the Model and ObservatioN Evaluation Toolkit (MONET) in collaboration with scientists in the Air Resources Laboratory, Chemical Sciences Laboratory, and Global Systems Laboratory at NOAA. MELODIES helps determine short-comings and uncertainties in models, assess new model developments, and identify where and what type of new observations are needed to improve our understanding of atmospheric composition and processes. This is accomplished through the design of a modular framework that integrates diverse atmospheric chemistry observational datasets with numerical model results for the evaluation of air quality predictions. Additionally, by making observational datasets and corresponding 3D chemistry model results more accessible, a larger community - including students - are engaged in atmospheric composition research.



EarthCube-funded Astrophysicists Use AI for Ten Year Study Regarding our Sun

Researchers from the New Jersey Institute of Technology, Moscow University and the Kislovodsk Solar Station published results of their EarthCube-funded solar corona research in The Astrophysical Journal. Entitled Machine-learning Approach to Identification of Coronal Holes in Solar Disk Images and Synoptic Maps, the study examined and analyzed solar synoptic maps for 2010 through 2020. The study allows researchers to better understand larger coronal holes, which often result in hefty solar wind speeds that can lead to adverse space weather conditions that negatively impact satellite and power distribution grids as well as other infrastructure in our own atmosphere.

EarthCube-funded Pangeo Scientists Publish Research in Frontiers in Climate Journal

Analysis-ready, cloud-optimized (ARCO) datasets unlock the power of cloud computing at scale for scientists tackling society's most urgent questions. The EarthCube-funded Pangeo Forge team published about their ARCO research in Frontiers in Climate. Their paper, entitled Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production explained how ARCO data production is a notoriously difficult task, demanding specialized scientific as well as computational expertise, a fact which has historically limited its production. And, how Pangeo Forge presents a foundation for the practice of reproducible, cloud-native, big-data ocean, weather, and climate science without relying on proprietary or cloud-vendor-specific tooling.

EarthCube Project Creates Data Broker to Share Information about Novel Volcano in Tanzania

Considered young in comparison to other volcanoes, Ol Doinyo Lengai has been viewed as a petrological mystery since its first recorded eruption in the late 1800s. While typical terrestrial magma consists of silicon and oxygen, Ol Doinyo Lengai erupts carbonatite magma. Monitoring this volcano's inflation and deflation in real-time has been one of the case-studies for the EarthCube project CHORDS (Cloud-Hosted Real-time Data Services for the Geosciences) for several years. Scientists from Virginia Tech, Ardhi University in Tanzania, and the Korea Institute for Geosciences and Mineral Resources (KIGAM) researchers set up the monitoring project to better understand the East African Rift and the hazards of the volcano.Their data broker is coined TZVOLCANO, and has been designed to collect and share the sensor information with the broader scientific community.

The Geoscience Paper of the Future (GPF)

GPF is an initiative developed by the EarthCube OntoSoft project to encourage geoscientists to publish papers together with the associated digital products of their research, including documentation of datasets, software and provenance and workflow for each figure or result. This practice will make science more open, promote fair credit of scientific contributions, and facilitate reproducibility.

Oceanographer Uses Systems Argo and GO-SHIP to Track Ocean Temperature

To examine the intricate changes in ocean temperature, EarthCube-funded researchers from Scripps Institution of Oceanography and CU Boulder created a data portal to search for co-located Argo and GO-SHIP data. They used this to measure how drastically ocean temperature is changing, and how to extend this novel EarthCube tool to manage and share data with worldwide climate scientists.

eODP: EarthCube-Funded Paleobiology Project Makes Progress

For more than 50 years, research vessels have been collecting data from worldwide oceans - providing insight into both present and past ecosystems. While many repositories exist to house these data, it has proven challenging for some scientists to access pertinent information related to their specific research endeavors.

An EarthCube-funded international team recently developed a unified database to help with this challenge. eODP (Extending Ocean Drilling Pursuits) is a collaborative repository encompassing information from multiple sources - both large and small.

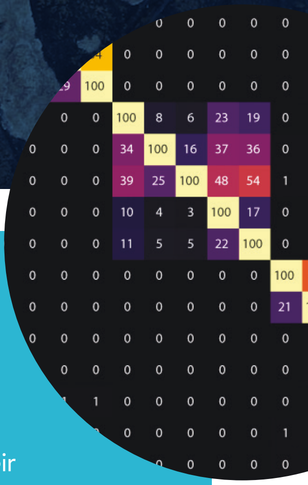
EarthCube's StraboSpot Develops Digital Tools for Geosciences Educators

In March 2020, EarthCube's StraboSpot team realized that their summer field camps were unlikely to occur as the world shut down due to the COVID-19 pandemic. The StraboSpot team worked with multiple community members to offer a variety of virtual field experiences, including repackaging and utilizing StraboSpot for online learning.

"Bringing together science and cyber is a unique EarthCube thing, and this is a hard thing to do."

EarthCube's PBOT Enables Scientists and Educators Access to Paleobotany Data

Detailed information about ancient vegetation has long been a mystery, and often when it is discovered, accessing the data is difficult - if even possible. An EarthCube project team developed a unique web client and database coined PBOT, short for Paleobotany Database, that aims to provide both scientists and educators easy access to these data. Led by the University of Wyoming, PBOT has been designed to change future research methods for paleobotanists as well as ways in which educators access and share data about ancient fossils.



Ocean Protein Portal Updates METATRYP to Encompass COVID-19 Peptides Data

EarthCube project researchers from Woods Hole Oceanographic Institution published their study findings related to SARS-CoV-2 coronavirus peptides in the Journal of Proteome Research. Specifically, the scientists discussed how SARS-CoV-2 has the most shared tryptic peptides with its closest bat precursor virus and, while the COVID-19 strain has some shared peptides with SARS-CoV-1, it is very different from the "common influenza".

Flyover Country makes geoscience data accessible to anyone by gathering information from scientific databases to highlight geological points of interest such as lakes, mountains, glaciers, and fossil locations. The app was primarily developed by a geology student at University of Minnesota.

QGreenland, EarthCube Initiative Develops Tool to Study Greenland and its Melting Ice Sheet

Greenland, the world's largest island, has long been known as a sparsely populated, ice-capped home for polar bears, reindeer and Arctic foxes. The Greenland Ice Sheet covers more than 80 percent of the island – stretching some 1500 miles long by 460 miles wide. However, conditions in Greenland are changing rapidly. According to NOAA's Arctic Report Card, last year's ice melt was similar to the alarming ice loss of 2012, and the ice sheet has lost ice every year for the last two decades.

To better understand the issues surrounding Greenland's declining ice sheet and how these changes are connected with land, people, plants, and animals, EarthCube has funded a team of scientists from the NSIDC ranging from ecologists and geologists to climatologists and software developers to create QGreenland, a free mapping tool to support interdisciplinary Greenland-focused research, teaching, decision making, and collaboration.

Researchers involved in the project are combining key datasets into a unified, all-in-one GIS analysis and visualization environment for offline and online use. A beta version has been released for testing and is available for download on the QGreenland Explore webpage: qgreenland.org.

IMPACT

The Community

BRIDGING GEOSCIENCE AND DATA SCIENCE

EarthCube's community is a unique blend of science and technical experts. Its multidisciplinary nature allowed geoinformatics projects to break out of domain silos to learn from projects with similar goals and a broad range of geoscience applications. Many of EarthCube's funded projects were essentially geodata agnostic and the solutions developed are applicable in many science areas. One of EarthCube's most notable achievements was the sustained engagement of research scientists in cyberinfrastructure development, guiding direction and facilitating awareness and uptake by geoscientists. EarthCube created support and motivation for researchers and builders to work together, and opportunities for them to build collaborations, e.g. at Annual Meetings. This community required substantial time to develop, and the EarthCube membership has emphatically called for it to continue.

FOSTERING NEW COLLABORATIONS

Across interviews with all funded projects, a clear consensus emerged that the highest perceived value of EarthCube was the opportunity for and products resulting from networking and community building. These attributes brought concrete value to participants through 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 and, for some, career-changing experience. Funded project participants saw the EarthCube community as unique in the degree to which it includes both technologists and geoscientists, and was multidisciplinary, as well as "friendly" and "nurturing." This sense of community has had concrete impacts on EarthCube participants: a poll given at the 2022 EarthCube Annual meeting showed that 61% of respondents had met a collaborator through EarthCube activities and, of those, 54% had formed a collaboration with an early career researcher.[16]

RESEARCH COORDINATION NETWORKS

During the first half of EarthCube, NSF funded several research coordination networks (RCN). These mechanisms, used throughout NSF, allowed for the formation of new networks around emerging scientific topics. These community activities broadened and deepened network connections, especially allowing for the inclusion of cyberinfrastructure staff to work alongside geo and computer scientists.

"It is a long, long bridge between technology and science."

> *It is valuable to see how other people are approaching similar problems. EarthCube, because it was cross-disciplinary, was particularly good at this.*

INCREASED OPPORTUNITIES FOR DIVERSITY IN GEOSCIENCES

EarthCube specifically strove to ensure that participants on committees and panels represented the diversity of the community. In addition, EarthCube events were open to all, and notice of upcoming events were disseminated broadly. Annual Meeting content reflected an emphasis on the community's diversity, such as the session at the EarthCube 2021 Annual Meeting, "Be FAIR and CARE: The CARE Principles for Indigenous Data Governance", addressing the crucial role of data in advancing innovation, governance, and self-determination among Indigenous Peoples.

During EarthCube, early career geoscientists were frequently invited to speak at conferences, and they were offered travel support to EarthCube Annual Meetings and scholarships to cover travel and registration at other conferences. These opportunities allowed those without deep technical expertise to engage with community members about the technical details of EarthCube projects. In response to the decrease in speaking opportunities for early career participants due to the pandemic, the majority of keynote speakers invited were early career researchers. Other activities included an early career lunch at the Annual Meeting to encourage greater networking of students and early career researchers with prominent geoscientists.

WORKING GROUPS

Early on, EarthCube leadership recognized the need to create rapid, low-overhead funding opportunities for small projects that were developing from the community. These funds were generally under $25K and followed a proposal submission process that was vetted and approved by the EarthCube Leadership Council. Working Group initiatives such as the Visiting Scientist initiative allowed new researchers to pair up with staff from funded projects to expand the reach and scope of EarthCube cyberinfrastructure in the geosciences. Other examples of Working Group outputs include creating the P418 project (a precursor to GeoCODES), offering travel grants for early career researchers, developing best practices and guidelines for EarthCube projects, providing support for notebook developers, refining the EarthCube architecture, and addressing the sustainability of EarthCube efforts beyond the lifetime of the program.

COUNCIL OF DATA FACILITIES

Identifying the need for more robust and sustainable cross-repository search facilities, the CDF launched Project P418, which eventually grew into GeoCODES, now a major discovery hub. The active and engaged CDF community has transitioned to an ESIP Cluster, and will live on as a legacy of EarthCube.

> “Geoinformatics moved forward as a field over the past decade but EC was a part of shining a spotlight on that area.”

CyberInfrastructure & Capacity

SOFTWARE, DATA, MODELS, NOTEBOOKS, SERVICES, APPS AND COMMUNITY RESOURCES

The projects funded during the almost 10-year NSF Earth-Cube program represent a substantial component of the recent geoscience cyberinfrastructure (CI) development in the U.S. Collectively, the set of EarthCube-funded projects provided deep and broad knowledge about effective strategies and critical considerations for geosciences CI from which future projects can benefit. EarthCube projects were uniquely multidisciplinary in that many of the developed tools spanned science areas. The CI experts often looked at data agnostically, and EarthCube-funded projects formed a network by interacting and sharing opportunities as well as new techniques. EarthCube projects have also produced a wide variety of products (*www.earthcube.org/products*) stemming from ideas that emerged from this social network. These include notebooks as scholarly artifacts (more below), use of open source software, shared codes, use of modern software development stacks, and common software repositories, among many others.

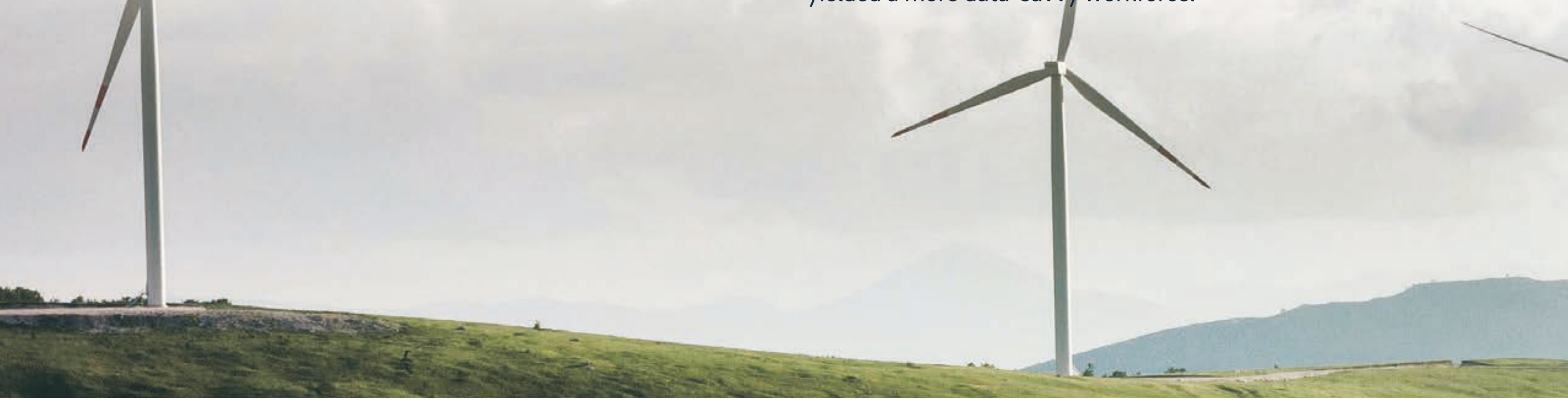
EARTHCUBE'S ACHIEVEMENTS IN FAIR INFRASTRUCTURE

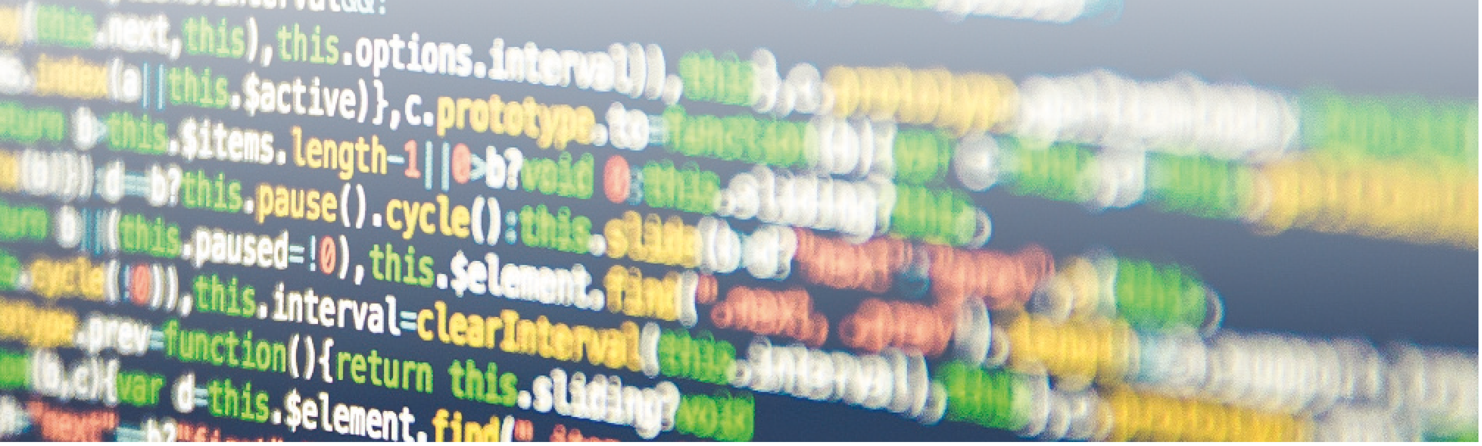
Noteworthy progress was made through EarthCube on promoting the FAIR (Findable, Accessible, Interoperable, and Reusable) principles for scientific data, conceptually extended to include software, notebooks, workflows, and additional geoscientific artifacts. FAIR-related EarthCube efforts that merit particular note are:

- **Council of Data Facilities** host, a potential forum for coordinating persistence and FAIRness across repositories;
- a GeoCODES JSON-LD crawler which enables **federated metadata** and searches across agency repositories;
- elevating the potential of **executable notebooks** to serve as persistent scholarly objects that, in particular, enhance reusability.

INCREASED CI AND DATA LITERACY

Over the course of the program EarthCube cultivated a community of nearly 2000 cross-domain professionals, enabling productive collaborations between researchers as well as data and technical professionals. Benefits included a corpus of new resources and tools that directly address geoscience research and education needs and that fill infrastructure gaps. Beyond their direct impact on capacity, deployment of these resources in EarthCube's community yielded a more data-savvy workforce.





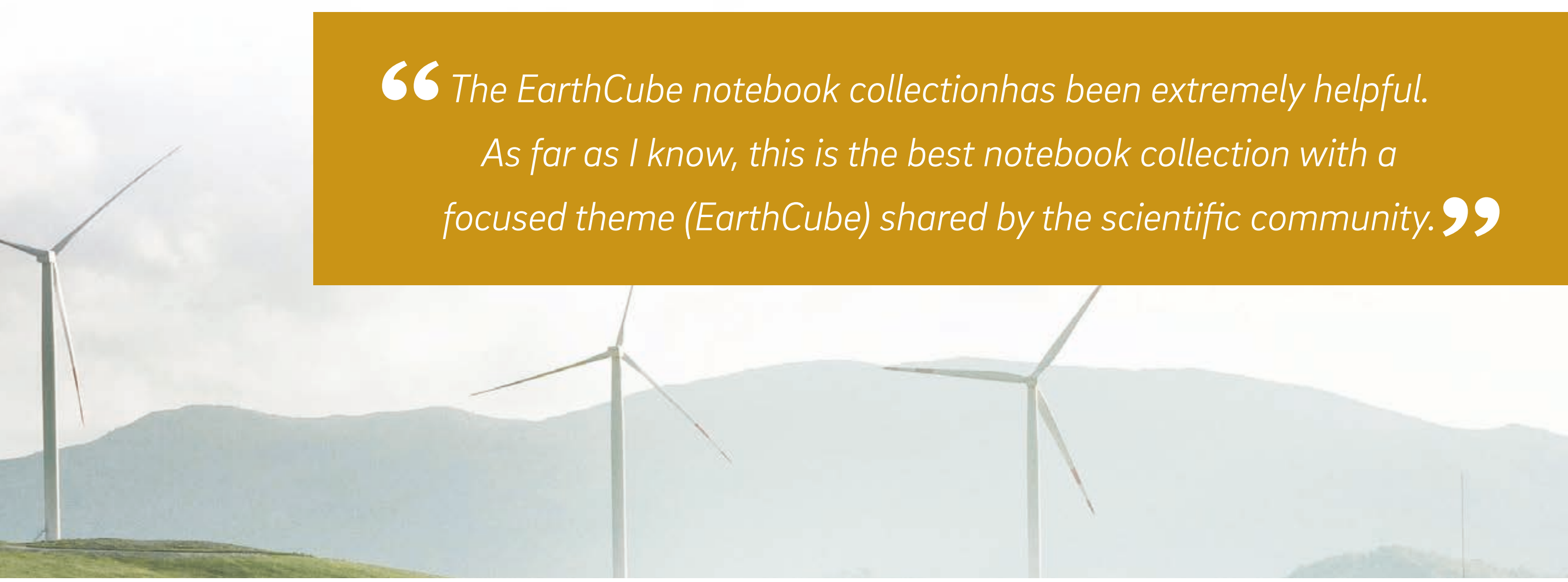
NSF program officers have observed that grant proposals indicate researchers' better understanding of data complexities. More scientists see value in making their data easily available and shareable; they now include DOIs for data connected to their published research results, which has resulted in an increase of citations related specifically to data and software within geoscience publications.

COMMUNITY BEST PRACTICES AND STANDARDS

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.

NOTEBOOKS AS SCHOLARLY OBJECTS

EarthCube has capitalized on the potential of executable notebooks to support learning and scholarship. The growing interest in peer-reviewed notebook presentations at the 2020, 2021, and 2022 EarthCube Annual Meetings became a locus of fruitful interaction between technological and geoscientific specialists in the EarthCube community and served as a catalyst for encouraging submissions from early career researchers. The educational potential of notebooks was also applied to learning about EarthCube capabilities per se, as well as their application to real (and often transdisciplinary) geoscience problems. The annual competitions served to build and keep fresh the EarthCube notebook directory of 44 published notebooks (*www.earthcube.org/notebooks*), easily accessed by researchers looking for examples of geo tools and datasets, as well as useful more broadly for geoscience educators.



"The EarthCube notebook collectionhas been extremely helpful. As far as I know, this is the best notebook collection with a focused theme (EarthCube) shared by the scientific community."

Research Outputs

SCHOLARLY OUTPUT FROM THE FUNDED PROJECTS

The EarthCube program has produced a wide array of scholarly assets, including papers, software, data and other project outputs. The publication metrics presented below were compiled to help assess the impacts and reach of the EarthCube program within its target scientific communities. While the publications examined were primarily drawn from the NSF's publication database, please note that the publication count is not complete because some publications were produced after the timeline of funded projects, some projects are still ongoing (as of this writing) and have not reported their final publication lists yet, or because publications reported to the NSF were not assigned DOIs or there were unresolved errors in the DOIs assigned.

DATA SUMMARY

215 grants

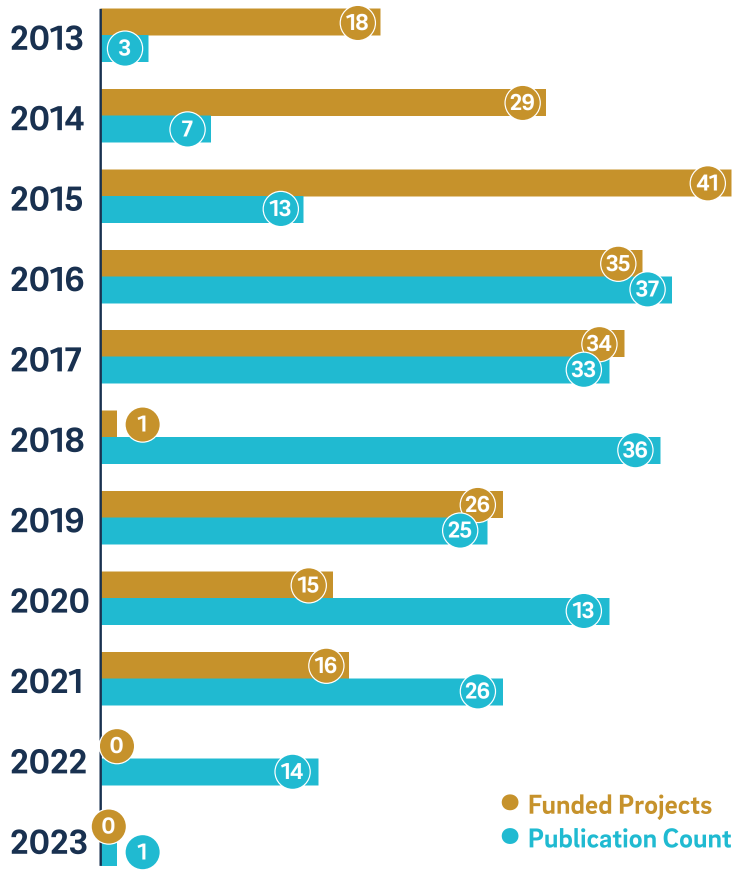
231 total publication DOIs analyzed

4248 total citations accumulated

18.4 citations per publication (on average)

6 citations per publication (median)

FUNDED PROJECTS AND PUBLICATIONS



TOP 10 MOST-CITED EARTHCUBE PUBLICATIONS

Citations	Publication
375	Yang, C., Q. Huang, Z. Li, K. Liu, and F. Hu, 2016: Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth,* 10, 13–53, https://doi.org/10.1080/17538947.2016.1239771.
356	Morlighem, M., and Coauthors, 2017: BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation. *Geophysical Research Letters,* 44, https://doi.org/10.1002/2017gl074954.
234	Roux, S., and Coauthors, 2018: Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology,* 37, 29–37, https://doi.org/10.1038/nbt.4306.
186	2017: A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data,* 4, https://doi.org/10.1038/sdata.2017.88.
140	Cervone, G., L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, 2017: Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renewable Energy,* 108, 274–286, https://doi.org/10.1016/j.renene.2017.02.052.
138	Williams, J. W., and Coauthors, 2018: The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. Quaternary Research, 89, 156–177, https://doi.org/10.1017/qua.2017.105.
113	Yang, C., M. Yu, F. Hu, Y. Jiang, and Y. Li, 2017: Utilizing Cloud Computing to address big geospatial data challenges. *Computers, Environment and Urban Systems,* 61, 120–128, https://doi.org/10.1016/j.compenvurbsys.2016.10.010.
94	Gil, Y., and Coauthors, 2016: Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. *Earth and Space Science,* 3, 388–415, https://doi.org/10.1002/2015ea000136.
92	Farley, S. S., A. Dawson, S. J. Goring, and J. W. Williams, 2018: Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience,* 68, 563–576, https://doi.org/10.1093/biosci/biy068.
79	Radcliffe, D. E., and Coauthors, 2015: Applicability of Models to Predict Phosphorus Losses in Drained Fields: A Review. *Journal of Environmental Quality,* 44, 614–628, https://doi.org/10.2134/jeq2014.05.0220.

TOP 10 HIGHEST ALTMETRIC SCORING EARTHCUBE PUBLICATIONS

(Altmetric scores indicate attention that articles get across social media, blogs, news outlets, and other web sites. In general, a score of over 100 typically puts a paper at or above the top 10% of all Altmetric tracked papers.)

Altmetric Score	Citation
981	2017: A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data,* 4, https://doi.org/10.1038/sdata.2017.88.
919	Morlighem, M., and Coauthors, 2017: BedMachine v3: Complete Bed Topography and Ocean Bathymetry Mapping of Greenland From Multibeam Echo Sounding Combined With Mass Conservation. *Geophysical Research Letters,* 44, https://doi.org/10.1002/2017gl074954.
581	Lepore, C., R. Abernathey, N. Henderson, J. T. Allen, and M. K. Tippett, 2021: Future Global Convective Environments in CMIP6 Models. *Earth's Future,* 9, https://doi.org/10.1029/2021ef002277.
275	Chen, B., and Coauthors, 2020: Measurement of magnetic field and relativistic electrons along a solar flare current sheet. *Nature Astronomy,* 4, 1140–1147, https://doi.org/10.1038/s41550-020-1147-7.
170	Roux, S., and Coauthors, 2018: Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology,* 37, 29–37, https://doi.org/10.1038/nbt.4306.
167	Williams, J. W., and Coauthors, 2018: The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quaternary Research,* 89, 156–177, https://doi.org/10.1017/qua.2017.105.
155	Husson, J. M., and S. E. Peters, 2017: Atmospheric oxygenation driven by unsteady growth of the continental sedimentary reservoir. *Earth and Planetary Science Letters,* 460, 68–75, https://doi.org/10.1016/j.epsl.2016.12.012.
145	Hulbert, C., B. Rouet-Leduc, P. A. Johnson, C. X. Ren, J. Rivière, D. C. Bolton, and C. Marone, 2018: Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience,* 12, 69–74, https://doi.org/10.1038/s41561-018-0272-8.
142	McNutt, M., K. Lehnert, B. Hanson, B. A. Nosek, A. M. Ellison, and J. L. King, 2016: Liberating field science samples and data. *Science,* 351, 1024–1026, https://doi.org/10.1126/science.aad7048.
107	Zheng, W., 2022: Glacier geometry and flow speed determine how Arctic marine-terminating glaciers respond to lubricated beds. *The Cryosphere,* 16, 1431–1445, https://doi.org/10.5194/tc-16-1431-2022.

EARTHCUBE AND THE GEOSCIENCES COMMUNITY ACROSS THE DECADE

The National Science Foundation launched the EarthCube initiative in 2011 with the bold ambition to develop "a framework to understand and predict responses of the Earth as a system." This called for culture change in the geosciences community, where data was sometimes seen as proprietary and disciplinary boundaries often limited collaboration. Upon conclusion of the EarthCube initiative in 2022, a team compiled and analyzed EarthCube and geosciences community survey responses from 2013, 2015, 2021, and 2022 - as well as data from a 2016 article in Data Science Journal entitled "Build it, but will they come?"[24]—that together indicate the following changes over the decade:

- **Increased amount and complexity of data:** During the ten years of EarthCube's funding by the NSF, the amount and complexity of data has increased dramatically and the vast majority of the geosciences community have indicated that it is still very important to find, access, and integrate data, models, and software.
- **Important Progress:** EarthCube advanced the geosciences community in important ways by creating more inclusive and collaborative forums for the array of disciplines involved with the initiative.
- **Increased Expectations:** In EarthCube's span, the community's needs increased and expectations grew for greater capability and integration of research computing infrastructure and data sharing within the community.
- **Continued Challenges:** There were persistent challenges related to the ease of finding, accessing, and integrating data, models, and software. Although EarthCube fostered substantial communication and collaboration between geoscientists and cyberinfrastructure developers, there remains important work to be done in growing these connections. In addition, participants felt that future endeavors should include more end user training.

The four data sets, 2013 (n=1,542), 2015 (n=449), 2021 (n=551), and 2022 (n=160), were each collected for different reasons from an array of EarthCube community participants. Baseline data are 2013 and 2015; current data are 2021 and 2022. Not all questions were asked in 2013, 2015, 2021, and 2022, so we include only the common questions here.

Culture Change in the Geosciences	**Baseline**	**Current**
Importance of finding, accessing, and/or integrating multiple data sets, models, and/or software. *(Percent responding "important/very important")*	82%	91%
Difficulty of finding, accessing, and/or integrating multiple data sets, models, and/or software. *(Percent responding "difficulty/very difficult")*	65%	60%

Culture Change in the Geosciences	**Baseline**	**Current**
EarthCube initiative is inclusive in the way it operates. *(Percent responding "agree/strongly agree")*	36%	62%
The balance between cooperation and competition in the culture of your field or discipline. *(Percent emphasizing cooperation)*	25%	48%
Degree to which success is primarily a product of individual effort or a product of collective effort. *(Percent emphasizing collective effort)*	42%	58%
Trusting that shared data, tools, models, notebooks, and software will be well-documented and reliable. *(Percent responding "agree/strongly agree")*	25%	29%

> *"...the impact of EarthCube was not necessarily the tools or research currently enabled, but rather the culture change that it has instilled throughout the research cycle to increase the value of collaborative work around data."*

Cooperation, Communication, Sharing, Training *(Percent responding "agree/strongly agree" in each case)*	Baseline	Current
A high degree of cooperation and sharing of data, models, and simulations among geoscientists.	23%	30%
Approximate number of data sets, tools, models, notebooks or software you made publicly available to other scholars in the past 5 years. *(Mean)*	2.8	8.6
There is sufficient communication and collaboration between geoscientists and those who develop cyberinfrastructure tools and approaches to advance the geosciences.	12%	8%
There is sufficient geoscience end-user knowledge and training so they can effectively use the present suite of cyberinfrastructure tools.	4%	5%

Looking Ahead

Respondents to the 2022 survey were asked **what is most needed to promote innovation in geoscience research.** The top responses centered on **funding.**

> "Funding and professional/career recognition for work to make data and other information artifacts reusable (FAIR)."

> "Similar initiatives to carry on the work. This isn't a one and done deal. There is still a lot more room for improvement in all disciplines."

A close second were responses emphasizing **community building, collaboration, and communication.**

> "Continued collaboration between geoscience investigators and data science and CI professionals."

Also of high importance were issues of **standards, documentation, and easy/open access to data.**

> "Ease of integration across data and models toward end use applications beyond basic research; integrating human systems, data and models."

> "Data standards that are enforced. Let's stop dumping "data" into repositories in formats that are not accessible."

Another cluster of common responses focused on **training, education, and career management.**

> "Partnership with groups working on equity and diversity in data science."

> "Education programs focusing on access and persistence of data science at the post-secondary level."

> "Career paths for staff software and data managers."

LESSONS LEARNED

ORGANIZATIONAL DEVELOPMENT

- The combination of a funded central office, complete with staff with specialized expertise, and a strong elected community leadership with clearly defined roles and decision making authority, was the most effective governance model.
- The flexibility to refine scope allowed EarthCube to embrace community-driven change and unanticipated opportunities.
- NSF site visits and reviews were crucial to major structural changes within EarthCube.
- Nimble, small-scale funding via Working Group proposals that were approved and overseen by EarthCube leadership was extremely beneficial for trying out ideas (e.g. the early career travel grants and furthering the DEI and sustainability efforts) and fostering key governance work.
- An early over-reliance on volunteer time was problematic. Providing key governance committee members with small honoraria and funding dedicated office staff to implement governance-approved decisions, policies, and programs prompted results.

LEADERSHIP

- Standing committees often struggled to identify unifying mission-driven work. Requesting 6- to 12-month work plans from committees kept them productive.
- Assessing the organization's culture through understanding our shared values facilitated gaining trust, pinpointing top priorities, and focusing finite resources for the greatest impact.
- Creating a small Cabinet with diverse expertise proved highly effective in managing a large collaborative project.
- Having one consistent program officer at NSF throughout the duration of the program was key, as this created the opportunity for an in-depth understanding of the vision and the reality of the process. Leadership was able to build on previous lessons learned rather than relearning the lesson at the expense of time lost or resources used less optimally.

SPECIALIZED EXPERTISE

- With a centralized office and dedicated staff available to serve the EarthCube community, it became possible to assemble a team with specialized expertise, as opposed to a volunteer community model staffed by researchers and their students.
- **Science writers** employed by ESSO and ECO ensured success stories and community profiles were highlighted and disseminated. The relatively modest investment paid outsized dividends in exposure and outreach.
- An **evaluator** was engaged by ECO to ensure activities planned were connected to impactful and measurable outcomes. This work continued through the first half of the project, rather than being clustered toward the end of the award, to ensure that any adjustments needed for stakeholder alignment could be made early and validated.
- Domain-embedded **information scientists** used data available to provide insight on impact, to craft guidance on software citation, and to create curriculum for just-in-time training and workshops on research data management topics.
- **Geoscientists** provided perspective on the priorities of domain scientists, and ways to communicate with them effectively. This was critical to funded projects, as well as to the office. Later EarthCube solicitations required geoscientists to be involved at a high level, which proved effective in connecting CI development to real community needs and uptake.
- In the last office, a **dedicated technical team** made up of geo-cyberinfrastructure experts implemented proof of concept infrastructure, worked with industry to propel collaboration on standards, and improved interfaces through user interface and experience (UI/UX) assessments.

BUILDING COMMON CYBERINFRASTRUCTURE

- Encouraging buy-in of cyberinfrastructure starts with small, flexible initiatives that can allow for building support for large-scale changes.
- For individual researchers to come together and work collaboratively required putting ego, competitiveness, and positioning aside to be successful.
- Developing a community architecture can be a slow and difficult process, with the benefits being realized by later projects, allowing stronger convergence among the tools.
- The successful building of common CI required support within scientific communities as well as NSF Directorates.
- With most data facilities having limited resources for new development, training and technical support is needed to implement the schema.org markup.
- Tagged software repositories helped build a CI community in ways that were not possible at the start of EarthCube, making the developing use of modern technologies another key pillar of success.
- An important step in developing a common CI such as GeoCODES included approval by the community.

COMMUNITY ENGAGEMENT

- Strategic stakeholder engagement needed to be implemented early and often to support the evolution of governance structure, and be given enough time for the process to evolve. Having an independent facilitator helped with that process. Stakeholder participation in design and implementation also helped to create a more transparent, relevant, useful, and sustainable program.
- The Annual Meeting encouraged buy-in to the community-driven process by providing ample opportunities for relationship-building and in-depth discussions on process and goals.
- Onboarding webinars for newly funded projects, followed by repeated opportunities for interactions in the Council of Funded Projects, fostered connections among the funded projects and a sense of community.
- Different types of data infrastructure projects require different levels of interactions with their communities. Database projects require engagement by trusted disciplinary scientists; framework projects are inseparable from their community; and middleware projects are less dependent on community trust.

SUCCESSFUL COLLABORATIVE PROJECTS

- While geosciences cyberinfrastructure projects require expertise in both CI and geosciences, the goals and needs of science and technology team members differ. What is cutting edge for one side may not engage the other. A successful project must recognize and support professional rewards for all team members
- The geoscience community needs increased education and/or skills in CI.
- The project phases of R&D, hardening, adoption and outreach required a variety of personnel filling a range of roles, and a longer time frame than 3 years.
- It is difficult to attract and retain technical staff, particularly for projects with short term funding requiring part-time expertise. Creating a pool of people with shared technical expertise and geoinformatics experience who can split their time across projects is one approach, though this was more tractable for larger research institutions.
- The transition from the initial development to operations phase of a CI project was a major stumbling block because different skills and management approaches are needed at different times.

BUILDING A USER BASE

- Creating a strong user base is critical for project success and sustainability and required effort and planning. It takes time to engage the user community and support uptake through training and outreach. Asking the community about their needs and being flexible enough to adapt plans accordingly was important. Use cases and user scenarios were particularly helpful tools.
- Making adoption as easy as possible was key. Meeting researchers where they were, integrating existing workflows when possible, making well-documented examples and targeted demonstrations and being flexible encouraged broad uptake.
- Notebooks and hackathons were helpful in building a user community.
- Being ready to repeat your message is important since it can take potential users many exposures to understand what is useful.
- A translator who can pitch to geoscientists without jargon was key, and required particular skills.
- Involving early career researchers became a priority. They were found to be more open to the adoption of new approaches, which they then shared with more senior colleagues.

BROADENING PARTICIPATION

- Set expectations for participation goals ahead of time; discuss, define, and track specific DEI targets.
- Co-designing activities with target audiences from the start was very effective, including designing a process for incorporating multiple perspectives and visions that is respectful and inclusive.
- While conference registration scholarships were valuable, providing full travel support (e.g. including airfare and hotel costs) is critical to assist early career researchers in attending in-person professional meetings.
- Collaborating with organizations already strongly engaged in DEI initiatives was productive. Initiatives can include having collaborators from a Minority Serving Institution, engaging with campus organizations, or partnering with programs set up to support participation from developing nations.
- Hold in-person events in the communities the project is seeking to engage or serve.

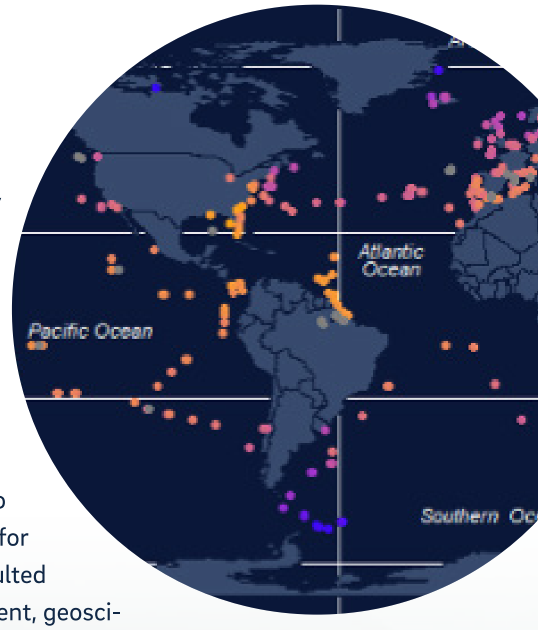
PUBLICATION METRICS AND MEASURING IMPACT

- Encourage uptake of tools that improve tracking, whenever they appear to have enough maturity to be accepted.
- Program outputs require broad scoping, since there are many different outcomes beyond just publications.
- Not all projects produced publications, instead producing posters, presentations, and/or infrastructure, software or data. Measuring the impact of these non-publications requires tracking tools that may not be verifiable or well-supported, but perhaps risks need to be taken to encourage novel tracking methods.
- Program evaluation needs to have visibility to all relevant program outputs, and some of the more frequent outputs (e.g. software and data) need to have strong support for their best practices to gain formal exposure.
- "Program success" needs to be calibrated to relevant outcomes, and relevant outcomes need to be articulated and tracked at program inception.
- Tracking project outcomes becomes much easier when persistent identifiers have been assigned.
- Automate as much as possible with human-in-the-loop and domain experts.

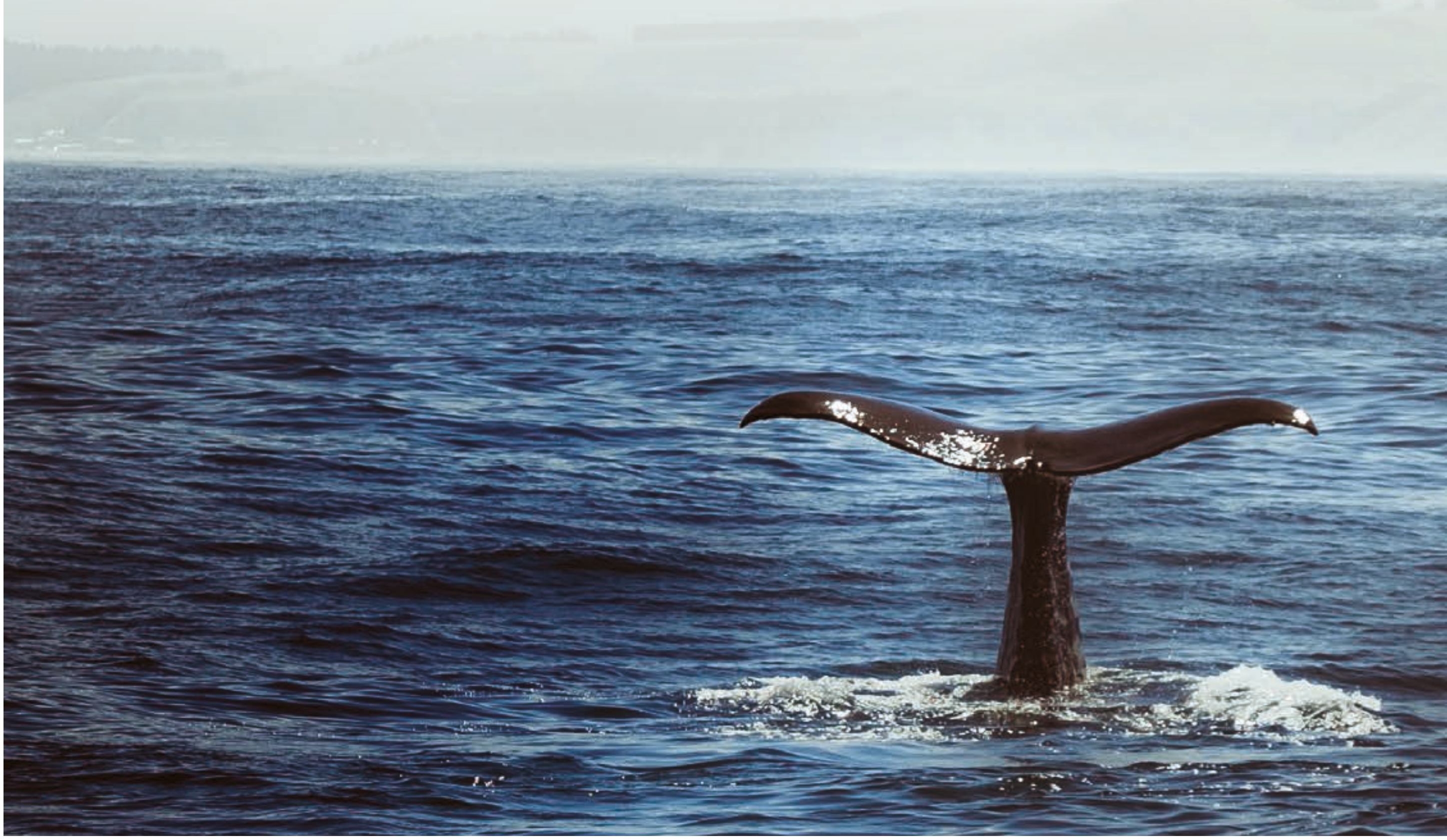
PROJECT SUSTAINABILITY

- Begin sustainability planning from the start.
- Consider development targets that have high value but low overhead, and can persist with minimal support (e.g. are hosted and run on a server with very little cost).
- Avenues for community use such as placing open source software on GitHub was great in concept but difficult to engineer and maintain in the long-term.
- Most successful examples of EC project sustainability involved embedding a resource in an existing institution with long-term support or using campus infrastructure.
- Creating a new organization, such as a not-for-profit or Limited Liability Company (LLC) can allow more diverse forms of funding, such as mandatory fee-for-use or voluntary contributions, though this approach is challenging.

CONCLUSIONS



Valuable cyberinfrastructure (CI) across the geosciences would certainly have emerged if the EarthCube funds had instead been distributed as individual research grants, connected only by the common characteristics of the associated request for proposals. However, EarthCube achieved more than could be gained from a portfolio of independent, disconnected projects. Indeed, key outcomes (e.g. GeoCODES, notebooks-as-scholarship, and standardization guidelines) would have had little value without significant community uptake. Annual meetings, standing committees, and early-career engagement were critical to such uptake. Further, while EarthCube solicitations focused on benefits for research, with no mention of science education, the ecosystem that resulted might well be characterized as a "learning community." In this environment, geoscientists and technologists, including many who were launching their careers, gained new understandings as well as new collaborative relationships. EarthCube was successful because it crossed disciplines and sparked collaborations among technical and science experts to a degree that was previously uncommon. Given the growing complexities in both science and technology, many of the lessons learned and successes from EarthCube will be critical if we hope to truly advance science at the scope and breadth that are required to answer our most pressing questions.



> "The benefits of EarthCube will continue to come out for years. It built the foundation of something hard and will take a while to unfold."

HALL OF FAME

Governance Leaders

LC- LEADERSHIP COUNCIL (2015-2022)
Basil Gomez (Chair), David Arctur, Danie Kinkade, Kerstin Lehnert (Chair), Lynn Yarmey, Janet Fredericks, Scott Peckham , Farzad Kamalabadi, Yolanda Gil, Ken Rubin (Chair), Simon Goring, Lindsey Powers, Tim Ahern, Sara Graves, Rebecca Koskela, D. Sarah Stamps, Anna Kelbert, Ouida Meier, Ryan Gooch, Mike Daniels (Chair), Isabel Cruz, Steve Diggs, Jed Brown, Maša Prodanović, Corinna Gries, Denise Hills, Emmanuel Njinju, Dave Fulker, Steve Kuehn, Leah LeVay

CDF - COUNCIL OF DATA FACILITIES (2014-2022)
Mohan Ramamurthy, Don Middleton,Tim Ahern, Danie Kinkade, Steve Diggs, Kerstin Lehnert, Corinna Gries, Sara Graves, Bob Arko, Jessica Hausman, Bernard Minster, Chuck Meertens, Doug Fils, Karen Stocks, Chris Crosby, Christine Laney, Jerry Carter, Bob Downs, Nicolas Jarboe

TAC - TECHNOLOGY & ARCHITECTURE (2015-2022)
Yolanda Gil, Jay Pearlman, Rebecca Koskela, Ruth Duerr, Dave Fulker, Donata Giglio

SC - SCIENCE COMMITTEE (2015-2020)
Basil Gomez, Ken Rubin, Emma Aronson, Kristen Rasmussen, D. Sarah Stamps

ET - ENGAGEMENT TEAM (2015-2020)
Marjorie Chan, Denise Hills, Simon Goring, Xuan Yu, Mimi Tzeng, Ryan Gooch, Dan Fuka

LT - LIAISON TEAM (2015-2020)
Lindsey Powers, Rick Ziegler, Joshua Young, Sarah Graves, Rowena Davis, Denise Hills, Leslie Hsu

CFP - COUNCIL OF FUNDED PROJECTS (2019-2022)
Basil Tikoff, Deborah Khider

SET - SCIENCE & ENGAGEMENT TEAM (2020-2022)
Denise Hills, Emma Aronson, D. Sarah Stamps, Steve Kuehn

Office PIs and Staff

ECTEG - EARTHCUBE TEST ENTERPRISE GOVERNANCE (2013-2016)
Lee Allison (PI), Kimberly Patten (Co-PI), Erin Robinson (Co-PI), Christopher Keane (Co-PI), Tina Lee (PI), Bruce Caron, Kate Kretschmann, Anna Katz, Rachel Black, Rowena Davis, Geneviev Pearthree, Randi Bellassai, Joel Cutcher-Gershenfeld

ESSO - EARTHCUBE SCIENCE SUPPORT OFFICE (2016-2019)
Mohan Ramamurthy (PI), James Davies, Julie Petro, Lynne Schreiber, Eric Lingerfelt, Sheri Ruscetta, Inkin Purvis, Doug Dirks, Joshua Young, Emily Villaseñor

ECO - EARTHCUBE OFFICE (2019-2022)
Christine Kirkpatrick (PI), Cathy Constable (Co-PI), Karen Stocks (Co-PI), Kenton McHenry (Co-PI), Lynne Schreiber, Catherine Cramer, Mike Bobak, Megan Carter, Kevin Coakley, Melissa Cragin, Mike Dwyer, Helen Evans, Doug Fils, Lisa Gatzke, Nancy Hoebelheinrich, Lucie Gustin, Daniel S. Katz, Rob Kooper, Kimberly Mann Bruch, Keith Maull, Matt Mayernik, Steve Richard, David Valentine, Ilya Zaslavsky, Summer Chao, Fangyu Zhou, Chris Battistuz, Julius Eshabarr, Erin Robinson (Co-PI), Rebecca Koskela, Ouida Meier

EAG - EarthCube Advisory Group

(2016-2019) Gwen Jacobs, Josh Semeter, Peter Wiebe, Tom Hoffman, Cathy Constable, Jeff Dozier, Jenni Evans, Dan Stanzione

(2019-2022) Gwen Jacobs, Ari Asmi, Nirav Merchant, Jeff Dozier, Cathy Constable

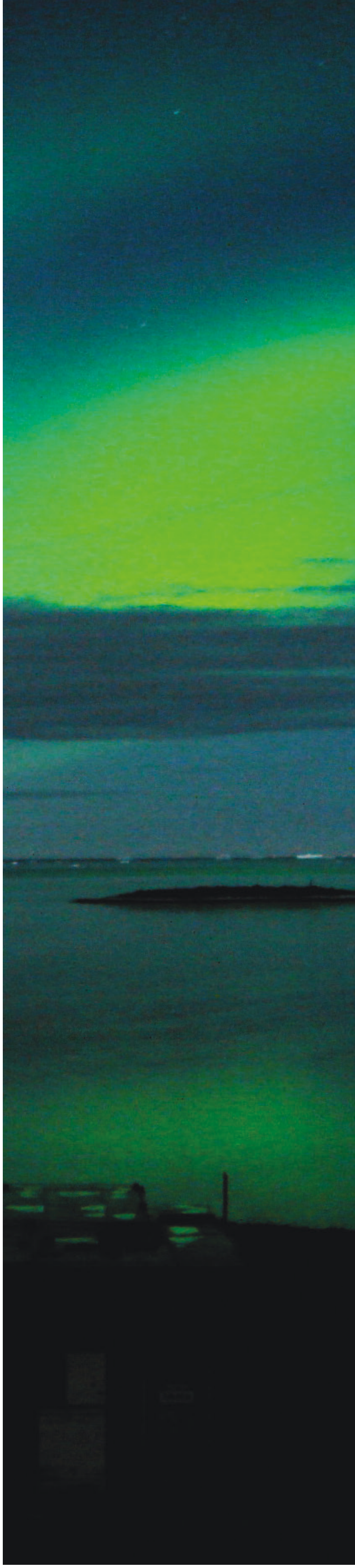
Key Partnerships

AGU, Arizona Geological Survey, EGU, ESIP, ESRI, GSA, NCSA, NSF Convergence Accelerator, ReSA, SDSC, SGCI, SIO, UCAR

Recognition Awards

COMMUNITY SERVICE AND LEADERSHIP AWARD
2015 - Simon Goring, Mimi Tzeng, Plato Smith
2016 - Raleigh Martin, Elisha Wood-Chatson, Emily Law
2017 - D. Sarah Stamps, Doug Fils, Dawn Wright
2018 - Leslie Hsu, Daniel Garijo, Dave Fulker
2019 - Matt Mayernik, Daniel Fuka
2022 - Emma Aronson, Deborah Khider

LEGACY AWARD
2019 - Yolanda Gil
2022 - Denise Hills



REFERENCES

1. Schreiber, Lynne; Evans, Helen (2022). EarthCube Funded Projects (2013-2024). In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J01C1X20*

2. EarthCube Leadership Council (2020). EarthCube Charter, 2020 version. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0C53JDQ*

3. Signell, Richard & Fox, Peter. (2012). Report on GeoData 2011 Workshop - Data Life Cycle, Integration and Citation. 9594-.

4. National Science Foundation (NSF). (June 21, 2011). Dear Colleague Letter: The "Earth Cube" - Towards a National Data Infrastructure for Earth System Science. Available: *https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf11065*.

5. National Science Foundation (NSF). (2011). Earth Cube Guidance for the Community. Available: *https://www.nsf.gov/pubs/2011/nsf11085/nsf11085.pdf*

6. EarthCube Office (2021). EarthCube End User Workshop Executive Summaries. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J03T9HC6*

7. Gil, Yolanda; Chan, Marjorie; Gomez, Basil; Caron, Bruce (Eds) (2021). EarthCube: Past, Present, and Future. EarthCube Project Report EC-2014-3, December 2014. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0Q52PHF*

8. Arctur, D., Barrett, J., Pearlman, J., Powers, L., Richard, S., Tricomi, M., & Zaslavsky, I. (2016). EarthCube Architecture Workshop 2016: Final Report and Recommendations. OpenSky. *https://doi.org/10.5065/2e84-c236*

9. EarthCube Advisory Committee (2021). EarthCube Advisory Committee Report (March 11, 2016). In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0VD6ZM8*

10. EarthCube Leadership Council (2016). Leadership Council Response to the EarthCube Advisory Committee Report. OpenSky. *https://doi.org/10.5065/45be-q612*

11. Daniels, M., Arko, B., Hsu, L., Koskela, R., & Pearlman, J. (2017). 2017 EarthCube Architecture Refinement Workshop Report. OpenSky. *https://doi.org/10.5065/52gx-w069*

12. Rubin, K. H., Kelbert, A., Stamps, D. S., Meier, O., & Koskela, R. (2019). FAIR Document. OpenSky. *https://doi.org/10.5065/ggbp-m642*

13. Rubin, Kenneth H.; Daniels, Mike; Fulker, Dave; Brown, Jed; Richard, Stephen; Meier, Ouida; Zaslavsky, Ilya; Willis, Craig; McHenry, Kenton; Kirkpatrick, Christine (2020). Recommended Standards and Specifications for EarthCube Projects. EarthCube Leadership Council (May 7, 2020). In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0QR4VMG*

14. Hills, Denise J.; Prodanović, Maša; Stamps, D. Sarah; Njinju, Emmanuel, A.; Goring, Simon, J.; Diggs, Stephen (2022). Community Participation Guidelines. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0R211KC*

15. EarthCube Organization (2020). EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0PG1Q4B*

16. Stocks, Karen; Evans, Helen (2022). Interviews with EarthCube Funded Projects: Perspectives on EarthCube's Impact and Sustainability Priorities. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0513ZC0*

17. Stocks, Karen; Evans, Helen (2022). EarthCube Project Challenges and Lessons Learned: Perspectives from Current and Past EarthCube Projects. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J08S4Q2Z*

18. Daniels, Mike; de La Beaujardière, J-F; Downs, Robert R.; Fulker, Dave; Hills, Denise J.; Jacobs, Gwen; Jha, Shantenu; Jones, Matt; Kinkade, Danie; Kuehn, Steve; Schreiber, Lynne; Walker, Doug; Cramer, Catherine (2022). EarthCube Sustainability Panel Report. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0CR5TJF*

19. Jarboe, Nicholas; Diggs, Stephen; Downs, Robert R.; Kinkade, Danie; Lehnert, Kerstin A.; Ramamurthy, Mohan; Seul, Martin; Schreiber, Lynne (2022). CDF Sustainability Task Force Report. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0P55NQJ*

20. Virapongse, Arika; Gallagher, James; Tikoff, Basil; Cornillon, Peter; Koskela, Rebecca; Shingledecker, Susan; Trabant, Chad; Hanson, Brooks (2022). Sustainability models for integrated digital Earth Science. In EarthCube Organization Materials. UC San Diego Library Digital Collections. *https://doi.org/10.6075/J0JH3MBN*

21. Carter, Megan (2020): Data Help Desk Planning Tips. ESIP. Online resource. *https://doi.org/10.6084/m9.figshare.12827717.v1*

22. Carter, Megan (2022): Virtual Data Help Desk Planning Guide. ESIP. Online resource. *https://doi.org/10.6084/m9.figshare.19674579.v1*

23. Maull, Keith and Mayernik, Matthew. 2022. EartCube Program Metrics Analysis 2013-2022. UCAR/NCAR. *https://doi.org/10.5065/vmfj-qy55*

24. Cutcher-Gershenfeld, J., Baker, K.S., Berente, N., Carter, D.R., DeChurch, L.A., Flint, C.C., Gershenfeld, G., Haberman, M., King, J.L., Kirkpatrick, C., Knight, E., Lawrence, B., Lewis, S., Lenhardt, W.C., Lopez, P., Mayernik, M.S., McElroy, C., Mittleman, B., Nichol, V., Nolan, M., Shin, N., Thompson, C.A., Winter, S. and Zaslavsky, I., 2016. Build It, But Will They Come? A Geoscience Cyberinfrastructure Baseline Analysis. *Data Science Journal,* 15, p.8. DOI: *http://doi.org/10.5334/dsj-2016-008*


EarthCube