



# Insights on Sustainability of Earth Science Data Infrastructure Projects

RESEARCH PAPER

ARIKA VIRAPONGSE 

JAMES GALLAGHER 

BASIL TIKOFF 

\*Author affiliations can be found in the back matter of this article

]u[ubiquity press

## ABSTRACT

We studied 11 long-term data infrastructure projects, most of which focused on the Earth Sciences, to understand characteristics that contributed to their project sustainability. Among our sample group, we noted the existence of three different types of project groupings: Database, Framework, and Middleware. Most efforts started as federally funded research projects, and our results show that nearly all became organizations in order to become sustainable. Projects were often funded for short time scales but had the long-term burden of sustaining and supporting open science, interoperability, and community building—activities that are difficult to fund directly. This transition from ‘project’ to ‘organization’ was challenging for most efforts, especially in regard to leadership change and funding issues.

Some common approaches to sustainability were identified within each project grouping. Framework and Database projects both relied heavily on the commitment to, and contribution from, a disciplinary community. Framework projects often used bottom-up governance approaches to maintain the active participation and interest of their community. Database projects succeeded when they were able to position themselves as part of the core workflow for disciplinary-specific scientific research. Middleware projects borrowed heavily from sustainability models used by software companies, while maintaining strong scientific partnerships. Cyberinfrastructure for science requires considerable resources to develop and sustain itself, and much of these resources are provided through in-kind support from academics, researchers, and their institutes. It is imperative that more work is done to find appropriate models that help sustain key data infrastructure for Earth Science over the long-term.

## CORRESPONDING AUTHOR:

**Arika Virapongse**

Middle Path EcoSolutions, US  
[av@middlepatheco.com](mailto:av@middlepatheco.com)

## KEYWORDS:

Earth Science data;  
sustainability;  
cyberinfrastructure; databases;  
funding; community

## TO CITE THIS ARTICLE:

Virapongse, A, Gallagher, J and Tikoff, B. 2024. Insights on Sustainability of Earth Science Data Infrastructure Projects. *Data Science Journal*, 23: 14, pp. 1–27. DOI: <https://doi.org/10.5334/dsj-2024-014>

## 1. INTRODUCTION

Increasing the impact of publicly funded science projects has long been a goal for US federal agencies. Oftentimes funded projects are asked to provide a sustainability plan as part of an effort to emphasize their potential long-term impact. Despite the emphasis on sustainability of projects, evaluation models and studies that assess sustainability in a systematic way, and the variables that affect sustainability, are generally lacking for science-related projects (Stirman et al. 2012).

While many science initiatives begin as projects (with a determined beginning and end), data initiatives in Earth Science now push the boundaries of whether or not projects merit more long-term governmental financial support. Cyberinfrastructure created by data initiatives is often expensive and resource-intensive to maintain but is also becoming increasingly critical for conducting science.

In this paper, we examine a set of 11 Earth Science data infrastructure projects that demonstrate longevity (ongoing for over 10 years) for the purpose of understanding: What aspects of these projects have contributed to their sustainability? Further, we seek to understand: What challenges did/do they face with sustainability? What does sustainability mean to these projects? While we selected most projects because of their successful continuity, we intentionally included one project that chose to close and one project that was being integrated into another larger project. We believe that such a breadth of perspective added to the richness of our study, as the interviewees that contributed to our study were familiar with success, failure, and drastic change within projects.

With the results of this paper, we hope to help improve sustainability of earth science data infrastructure initiatives and promote system-wide sustainability. Such results can be useful for researchers developing and leading projects regarding Earth Science data infrastructure, and agencies that provide and manage funding for such projects. More broadly, this paper highlights a gap in science that has implications for society. In an era defined by intense environmental change and increasing data volume, Earth Science data infrastructure is needed to help support decision making about our human relationship with the world.

### 1.1 BACKGROUND

The original motivation for this work was based on discussions within the Council of Funded Projects of the EarthCube program funded by the US National Science Foundation. The Earthcube program, which ran from 2011 to 2023, was a community-driven initiative with the goal to transform ‘the conduct of geosciences research and education by building a well-connected cyberinfrastructure’ (EarthCube Office; EarthCube Leadership Council 2022). As the EarthCube program was coming to an end, the Council of Funded Projects sought to explore what the next steps might be, and this study was commissioned as one result.

Most Earth Science data infrastructure initiatives in the US begin their life as a project with short-term funding, providing the opportunity for a project to demonstrate a proof of concept. To sustain the value of a project after its initial funding period, a project typically moves towards becoming an organization. An organization is a legal and administrative body that separates individual goals from the goals of the project itself. We assume that an *organization* is something that does not have a defined end point but is instead the continuation of an effort toward an open-ended goal. This idea is fundamental to how we define sustainability here: the ability of a project to maintain its core mission and functions beyond the initial seed or start-up funding, and despite a change over time of the individuals who make up the organization. This definition of sustainability incorporates key elements of sustainability that are often used in other studies such as: maintenance of core elements, benefits after initial support has been withdrawn, modifications to a project’s benefits, and ability to function in order to maintain benefits (Stirman et al. 2012).

Earth Science presents a unique case study for understanding sustainability of organizations. Here, we use ‘Earth Science’ to include the study of all fields of natural sciences related to the planet Earth, including the atmosphere, hydrologic system, and oceans (NSF 2023). Both meteorology and oceanography, in particular, have come to rely on shared data resources. Often those data are the product of government agencies (e.g., data from the satellite-borne

Moderate Resolution Imaging Spectroradiometer (MODIS) (Kaufman 1997) sensors) or derived from data produced by agencies and then shared with the NASA Science Investigator-led Processing Systems (SIPS) program (NASA 2020). Roughly equivalent statements can be made for the atmospheric science research community.

Sustainability efforts in Earth Science have much to gain from advances made in other fields. The ecological community, in particular, has been engaged in major efforts on sustainability in data science (e.g., ESA 2018; Maron 2018; Volentine et al. 2021). Other work has addressed the vexing problems of how to financially support and intellectually maintain open data repositories (Erway 2023; Goben & Sandusky 2020; Kitchin et al. 2015). There is recognition that critical aspects of community buy-in extend beyond data stewardship to the attitudes and operating procedures of the people and organizations maintaining the digital systems (Lin et al. 2020). Finally, commercial and free/open-source software development efforts, which both follow similar growth patterns in terms of technical complexity (Paulson 2004), have long been compared in regard to their structure and governance (Bitzer 2006; Ferraz & Santos 2021; Raymond 1999).

Earth Science often addresses complex, vast, and interdependent systems with large multi-dimensional heterogeneous data and information that are too challenging for a small group of people (with inherently limited perspectives) to understand and address. For example, satellite data and model data are often ‘large,’ always pushing the edge of computational and storage systems. To succeed in a better understanding of complex Earth systems, Earth Science is faced with the challenges of both breaking down silos through collaboration among people and fields and integrating datasets together through technical and social interoperability.

## 2. METHODS

### 2.1 APPROACH

We examined Earth Science data infrastructure projects that successfully outlasted their initial funding period in order to find common traits that contributed to their organizational longevity—our initial definition of ‘sustainability.’ We use the term ‘project’ to describe our case studies, since most efforts in our study began as projects. We do acknowledge that over time, most study projects transitioned into more formal entities, which would technically be better described as ‘organizations.’

Our study was designed as qualitative team research based on grounded theory. With grounded theory, data collection, analysis, and theory are closely related; theory emerges as data are collected and analyzed during the study (Strauss & Corbin 1998). We applied grounded theory by using a scaffolding approach to select interviewees, analyze the data they provided to us, and develop theories to explain what we understood. These insights then determined who we would interview for the next round of data collection, so we could further explore and test our theories as the study progressed.

Our study consisted of three phases:

Phase 1: Project design (February to November 2021; 5 months).

Phase 2: Analysis and drafting an initial report for the funder (January to May 2022; 10 months).

Phase 3: Fine-tuning analysis and drafting of this peer-reviewed paper (June 2022 to February 2023).

The first phase of the project included forming a ten-person research team that designed the research process, methods, and data collection templates, and identified the sample group. Team members provided diverse perspectives and experiences relevant to understanding the broader context around the project; they had a spectrum of previous involvement in Earth Science including publishing, research, national labs, funded projects, and social science. A core team of three individuals (co-authors AV, JG, and BT) was established to provide overall coordination, research design, project management, and accountability for the project. Interviews were conducted by the core team.

For the second phase of the project, the research team was reduced to eight individuals. Outputs of this phase included 11 completed interviews, analysis of results, and writing of the final report of the project (commissioned by the US National Science Foundation (NSF) EarthCube program). Throughout this data collection and analysis phase, the research team met weekly or biweekly to develop and implement the project—often alternating between core group meetings (JG, BT, AV) and full team meetings (8 people total)—as we worked together to refine our process, gather information, and analyze data. The third phase of the project consisted of the core group reviewing and fine-tuning the data analysis and drafting this final peer-reviewed paper.

The interview process and questions were reviewed by the Institutional Review Board (IRB) at the University of Wisconsin – Madison and considered ‘exempt’ from further review. In keeping with the intent of the IRB, we utilized an informed consent with the interviewees, which they agreed to verbally before beginning the interview. We sought to maintain anonymity and privacy of the interviewees; therefore, primary data (i.e., audio recordings, interview notes, and raw transcripts) are not available outside of the research group. A detailed methodology for our study can be found in Virapongse et al. (2022).

## 2.2 SAMPLE GROUP

The sample group consisted of 11 key informants. The sample group was selected based on a two-tier process that focused on selecting the project first, and then selecting an individual to interview as a representative of that project. The project and individual representing the project were selected based on the criteria detailed in Box 1. We used a sampling approach known as stratified sampling, which is a purposive sampling technique that targets informants from defined sub-sample groups (Robinson 2014). As opposed to random sampling, a purposive stratified sampling approach helps to ensure that sub-sample groups are represented in a study; this is especially important in qualitative studies that rely on small sample sizes due to the depth of their data collection. Stratified sampling was used to ensure that the projects we included in our study were representative by sub-sample group and project size, and that the individuals we interviewed were representative by gender and career stage.

**Box 1** Sample Group Criteria.

### Project sample group criteria:

- Relevant to Earth Science data
- Location of the project is in the US (e.g., main office, leadership)
- Existed for 10+ years
- Not a government-based project or national labs
- Stratified sampling considerations:
  - Sub-sample group designation: Database, Middleware, Framework
  - Size of the project: number of staff; small to large

### Individual sample group criteria:

- Had/has a strategic leadership role in the program
- Held their leadership role for at least 2 years
- Stratified sampling considerations:
  - Gender
  - Career stage: early, middle, late

Using a scaffolding approach based on grounded theory, the team first identified three initial sample group targets, then we completed the interviews and conducted a first round analysis of the data (resulting in the first-level derived product). Once that was completed, the team then focused on identifying the next few sample group targets. In this way, the team sought

to reduce redundancy and fill in gaps in knowledge for the overall study. This approach also resulted in our identification of three sub-samples groups—Database, Middleware, and Framework—that became a key element of both our research design and our results.

Table 1 shows the projects that were included in the study, and some key characteristics that were relevant to our sampling approach. The projects were selected to represent sub-sample groups of Database (5 projects), Middleware (3), and Framework (3). Database projects aimed to bring together data and data resources for use. Middleware projects sought to develop software and technology. Framework projects focused on developing best practices.

PROJECT NAME & WEBSITE	SUB-SAMPLE GROUP	ORGANIZATIONAL STRUCTURE	YEAR FOUNDED	CURRENT STAFF
<b>BCO-DMO</b> Biological & Chemical Oceanography Data Management Office <a href="http://www.bco-dmo.org">www.bco-dmo.org</a>	Database	University hosted, NSF funded	2006	5
<b>ESIP</b> Earth Science Information Partners <a href="http://esipfed.org/">esipfed.org/</a>	Framework	501(c)3	1998	5
<b>Force11</b> Future of Research Communications and e-Scholarship <a href="http://force11.org">force11.org</a>	Framework	501(c)3	2011	16
<b>HDF Group</b> Hierarchical Data Format Group <a href="http://www.hdfgroup.org">www.hdfgroup.org</a>	Middleware	501(c)3	2006, NCSA 1988	20
<b>IEDA</b> Interdisciplinary Earth Data Alliance <a href="http://www.iedadata.org">www.iedadata.org</a>	Database	University hosted, NSF funded	2010 (web site copyright)	14
<b>IRIS</b> Incorporated Research Institutions for Seismology <a href="http://www.iris.edu/hq/">www.iris.edu/hq/</a>	Database	NSF funded	1984	50
<b>OGC</b> Open Geospatial Consortium <a href="http://www.ogc.org/">www.ogc.org/</a>	Framework	501(c)3	1994	20
<b>OPeNDAP</b> Open-source Project for a Network Data Access Protocol <a href="http://www.opendap.org">www.opendap.org</a>	Middleware	501(c)3	2000, University of Rhode Island 1993	5
<b>PaleoDB</b> Paleobiology Database <a href="http://paleobiodb.org">paleobiodb.org</a>	Database	NSF Funded	1998	3
<b>SERC</b> Science Education Resource Center <a href="http://serc.carleton.edu/">serc.carleton.edu/</a>	Database	University hosted, NSF funded	2001	19
<b>Unidata</b> <a href="http://www.unidata.ucar.edu/">www.unidata.ucar.edu/</a>	Middleware	UCAR hosted, NSF funded	1984	20

**Table 1 Study projects.**

Projects included in the study and some defining characteristics.

NSF = National Science Foundation; 501(c)3 = a type of legal status for a non-profit organization in the United States; NCSA = National Center for Supercomputing Applications; UCAR = University Corporation for Atmospheric Research.

Efforts were made to select a demographically diverse sample group. We defined diversity as gender, career stage, and any other self-reported characteristics that the individual wished to share. The demographic data of the individuals in the sample group were all self-reported. The sample group consisted of 6 women and 5 men. When asked if they felt that they were part of an underrepresented group within the Earth Sciences and Data Sciences, interviewees identified the following characteristics of themselves: not being a US citizen (n = 2), 'out and

proud gay man,’ ‘not having a PhD,’ ‘I’m a neuroscientist,’ ‘I grew up in a rural environment,’ and being a professional musician.

Interviewees reported having these terminal roles in their project: Director (n = 3), President (n = 2), CEO (n = 2), Funding Program Manager, Lead, Senior Research Scientist, Executive Director; all were paid positions with the exception of ‘Lead.’ The years that interviewees have been involved in their project ranged from 5 to 30 years (average 18 years). Interviewees self-reported (based on their own definition) that they were in the following career stage when they started working on the project: Early (n = 2), Early/Middle (n = 3), Middle (n = 3), and Middle/Late (n = 3).

### 2.3 DATA COLLECTION AND ANALYSIS (BOX 2)

Interviews with 11 key informants were used as the source of primary data. We conducted a one-hour interview with each key informant. The interview followed a script, question template, and general order (i.e., a semi-structured interview approach; interview process, script, and template is available in [Virapongse et al. 2022](#)). This approach provided consistency as well as flexibility so that (sub)topics relevant to a particular interview could be examined in more detail through follow-up questions. The interview questions focused on topic areas that we believed were relevant facets of sustainability, including governance, community engagement, project assessment, business models, and strategic development. Other studies have used similar facets to assess sustainability ([Arp & Forbes 2018](#); [Hsu et al. 2019](#)). The interviews were recorded and automated transcripts were generated, saved, and reviewed for errors.

**Box 2** Summary of the Methods.

#### Data collected/generated:

- Recording of the interview, resulting transcript (not publicly available because this information is protected by IRB requirements)
- Notes by interviewers during the interview
- Publicly available background information about the project

#### Analysis process of the interviews:

- Coding of the transcript by 3 to 6 (average 5) team members, to identify main themes
- Group discussion of the coded transcript by 4 to 8 (average 5.9; one interview was not discussed by the group due to time/schedule constraints) team members
- Collection of all of the quotes, codes, and discussion notes into a summary document (first-level derived product)
- All of the coded quotes from all of the interviews were collected into a single document, and quotes were clustered together based on commonalities (second-level derived product)
- Clustered quotes were given a title and description, and these composed the final results

Each interview was conducted by two or three core team members (AV, BT, and JG), with one taking the lead and asking questions while the other one or two listened and took notes; core team members rotated through the roles, and all were present at most interviews to ensure consistency in how the interviews were conducted. The recording, transcript, and notes were made available to the full analytical team (8–10 individuals total), who annotated the transcript (known as ‘coding’ in qualitative analysis). In a subsequent meeting, the full team discussed their reflections about the interview, while one person from the team facilitated the discussion and took notes. The team then prepared a summary of the interview (consisting of quotes from the transcript, annotations, and notes from the facilitated discussion), and this served as a first-level derived product from the raw data of the interview itself.



Most members of the team rotated through roles, so that the roles were distributed across the team. By having numerous people participate in the coding and discussion of the interviews, a broader breadth of perspectives was represented within the analysis.

The first-level derived product was a document that brought together quotes, insights about the quotes (coding), and notes from the facilitated discussion among the team. Next, all of the quotes and insights about the quotes from the first-level derived products were compiled together into a single aggregated second-level derived product. Then, the core team grouped these data according to common themes. These themes were given titles and descriptions (known as a clustering approach; second-level derived product). This final derived product formed the basis for the Results presented in this paper. To improve readability, the quotes from the different sections of Results were removed from the main text and placed in Appendix 2. The notes from facilitated discussions among the team helped to inform the Discussion of this paper.

The 11 interviewees provided feedback on the initial report of this study (provided to the NSF as [Virapongse et al. 2022](#)), and their feedback was solicited for incorporation into this peer-reviewed paper.

### 3. RESULTS

The results are organized by five main topic areas: governance, community outreach and participation, assessing project success, business model, and strategic development. The results focus on reporting commonalities among interviewee responses, as well as some key outliers.

In parentheses, the frequency of interviewees included in a cluster is noted with ‘n,’ and the specific interviewees in the cluster are noted using the convention of Interviewee Number (ranging from 1 to 11) + Subsample group (D = Database, F = Framework, M = Middleware). To help protect the interviewee’s identity, we attempted to exclude any information that might link a specific organization to an interviewee code.

#### 3.1. GOVERNANCE

##### 3.1.1. Governance is intentionally created

All of the Framework projects and some Database projects (n = 5, 2F, 4D, 5F, 6F, 10D) noted that an intentionally created governance model was key to their project’s effectiveness. Both sub-sample groups worked closely with their communities (e.g., participants and collaborators in the projects), and they relied on governance models to integrate community input into a project. Likewise, poor governance could be detrimental to a project; sufficient attention and resources were needed to develop an effective governance model. Governance often evolved over time (2F, 4D) as a project established its identity, goals, and mission. In particular, new leadership and governance structures for decision making were sometimes needed after key milestones were achieved.

A main element of governance was to determine how decisions were made in the project (n = 4, 2F, 4D, 5F, 6F). Voting was used by the Framework projects, which all had large, open-ended communities, for things like selecting new volunteer leaders and community members. A consensus approach was used by a Database project (4D), which was notably a loose collection of researchers working together on the project. There was also attention given to achieving equitable decision making. To bring community perspectives into decision making, formal groupings like a policy committee, an advisory board, an executive committee, working groups, and strategic members were used (n = 5; 3M, 4D, 5F, 8D, 9D). These groups often provided advice and guidance for paid leadership to include in their implementation. They also provided a way for volunteers to organize and work on issues relevant to the project (e.g., working groups, 4D).

##### 3.1.2. Good documentation is often a key component of successful projects

Without any prompting, five interviewees (n = 5; 2F, 3M, 5F, 8D, 11M) mentioned the importance of good documentation in their project. Among the Framework and Database projects (n = 3; 2F, 5F, 8D), documentation was noted in the context of keeping a record about how and what decisions were made for governance of the project. In contrast, for the Middleware projects (3M, 11M), documentation was noted as important for technical development.

### 3.1.3. Top-down and bottom-up governance is balanced

When discussing top-down governance, interviewees often referred to the involvement of project staff and individual leaders (paid and unpaid, respectively). Top-down decision-making was often used for handling project funding, requirements of the funders, business decisions, and supporting and coordinating the community. In contrast, when discussing bottom-up governance, interviewees often referred to stakeholders in a project, end users, and general community that interacted with, contributed to, and had interests in the project. Bottom-up governance had the role of eliciting new ideas, innovating along the edges of the project scope, and increasing engagement of broader community members.

More than half of the interviewees noted that their project's governance models represented a 'balance' or 'mixture' of both top-down and bottom-up governance (n = 7; 2F, 5F, 6F, 7D, 8D, 9D, 11M). One interviewee (2F) noted that the federal funding program manager was heavily involved in the project with the purpose of supporting the community's own independence and self-reliance by bringing outside examples and expert advice to the community. Some interviewees from community-oriented projects noted that leaning more bottom-up was better for their projects (n = 5; 2F, 3M, 4D, 6F, 7D). They recognized the importance of community ownership, the role of community in helping their project stay relevant (e.g., maintaining funding streams), and for meeting their project mission. They also hoped to increase the autonomy of their community and encourage their contributions to the project.

In contrast, a more top-down governance approach was considered to be better for other projects (n = 3; 3M, 5F, 10D) because it provided necessary oversight and good decision making by carefully selected leaders. Top-down governance also provided more long-term and consistent commitment and accountability for a project; characteristics that were not often associated with volunteer contributions. One interviewee emphasized that when a community has too much decision-making power, the goals of a project can become diluted.

### 3.1.4. Leadership qualities are considered based on project needs and community representation

Project governance was led and implemented by individual leaders like hired staff and unpaid community champions, and small groups like Boards and working groups (also typically unpaid). When selecting leaders, projects often balanced their needs for star power, hands-on work, and domain area representation (n = 8; 2F, 3M, 4D, 6F, 7D, 8D, 9D, 11M). A leader with star power helped add credibility to a project so it could more easily capture funding and attract members (3M, 6F); however, these characteristics could be at odds with projects that sought greater community ownership (6F).

Leadership needs varied depending on what the project was aiming to accomplish for its mission and/or particular development stage. Leadership qualities—particularly in the context of after a project grows out of a planning phase—included coordination, administration, organization, project management, and convening groups (2F, 6F, 8D). Commitment and dedication by leaders were also highly valued (4D, 7D). Many of these qualities were learned while on the job, since leaders were often initially selected for their domain expertise (n = 2; 3M, 8D). Interviewees pointed out that projects within academia/science often required a leadership model that differed from conventional business leadership models. For example, many personnel in the study projects were highly educated academics. They often stayed in specific professional roles for only short periods of time, for example as post-docs, and their motivations were typically driven by their interest in a project. Unlike for conventional business models, most personnel did not seek to climb a career ladder within the study projects. Therefore, study project leaders had to develop their own leadership model to attract and retain personnel in the projects.

Earth Science data infrastructure projects required leadership by people with a high level of expertise in both a domain area and technology. Leadership that represented both areas was often needed to strategize effectively, connect with end users, and solve logistical problems. To address this challenge, one interviewee described the value of co-management, where two leaders bring their science background and technical know-how, respectively, into managing their project (9D).

Volunteer leadership was mentioned (n = 3, 2F, 6F, 9D) as a way to provide domain area representation, community input, build credibility, identify new opportunities and ideas for a



project, and engage the community broadly. Volunteers could also help reduce the amount of resources needed for staff because they could take on discrete but important tasks, such as representing the project publicly.

Social diversity in leadership and decision-making was valued by some interviewees, and particularly by Database projects ( $n = 5$ ; 4D, 6F, 7D, 9D, 10D). The types of diversity included geographic representation ( $n = 3$ ; 4D, 6F, 10D), domain area expertise ( $n = 2$ ; 4D, 7F), and people that are historically marginalized due to their career-level, nationality, and cultural background (6F). Database projects often relied on a community of domain experts to contribute and use data in their project, so they often sought leaders that well-represented their target audience. Most interviewees mentioned social diversity in the context of volunteer leadership positions like Boards, although interviewee 9D also discussed the importance of domain area diversity among hired leadership. Intentionally broadening diversity within a leadership group led to conflict among at least one of the sample projects (e.g., it reduced star-power leverage), but ultimately the interviewee believed that social diversity was beneficial overall for ensuring that the project met its mission as being community-centered (6F).

### 3.1.5. Leadership succession plans are a part of project sustainability

Succession plans for both paid and volunteer leaders helped avoid a leadership gap and ensured that the right people were placed in the role ( $n = 7$ ; 1M, 2F, 3M, 4D, 6F, 9D, 10D). For paid positions, new leaders were found/cultivated by recruiting through personal (1M) and professional (9D) networks. There was also the strategy of placing a potential leader into a lower-level position and then promoting them into a leadership position ( $n = 3$ ; 2F, 3M, 9D); it could take up to four years for an individual to be prepared to step into a leadership position (9D). One way to shore up a succession plan was to include two people per role, so if one person left there would be another person available to step in.

One volunteer-oriented project (4D) implemented a few different strategies to support volunteer leadership: formal groupings (e.g., working group) that help people gain an understanding of the project in a low risk way, rotating people on and off of leadership positions (3 years for each term), and current leaders nominating/recruiting new leaders.

### 3.1.6. Tensions in leadership led to crises and failures

Five of the interviewees (2F, 4D, 6F, 10D, 11M) noted that periods of crisis in their projects were due to leadership and governance conflicts. Such a crisis was sometimes due to individuals—such as a poorly selected Executive Director, a founding Principal Investigator, or unpaid community leader, whose vision for the project did not align with the rest of the group—who were looking out for their own personal interests, such as putting resources into passion projects or supporting their allies, and were not putting the mission and health of the project first. In at least one case, a former retired leader was asked to return to replace a successor that did not work out. Another interviewee believed that the lack among leaders of a consistent vision of a governance model for the project led to the eventual failure of the project.

One interviewee noted disagreements between previous leadership and project community, including about values of the project and representation among leadership. Ultimately, the grassroots constituents took on more of a leadership role, which led to systemic changes in the project including how funding was obtained and used.

## 3.2. COMMUNITY OUTREACH AND PARTICIPATION

Community is defined as stakeholders and participants in a project, including funders, end users, target audience, people attending meetings and/or contributing data to a project, and anyone else not being paid by a project to contribute to and interact within the project. ‘Outreach’ is defined as a project’s activities to reach out to community members. ‘Participation’ is defined as community members actively engaging with a project.

### 3.2.1. Projects benefit from stewarding and interacting with a community

All of the study projects had some kind of community that they cultivated, although their efforts to engage and integrate the community into their project varied. Database projects relied on their community to provide and use data. Middleware projects focused on delivering products

to their end users. Framework projects created spaces for their community to engage, develop best practices, and collaborate on work products. Other mentioned benefits of community included helping the project stay strategically positioned through better understanding of the broader landscape that their project is embedded in, as well as being able to leverage more opportunities (like funding), problem solve, and create work products (n = 3; 3M, 5F, 6F).

Community members helped provide support for project tasks (especially when the project was experiencing a lean period) (3M), demonstrated product impact through their use of a product (5F), and added features to a software to help increase interoperability and accessibility of the product (11M). Two Middleware interviewees (1M, 11M) mentioned that they would like to increase end user engagement in their development of technology.

### 3.2.2. Intentional strategies are needed for community outreach and enhancing participation

More than half of the interviewees noted the importance of having intentional community outreach strategies (n = 6; 2F, 3M, 4D, 7D, 9D, 10D, 11M). All of the interviewees that mentioned this theme (n = 6) cited an outreach strategy of meeting community members where they are at. For example, interviewees mentioned organizing meetings and workshops in locations convenient for participants, attending and hosting booths at domain area conferences that users often attend, and joining field research excursions to shadow scientists (10D). Another strategy included encouraging community members to reach out to others who might benefit from the project and community (2F).

Community participation was often the next step after outreach, and more than half of interviewees mentioned strategies for enhancing community participation in their projects (n = 8; 2F, 3M, 4D, 5F, 6F, 8D, 9D, 11M). The most frequently mentioned strategy was to create a space where experts exchanged ideas and best practices and learned from each other (n = 5; 2F, 3M, 5F, 6F, 8F). Low barriers to participation helped encourage the use of data and software services (n = 4; 1M, 3F, 4D, 9D), such as not requiring participants to log into platforms with passwords or submit credentials.

Four other strategies were the next most frequently mentioned participation strategies (n = 3 each). One strategy was incentivized participation (2F, 8D, 9D) that leveraged requirements for funded projects to participate in the project by, for example, contributing data/information to a project and/or attending meetings. Another strategy was to make the benefits of participating in a project more obvious (4D, 5F, 8D), such as by emphasizing that participation makes the participant's papers/data more accessible to others or provides them with more access to project funding. Projects hosted workshops and conferences to help people improve their use of project services (3M, 4D, 11M) (i.e., capacity building). Paid leadership directly interacted with community members in order to identify their needs and then create new products, tools, or functionality, and provide technical support and tutorials to help them overcome skills-related obstacles (6F, 9D, 4F). Finally, one interviewee (4D) mentioned that emphasizing community ownership of a project helped build trust among other potential participants, so they are more willing to participate and contribute data.

Community engagement was difficult, and failures were common (n = 3; 8D, 9D, 11M). One failure stems from when members were unclear how they might benefit by participating in the project (8D). Project participants were also often busy professionals, so it was difficult for them to prioritize a project in their workflow (9D). Another challenge was maintaining close connections with end users, including having skilled staff to coordinate and support project volunteers.

## 3.3 ASSESSING PROJECT SUCCESS

### 3.3.1 Sustainability is associated with providing an essential service

Most interviewees (n = 9; 1M, 2F, 3M, 4D, 5F, 7D, 8D, 9D, 11M) provided input on how they define sustainability. Five of these interviewees (2F, 3M, 4D, 9D, 11M) defined project sustainability as when a project becomes seen as an essential part of the overall landscape. Interviewees 4D and 11M used the term 'relevancy' as part of their sustainability concept. Likewise, three of these interviewees (3M, 5F, 8D) considered sustainability to be related to a project's long-term value, and they contrasted this notion with the immediacy of return-of-investment goals.

Finally, three of these interviewees (1M, 5F, 7D) mentioned how important it was for the overall mission of the project to be sustained. Interestingly, these three definitions were represented by all of the three subsample groups, and with almost unique groupings of interviewees.

Some of these nine interviewees offered additional characteristics of project sustainability; most of these interviewees represented Database projects. Two Database project interviewees (8D, 9D) noted that access to content was their sustainability goal. One Middleware project interviewee (3M) mentioned the linkage between sustainability and interoperability. Two Database interviewees (7D, 8D) raised the question of whether or not a project should be sustained; they mentioned that quality (7D) and impact (8D) of a project mattered most for evaluating what to sustain from a project.

### 3.3.2 Metrics of success are often related to user involvement

Almost all of the interviewees provided specific examples of metrics used to assess project success ( $n = 10$ ; 1M, 2F, 3M, 4D, 6F, 7D, 8D, 9D, 10D, 11M). The most often used metric involved the number of people using the service of the project ( $n = 7$ ; 1M, 2F, 4D, 7D, 8D, 9D, 10D), and this included people, IP addresses, and/or data centers using software and downloading / uploading data ( $n = 6$ ; 1M, 4D, 7D, 8D, 9D, 10D), and people joining the community (2F). The next most often used metrics included how much a project's software was used in applications ( $n = 3$ , 1M, 7D, 11M), the number of citations of a project's datasets ( $n = 3$ ; 4D, 7D, 10D), organizational/model duplication ( $n = 3$ ; 2F, 8D, 10D), and successful change of leadership ( $n = 3$ ; 2F, 8D, 9D). The last categories of metrics were number of data downloads ( $n = 2$ , 1M, 10D), specific products like written documents ( $n = 2$ ; 6F, 7D), responses to service issues ( $n = 2$ ; 3M, 4D), and overall service performance ( $n = 2$ ; 4D, 11M). Finally, other metrics were mentioned, such as program effectiveness and enthusiasm from funders (3M), good abstraction in software development by hiding complex details and focusing on essential concepts (3M), data quality (7D), and diversity of funding sources (8D).

### 3.3.3 Measuring project success is often challenging

Three interviewees ( $n = 3$ ; 1M, 4D, 7D) spoke of the challenges of measuring success. Interviewee 1M, whose project worked with open-source software, noted that it is difficult to measure the number of users of software. Interviewee 4D represented a PI-run project that depends on volunteers, so there were limited human resources to both assess metrics and provide oversight into the quality of data being contributed to the database by volunteers. Finally, 7D mentioned the importance of qualitative assessment of a project's success, considering that much of their work included collaboration and community building.

## 3.4 BUSINESS MODELS

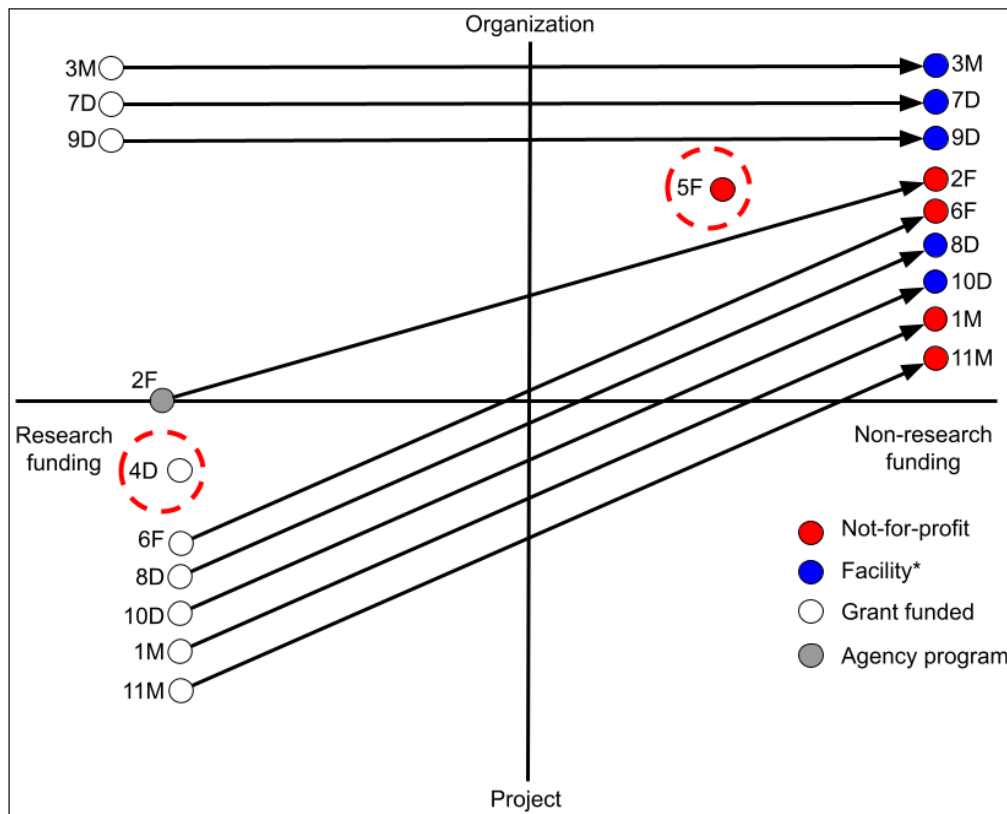
A project's business model refers to the mechanisms employed to conduct the day-to-day operations intended to fulfill the strategic goals of the project.

### 3.4.1 Organizational structure

In [Figure 1](#) the structure and funding of the 11 study projects are compared as they changed over time from inception to the time of the interview. Of those projects, nine transitioned from funding that was primarily based on research grants (limited to at most three years) to some form of long-term funding. Of those nine, four formed not-for-profit corporations (i.e., 501(c)3 companies) and five became a 'facility' either via funding that specifically designated them as such or using funding that established them as a *de facto* facility (here we mean *facility* as an organization regarded by a funding agency as providing long-term resources necessary for others to carry out research).

Two of the projects in [Figure 1](#) are outliers in regard to this notion of transition. One Framework project began as a not-for-profit corporation, and there was no change to that structure or to the nature of its primary source of funding. However, the project did increase in size and, since it derived significant funds from membership fees, that growth enhanced overall stability. The other outlier is a Database project that, at the time of this study, still continued to receive most of its funding from research grants, and there were no plans in place to change that pattern. Thus, despite not following the project-to-organization trajectory that the other projects did, these two outliers achieved funding stability, as well as overall project sustainability.

One project deserves special mention because while it transitioned funding, its origin was not formally as a funded project, but instead as a *program* that funded other projects. Since this ‘project’ started as something that was distinctly neither a project nor an organization, it has been placed straddling the boundary between these two concepts in Figure 1.



**Figure 1 Evolution of the structural changes over time of the studied projects.**

Nine of the eleven projects transitioned to some form of stable funding and all but one either formed or became part of a formal organization capable of long-term operation. Note that the exact position within the quadrants has no particular meaning, except for 2F, which started as neither a project nor an organization but as an agency program.

\*Facility is used here in vernacular sense and is not limited to the way funding agencies (e.g., NSF) define the term.

The projects have different reasons for their transitions to organizations. Two Middleware projects started as research projects within a university and left the university because of conflicts between the initial sources of funding—rooted in scientific research—and the need to fund the development and maintenance of a software technology as a distinct entity ( $n = 2$ ; 1M, 11M). The third Middleware project was able to remain within its parent organization because, in part, it transitioned from research to facility funding (both from NSF) (3M).

The transition from research grant to facility funding was experienced by other study projects ( $n = 4$ ; 3M, 8D, 9D, 10D), but it depended on convincing the funding agency (NSF in all the cases studied) that such a move would benefit the NSF and its scientific research grant recipients. One project (7D) decided at the outset to not be associated with a specific institution so that researchers would feel more at ease with contributing data to its data holdings. Another project (8D) described ‘growing pains’ associated with an increase in size and scope. Finally, one project (10D) described the complexities of managing a facility that was formed from two distinct projects. The differences in the two projects led to increases in cost and ultimately the resulting project was re-structured, resulting in separation of the two entities.

### 3.4.2 Staffing the organization

The study projects were all relatively small (30 or fewer employees). Paid staff often worked in more than one area and across domains including some who were conduits to domain-specific end users ( $n = 7$ ; 3M, 4D, 6F, 8D, 9D, 10D, 11M). Projects that focused on technology development and/or maintenance (Middleware and Database) often sought out people with better software engineering skills. At the same time, those technology-focused projects sometimes hired people who worked part-time because of the difficulty in finding or affording qualified full-time employees, as well as having enough consistent work for them ( $n = 2$ ). In many cases ( $n = 5$ ), projects depended on volunteers with other hard-money jobs. This typically manifested itself when a project leader held a full-time faculty position and volunteered to lead the project/organization.

Due to the small size of the projects, there was little repetition of expertise across staff. For many of the projects ( $n = 5$ ; 1M, 4D, 6F, 8D, 11M), staff turnover often meant that a project lost critical expertise, utility, and/or access to user groups. Staff turnover was also exacerbated by short project cycles, limited financial resources, and issues rooted in the academic environment where graduate students and postdoctoral fellows were working in low salaried, time-limited positions. Retaining staff members was also hard because staff were often highly educated/skilled people who stayed only if the work remained interesting to them.

One interviewee (3M) reported that their project was formed as part of a much larger institution that employed many people with strong technical skills (especially regarding software development), and this provided a staffing boost because there was ready access to people with hard-to-find skills. Such a partnership also provided the organization's leader with experience that would have been otherwise hard to get.

### 3.4.3 Partnerships often strengthen projects

Most of the study projects used partnerships to increase their strength or resilience ( $n = 9$ ; 1M, 3M, 4D, 5F, 7D, 8D, 9D, 10D, 11M). The partnerships these projects made can be categorized as: consortia, cross-sector, and resource pooling. Each of these approaches enabled relatively small projects to increase their effective size without the additional overhead or change in organizational dynamics that accompany an actual project change in size.

The consortia model was used by three projects (5F, 6F, 7D) and all of these were Database or Framework projects, none of the Middleware projects used this partnership model. A consortia model consisted of bringing together groups to address specific needs and/or problems identified by the project.

The cross-sector model was used by all the Middleware and Framework projects and some of the Database projects ( $n = 5$ ; 1M, 5F, 6F, 7D, 9D, 10D, 11M). In this context, cross-sector refers to different classes of projects working together, such as independent not-for-profit groups working with for-profit companies, government agencies, and academic institutions.

The resource-pooling model was used by mostly Middleware projects ( $n = 4$ ; 1M, 3M, 4D, 11M). Three Middleware projects pooled together human resources (personnel) to tackle a common problem and one Database project worked with another database system to leverage different but complementary data assets.

Related to partnerships was the use of 'side projects' to expand the organizations' opportunities. These side projects provided extra funds that helped cover operating expenses.

Not all partnerships were reported as successful. One project, as mentioned before, described a partnership intended to reduce management overhead through resource-pooling. However, because the projects were not otherwise well matched, the partnership failed (10D).

### 3.4.4 Funding Models

#### 3.4.4.1 Grant funding

Most of the study projects began with grant-based or similar sponsored funding ( $n = 8$ ; 1M, 4D, 6F, 7D, 8D, 9D, 10D, 11M). For the remaining three projects, one project began as an effort within a large grant-funded NSF facility (3M), one started as a grant-funded program within a federal agency (2F), and one started as a not-for-profit organization (5F). Thus, all of the study projects had their origin closely tied to the research grant funding culture with one exception among the Framework projects (5F, the sole organization that formally considers itself a consortium).

For the projects that depended on grant funding, the short-term nature of the funding (typically 3-year cycles; sometimes 3-year then 5-year cycles) was a significant impediment to their transition to an organization. For many of the projects, the time needed to become mature and more sustainable was significantly longer than the grant funding cycle.

The tension between the short-term nature of most grant-based funding and long-term needs divided the study project into two groups: those that continued to receive the majority of their funding from grants ( $n = 7$ ; 2F, 3M, 4D, 7D, 8D, 9D, 10D) with five ( $n = 5$ ; 3M, 4D, 7D, 8D, 10D) receiving funding *only* from research grants, and those that diversified to use funding from other non-grant sources ( $n = 4$ ; 1M, 5F, 6F, 11M). All of the Database projects were funded primarily



through grants from agencies that typically funded research. Middleware and Framework projects tended to move toward other forms of funding.

All 11 projects reported sustainability concerns regarding their reliance on research grants, even though seven continued (at the time of the study) to receive the bulk of their funds from either research grants or agencies that primarily funded research. For those that received the bulk of their funding from research grants, a couple ( $n = 2$ ; 3M, 7D) reported forming close ties to the funding source to better understand funders' needs.

The nature of the NSF research grant funding, as perceived by several projects ( $n = 4$ ), has changed over time. From one perspective, the NSF has moved away from funding data centers and instead looked for different ways to ensure data are archived and accessible (2F). From another perspective, the NSF was originally formed for the purpose of ensuring that nuclear physics research would be supported. Other disciplines, especially the Earth Sciences, require significantly different funding models (7D, 10D). At the same time, the NSF was seen as willing to take risks on technology development and provide support where positive outcomes are far from certain (3M).

#### 3.4.4.2 Alternatives to grant funding

The study projects that did not rely predominantly on grant-based funding ( $n = 4$ ; 1M, 5F, 6F, 11M) pursued a number of funding strategies including donations and membership fees; selling services and products (which included working as contractors); funding from private foundations; and registration fees from running meetings. Most of these approaches were insufficient for providing the majority of funding needed for project support although both of the Middleware projects did receive sufficient funds from contract work for the periods when those contracts were in effect.

Individual-level donations and membership subscriptions did not result in sufficient funds to justify the effort ( $n = 3$ ; 1M, 4D, 6F). One project successfully used an institutional-level membership model as part of their funding stream (5F), while another tried this model and abandoned it early on (7D). Thus, of the 11 projects studied, only one received the majority of its funds from membership fees and those members were mostly large organizations and government agencies (5F). No study project received significant funds via donation.

Private foundation funding helped only to fill gaps ( $n = 4$ ; 4D, 7D, 8D, 10D), so it was not a majority funding source for any of the study projects. Only Database projects reported any experience with private funding. Six of the study projects reported selling services ( $n = 4$ ; 1M, 11M, 5F, 8D) and running meetings to obtain a significant fraction of operating funds ( $n = 2$ ; 2F, 6F) once they stopped relying entirely on research grants.

#### 3.4.4.3 In-kind support

Large host organizations, including academic institutions like Carleton College and University of Rhode Island and non-universities like the University Corporation for Atmospheric Research (UCAR) and the National Center for Supercomputing Applications (NCSA) (see [Table 1](#)), played key roles in helping projects get started or reducing the administrative costs for small projects by providing business support, such as help with taxes, payroll, and legal issues.

In particular, three projects reported significant in-kind support from large 'parent' organizations (1M, 4D, 9D). Two projects described themselves as benefiting from university faculty with full time employment (1M, 4D), either because a tenured faculty member led the project or by depending on volunteers with other employment. The nature of tenure at an academic institution played a major role in these two projects because the faculty were effectively upper-level management. That is, their grants did not have to explicitly request funding for management costs. Without this in-kind 'contribution,' there would have been no other way to maintain the kind of management the projects needed.

#### 3.4.5 Funding and uncertainty

The lack of financial reserves was a major weakness for most projects because it impacted their ability to retain staff and withstand major shocks ( $n = 6$ ; 1M, 4D, 6F, 8D, 10D, 11M). Four of the studied organizations reported financial issues that nearly caused the projects to shut down ( $n = 4$ ; 1M, 6F, 8D, 11M).



Two Middleware projects found that it was impossible to build financial reserves using grant funding alone (n = 2; 1M, 11M). Both projects tried to build a 'tax' into their billing process to generate working capital, but discovered this was not allowed when using grants from the US government. They ultimately used contracts with for-profit corporations to generate the cash reserves they needed for financial stability.

At least one project used short-term bridge loans from an individual to cover periods between when one grant or group of grants ended and another set of grants started (1M). These loans enabled the project to maintain its payroll during funding gaps.

A number of projects existed as part of a larger 'umbrella' institution that helped provide financial stability (n = 4; 3M, 8D, 9D, 10D). The projects that remained within a larger institution were mostly Database projects and functioned as if they were NSF facilities (although only a subset actually received that type of funding). Even with the support of a larger institution, maintaining sufficient funds to pay staff was an ongoing challenge for all projects.

### 3.4.6 Value-proposition

The study projects discussed the role of business value-proposition and sustainability in the context of interoperability and gap filling.

#### 3.4.6.1 Interoperability

Nine of the projects (n = 9; 1M, 2F, 3M, 4D, 5F, 6F, 8D, 10D, 11M) cited building infrastructure to support interoperability with the social frameworks and/or technical systems (software or databases) developed by similar projects as part of their enduring value-proposition. All three of the Middleware projects were represented here, and they mostly described building infrastructure by developing (software) technology. The Framework and Database projects talked about community building and their impacts as developing infrastructure to support science.

#### 3.4.6.2 Gap filling

Four projects described filling gaps in Earth Science and then remaining on the leading edge of providing that service (n = 4; 2F, 3M, 6F, 8D). As with interoperability, gap filling was shared by all three types of projects. Each project initially perceived that an important type of support for the Earth Science community was missing or the problem was explicitly called out in review documents developed by government agencies.

For three projects (6F, 7D, and 8D), gap filling was accomplished by bringing together members from different communities, either directly (n = 2; 6F, 7D) or indirectly (8D), such as by having people who had previously been staff members join the larger community in different professional capacities, thus filling knowledge gaps in those organizations. Four projects started with the goal of solving an explicit problem and this led them down a path that made them both successful in the short-term and influential in the long-term (2F, 3M, 5F, 8D). These projects noted that timing and technology likely contributed to their ongoing sustainability. The immediacy of the initial problem they set out to solve served as a catalyst. Several projects described that they ultimately formed a community where people could expect to come and solve (new) problems.

Along these same lines, four projects reported that capacity building was an important part of their value-proposition (3M, 4D, 7D, 8D); no Framework projects called out this specific item. In this context, capacity building meant expanding the knowledge base of members so they could take advantage of new opportunities. In effect, the projects taught their members new skills.

## 3.5 STRATEGIC DEVELOPMENT

### 3.5.1 Staying mission-focused helps to navigate forks in the road

The majority of interviewees (n = 8; 2F, 3M, 4D, 5F, 6F, 7D, 8D, 11M) discussed encountering a major decision-making point during the life of the project.

Some projects were torn between building technology ('a platform') and organizing their members (2F, 3M, 6F). These choices often had strong advocates on both sides and while 3M did ultimately build important technology, it also built a strong community. The two frameworks that explicitly mentioned this choice (2F, 6F) ultimately did not build a technology but instead

focused on community organizing including the formation of working groups or the equivalent. In addition, interviewee 5F described how their project sought to stay their course by providing robust climate data (rather than focusing on profit margins).

Among the Database projects, interviewee 4D's project decided not to limit their database to specific sub-domain areas, interviewee 7D's project focused on the development of facilities rather than on scientific goals, and interviewee 8D's project chose not to switch to a common content management system and instead continued to improve their unique community authoring system. The two Middleware interviewees discussed the importance of staying mission-focused overall when faced with a strategic challenge.

Two interviewees (2F, 3M) described failed paths that their project took. One example included attempting to develop a specific type of software that would be interoperable on different computers, but the effort was deemed to be less successful than expected.

### 3.5.2 Being an early pioneer is a way to capture new opportunities

Most interviewees ( $n = 7$ ; 2F, 3M, 4D, 5F, 7D, 8D, 11M) described how their project was breaking new ground. Both 3M and 4D described doing things 'by the seat of our pants.' Interviewee 2F noted the development of a new research domain, namely: 'data as its own research, as a co-equal to oceanography or meteorology is something that has really come into its own in probably the last 20 years.' Other examples included the novelty of building a library of software tools (3M), open access to instruments and data (7D), ability to develop international collaborations (7D), and distributed tools for website development (8D).

### 3.5.3 Leaning into the unexpected encouraged innovation

Five of the interviewees (2F, 3M, 5F, 6F, 8D) discussed how their projects evolved in unexpected and unintended ways. The Middleware project interviewee (3M) discussed evolution in a broad way. The Framework projects leaned on their communities to lead the way. One project had not expected their community to become the main feature of their project. Another project's community 'bubbled up' unexpected goals for the project. Finally, yet another project's community leaned into social equity in a way that changed the overall mission and structure of the project.

Among the Database projects, one project noted that cloud-based collaborative workspaces really challenged (negatively) their project since they had heavily invested in developing their own system for managing such collaborative website development. Another Database project mentioned that their end user group was broadening, such that they are not sure if their model could be scaled up fast enough to be able to provide services for them.

### 3.5.4 Staying relevant is a challenge for technologically centered projects

Relevancy was particularly important for Middleware and Database projects ( $n = 4$ ; 3M, 9D, 10D, 11M), because they were challenged to stay up-to-date in the face of a quickly changing broader technical landscape. With limited resources, projects often contend with the question of: 'where do you invest in keeping with the state of the art and where are things becoming, if not obsolete, they're not necessarily the wisest choice of how you use the limited resources you have' (3M).

Technical debt eventually catches up with any software that is in use for a decade or more. To overcome this debt, considerable funding and resources are needed. For example, one Middleware project estimated that 3–4 years are needed to create a new version of a flagship project software. Therefore, projects often chose instead to use a scaffolding approach, which consisted of patching up software. In other words, 'by doing the tweaks you bring more people under the tent' (11M). At the same time, there was sometimes an expectation that a project maintains legacy software.

### 3.5.5 Competition is inevitable

Three interviewees (1M, 3M, 6F) specifically discussed competition within the Earth Sciences and how that both helped and hindered their work, and ultimately affected sustainability. Competition was seen as either inevitable or desirable. Both 1M and 3M felt that competition

was not a zero-sum game and instead used it as an opportunity to develop partnerships. The Framework project actively sought to minimize competition among their members because they believed it would threaten the group dynamic of equality.

## 4. DISCUSSION

The success and failure of organizations, and why they occur, have been extensively studied in fields like organizational economics, business management, and behavioral sciences. Topics of study have ranged widely, but include organizational learning (Saadat & Saadat 2016), leadership composition (Charan et al. 2013), decision-making (McMillan & Overall 2017), and risk management (Durst et al. 2019). While most of these studies have focused primarily on for-profit business, nonprofit organizations are also an area of study. With nonprofits, issues like governance (Molk & Sokol 2021), volunteer motivation (Ainsworth 2020), and overhead spending (Altamimi & Liu 2022) are additional topics of interest.

While existing studies on organizations have some relevance to our study, our results indicate that the sustainability of organizations that support Earth Science data infrastructure presents a unique case. One main difference is that science thrives on cooperation between individuals and groups. Scientific conferences, for example, where people regularly share and test ideas, are a key aspect of science. As such, transactional behavior, which underlines for-profit businesses, is often not a good fit for science initiatives. While nonprofit organizational structures might be better aligned with scientific endeavors, Earth Science data infrastructure is strongly tied to governmental resources. Indeed, Earth Science data is often viewed as a public good, and universities often provided crucial in-kind support for many of the projects that we examined.

Another unique quality of our study projects was the type of people involved. Many of the key people in the projects and members of the communities involved with the projects represented some of society's most educated individuals. This presented an advantage for the development of leading-edge ideas, but also introduced instability in terms of human resources. As interviewee 8D described the situation: 'really interesting people only stay as long as the work is really interesting.'

Studies suggest that the sustainability of community-based programs presents a special case. Community-based programs have high community involvement, require socio-cultural acceptability within the relevant community, and seek to achieve a long-term presence. When addressing sustainability within community-based programs, the quality of the programming is often most important (Ceptureanu et al. 2018). For community-based software development, which is particularly relevant to our Middleware projects, governance issues are particularly important. Core processes, like how contributions to software are made, are often formalized. More informally, there is a high concern for the socialization of members and their feeling of belonging within the project (Ferraz & Santos 2021).

A main finding of our work was the notable differences among the three subsample project groups of Framework, Middleware, and Database. Each of these project groups is an important part of the broader Earth Science data infrastructure landscape. The purpose and goals of these different subsample groups impact how each project approaches sustainability. The subsample groups represent a range of projects. Some projects are extreme examples of a subsample group, while others lean towards and/or overlap with another group (see Figure 1).

Framework projects are mostly aimed towards organizing a community (e.g., bringing people together). Since these projects prioritize the community and their needs, they often give as much decision-making power as possible to their community. Such an influx of diverse perspectives can push a project towards edgier ideas. At the same time, however, Framework projects are open to higher risk if the community is not well managed, coordinated, or aligned. For example, Framework projects can be diluted by too many opinions and self-appointed leaders, be easily swayed by certain factions of the community, and be led in a different direction than what was originally intended by a project.

Database projects are focused on serving the data needs of a community, often with the goal of supporting more efficient and broader use of data through sharing, storing, and curating datasets. Database projects present a middle ground between Framework and Middleware projects. Database projects often have stronger top-down leadership than Framework projects,

but they also depend on more community participation than Middleware projects. A participatory approach is commonly used among Database projects because they depend on researchers to voluntarily deposit and use data from a shared database. A participatory approach has the benefit of being able to involve a large number of people in the data collection at a low cost, while making sure that the end products meet the needs of the users. It can also help increase the diversity of different research initiatives that are able to get off the ground and find success (Fleming et al. 2023). Databases are challenged to sustain management and funding, because they are often domain specific; therefore, they heavily rely on support from their research domain (participation from researchers, sustained funding from research funding agencies).

Middleware projects focus primarily on the development of software. They often use a traditional for-profit business structure, while being closely tied to scientific needs. Middleware projects are challenged to balance both fast-paced, highly technical software development and often slower relationship-driven sciences. The rapid pace of technological innovation, which can often take the form of disruptive innovation (Christensen 1997), can pose special challenges in management. Middleware projects are generally top-down governed, such as by a board of directors or advisors and a Chief Executive Officer (CEO) plus Chief Financial Officer (CFO) model with hired staff. For Middleware projects, the community component of their work often presents as ‘end users,’ who provide one-way feedback on their products. Because they are working with the development of high-investment and highly technical projects (software development), Middleware projects often cannot afford the slower, unpredictable nature of community-based decision-making. For Middleware projects, partnership with other entities can take the shape of cooperative development where people working for one project do critical work for another project. The studied Middleware projects all leveraged externally developed open-source and commercial software because they could not exist without those connections.

#### 4.1 GOVERNANCE AND LEADERSHIP

Among all three groups, a governance structure that is explicit, intentional, and flexible to change is critical. Developing an appropriate governance structure for the project type is also key. For example, the Middleware projects cannot be run as if they were Frameworks and vice-versa. All of the subsample groups valued the documentation of decision-making (ranging from governance to technical development), so a team can work collaboratively and effectively.

For all of the subsample groups, good leadership is highly valued. The projects were challenged to find leaders that balance both technical and scientific domain expertise, as well as management and business competence. Getting this balance wrong resulted in near failure for some of the studied projects. Leadership succession plans were noted to be integral to project success. For at least one project, it took multiple years to train and prepare a selected leader.

**Differences among project types.** The Middleware projects studied all used fairly conventional hierarchical internal governance structures but also blended into those ‘lateral’ community-based governance. This matches the observations of Ferraz and Santos (2021) and the reasons listed therein: that hierarchical governance provides an effective way to ensure the needed characteristics of software—such as ‘correctness’—while the community can provide direction based on their members’ needs.

Governance was discussed much more among the Framework and Database projects, than by Middleware projects. This is likely because among Framework and Database projects governance is critical for bringing their community’s needs, perspectives, and participation into the project.

Framework and Database projects considered the advantages and disadvantages of different governance practices, particularly for bottom-up (more community participation) versus top-down (more executive decision-making) governance approaches. Framework and Database projects draw on community members for essential aspects of their decision making (e.g., via a membership-elected president) combined with some governance controlled and/or managed by paid personnel. This combination helps to provide stability, while volunteer members provide a direct conduit between the project and their community. Framework and Database projects sometimes create smaller groups for volunteers, such as an advisory board, to allow for more decentralized decision-making. With Framework and Database projects having changing and inconsistent volunteer leadership, institutional knowledge is easily lost, so documentation is particularly important.

The tension between top-down and bottom-up governance is sometimes acutely felt among the Framework and Database projects. In at least one project, this tension led to great disagreement in the project (e.g., the constituents disagreeing with the vision of top-down leaders). The main takeaway is that both directions of governance must be well-balanced, depending on the goals of the project. On the farthest end of the spectrum, low community involvement consists of having mechanisms for the community to provide opinions that were considered by executive decision-makers. High community involvement means greater rotation of individuals in the decision-making positions, and true decision-making power among all of the people involved in the project (e.g., voting among all of the constituents).

For example, one Database project was created and managed by a core group of users themselves (researchers), because they felt that their discipline needed what the project could provide. One downside of this approach, however, is that the in-kind support provided by participants can be interpreted by funders as self-sustaining; thus, it is challenging for them to justify why they need more research funding to support their work. Projects leveraging participatory approaches often find it challenging to gain access to funding, although recommendations suggest that funders could support such projects better if they take on roles as partners, rather than as top-down managers (Minkler et al. 2003).

For projects with high community involvement, such as for Framework and Database projects, governance structures evolve over time so they can be molded and customized to the needs of the project. It often takes at least a few years to clarify the needs and vision of the community and the mission of the project, so creating a governance model too quickly can be detrimental to the project.

Among the Framework and Database projects, leaders are both paid (staff) and unpaid (volunteer). With this type of model, it is necessary to parse out what is appropriate work for paid staff versus for volunteers. It is often necessary to ensure that volunteers are motivated to participate by doing things that are of interest and of benefit to them. At the same time, volunteers are often more than just advisors and advocates—they provide direct benefits to a project by lending a hand and helping a project meet its goals.

Because Framework and Database projects often represent a constituency, it is important for their projects to also represent this diversity among their leadership. This is especially important among Database projects, most likely because they rely on the long-term and consistent commitment of their community to deposit and use data. In contrast, Middleware projects do not emphasize broad participation in their decision making.

## 4.2 COMMUNITY ENGAGEMENT

The study projects focused their community engagement primarily on people who could directly contribute to, participate in, or otherwise influence their project. In the Earth Sciences, such community engagement activities are often associated with network development, capacity building, training, convening, and topic-based working groups. These types of engagement activities differ from outreach in sciences that seek more one-way directional influence on the public, such as for the purpose of influencing social change (Blue & Davidson 2020). Most of the study projects believed that successful engagement meant meeting users where they already are, and making sure that the projects had a full understanding of domain practices (e.g., hiring people from a target community, like a domain scientist). Study projects' community engagement goals were often aimed towards more participatory approaches, where users became directly involved in the project (such as through governance structures) or developed collaborations with each other through the project (such as through working groups).

All of the subsample groups sought to move from outreach to more active participation. Participation in projects can range from sharing knowledge, using the project's platform/service, contributing data, recruiting other users, volunteering their work effort (e.g., crowdsourcing), and helping to make decisions. Overall, community/end user participation in a project shows that the project is useful and essential. In other words, their participation justifies why a project should be sustained. The most successful projects developed a growing group of motivated, engaged, and devoted participants, and had a clear value proposition for their community.



For the Framework and Database projects, the opportunity for members of the community to have some level of ownership and control over the projects is critical to the projects' value. This is true on several levels, including the increased ability to control the direction of critical infrastructure relevant to a member's field. In addition, there is a social component to membership and the deeper personal connections that intensified involvement in the project provides. In both cases, the value proposition of the project includes benefits that may be hard to define *a priori* for any given individual. Importantly, these study projects often provide a unique place for people and groups to gather, convene, and share ideas that are crucial for the progression of science and scientific technology.

**Differences among project types.** Each of the subsample groups had different asks for their community. Framework projects mainly asked their community to interact with each other, Database projects mainly asked their community to contribute to their database, and Middleware projects asked for feedback about their products. Both the Framework and Database projects depended on members for essential aspects of their work. For example, Database projects must build up a large and relevant data store and all of those studied here did so by drawing on their community members to provide in-kind support. Without volunteers to provide data, Database projects would not work.

### 4.3 ASSESSING SUCCESS

Hsu et al. (2019) defined sustainability in the digital Earth Sciences as an impact that persists beyond the life of a seed-funded project, and sustainability can occur on individual, organizational, and community levels. In our study, we found that sustainability was often interpreted by interviewees in two ways: sustainability of the project, and sustainability of the product/mission. Sustainability of the project was primarily around how to maintain the operations of the project over the long-term. Sustainability of the product/mission was about the product/mission of a project continuing to have life outside or beyond that of the original project.

All but one ('Code repository used') of the seven influences on sustainability used by Hsu et al. (2019) appeared prominently among our sample group (other influences were: Outputs modified, Champion present, Workforce stability, Support from other organizations, Collaboration/partnership, Integration with policy). 'Code repository used' appears to be specific to projects that develop a tangible product (like Database or Middleware projects).

Since study projects mainly sought to be useful to their target audience, most projects focused on measuring and assessing the use of their products and services. Common approaches were quantitative metrics based on use (e.g., number of users, downloads, etc.) and/or perspectives by users (e.g., surveys asking users about their experiences). However, these results can be ambiguous since downloads do not equate to use and long-term users can be hard to differentiate from causal non-committal users.

Measuring success is needed to justify a project's continued existence, so it can receive the funding and community support it needs to operate. This tug-of-war between the demand on resources placed by short-term needs and long-term requirement for metrics was cited by three of the projects.

Two interviewees (both from Database projects) questioned whether or not a project should be sustained at all. Indeed, federal research funding from agencies like NSF often require projects to have a sustainability plan. Perhaps, we might consider that some projects are meant to just simply be projects. In other words, after meeting their project goals there is no need for continued sustainability.

**Differences among project types.** In comparison to Framework projects, Database and Middleware projects have stronger interpretations of what sustainability means for their project. Database and Middleware emphasize that sustainability is linked to being seen as an essential service.

While being viewed as essential is a great benefit to the project, the flip side is that its existence can be taken for granted. In other words, project leaders sometimes feel trapped into maintaining their project's existence. Yet, these projects might struggle to find sustained funding and support. Database projects are particularly susceptible to this—they help to solve a science maintenance issue that is often outside the scope of 'research' funding. There is also



sometimes a feeling of a lack of control over the project, particularly among interviewees who feel that they are unable to choose which parts of the project they feel most compelled to keep driving forward.

Metrics of success are emphasized more by Database and Middleware projects, in comparison to Framework projects. Database and Middleware groups are mostly concerned with measuring use of their product/service, while Framework projects consider more process-oriented metrics, such as the successful change in leadership or documentation of decision-making.

#### 4.4 BUSINESS MODEL

In digital Earth Science, short-term funding of 6 to 24 months is a common way to test the value and staying power of an initiative (Hsu et al. 2019). Likewise, 10 of our 11 case studies began as study projects, and nine of them underwent a transformation from a grant-funded project to an organization. One of the subjects started out as a formally defined organization. This organization was formed as a 501(c)3 (not-for-profit) corporation and started with a formal business model based on membership—two features unique to the study group (6F). One of the database projects can be seen as an organization although it lacked a formal structure as such and at the time of the study it remained primarily grant funded (4D).

For most of our sample group, the transition from project to organization was accomplished in two ways. Most database projects became organizations that relied on ‘facility’ funding, which is funding from a granting organization not associated with research per se but intended to support a general class of research projects. In some cases, this was formally recognized while in others it was not. Regardless, the effect is the same. These projects have enough financial stability to provide the means to set and work toward open-ended objectives. The second group, which is composed of Middleware and Framework projects, all formed not-for-profit corporations (with notable exceptions). These are all formally incorporated as 501(c)3 organizations, which simplifies forming long-term objectives while enabling participation as a principal investigator in research grant funding. Most of the 501(c)3 projects use a mix of funding that includes a minority component of research grants, often in collaboration with one or more people that represent one of their stakeholder groups.

The financial transition from grant-funded to a more diverse portfolio of revenue streams is often one of the greatest challenges for digital projects, as was the case for all but two of the studied projects. The project represented by 4D remained largely funded by research grants at the time of this study. The project represented by 5F was never funded that way since it started as a consortium project run by a not-for-profit corporation. Common diverse revenue streams used by digital projects include direct pay by participants, consulting/contracting, corporate sponsorship, host institutional support, membership models, licensing, pay-per-use, philanthropy, and subscription (Maron 2014). While these revenue streams might be useful in the for-profit world, most of these mechanisms—and particularly those that were the most transactional—were a poor fit for the context of our study group.

As depicted in Figure 1, there is one Database project that had not made the transition from project to organization at the time of the study. Unsurprisingly, it is the project that also struggled the most with sustainability, both in financial support and human resources. While it appeared to be an outlier in our study, it is not an outlier among projects in digital Earth Science. Indeed, many Database projects struggle to find the resources that they need to provide increasingly essential support for science research: a domain-specific database that allows researchers to share and reuse scientific data. Such Databases often depend on the in-kind support of scientific communities, as such activity typically goes unfunded. Introductory remarks at a membership meeting for IRIS provides one example of how funding can encourage the shift of a project towards a facility, while their activities with community building and collaboration may be less valued by funders (Simpson 2013).

The projects studied are structured around all or a subset of the following: service, interoperability, collaboration, academia, and community-based approaches. These characteristics can make a project a poor fit for conventional for-profit business models. Thus, when the study projects tried to apply models used successfully by for-profit businesses, it was not enough for success. In one case, there was substantial financial loss and organizational instability. Techniques like fee-for-service and membership fees were difficult for the study projects to manage; they did

not often lead to sufficient funds to run the project alone. Indeed, the culture of academia and science, which is highly collaborative and dependent on in-kind support, is not often receptive to *quid-pro-quo* approaches. These revenue generating mechanisms, however, can augment other sources of funding, and the combination can yield sufficient funds for stability.

Research grant funding sources are not well suited for projects like the ones we studied. Because these study projects are intended to be run long-term, they do not fit the commonly used research model where a study is performed for a finite period (the true definition of a ‘project’) capped with the publication of results and recommendations for further work that will take place as part of a subsequent project. On the other hand, a sustained data infrastructure project, like a corporation or a research program, is intended to exist until it is no longer useful, with no fixed end date.

All but two of the study projects began with a grant from a federal funding agency that primarily funds scientific research projects. Most of the study projects engaged in mixed funding models in order to leverage the strengths of one to offset the limitations of another. How the projects managed this vary by project type. Middleware projects generally mixed contract work from both government agencies and private companies, although one Middleware project is almost solely funded by the US National Science Foundation (NSF). The studied Database projects tend to derive the bulk of their funding from NSF. Framework projects use a mix of grants spread across agencies along with private funding sources, membership fees, and meeting fees.

Geological/Earth Science data can be so highly heterogeneous and specialized that ‘silos’ often unintentionally form between scientific domains. All three studied project types generally aimed towards interoperability as a way to break down those silos and remove barriers to interdisciplinary research. However, Middleware projects tend to focus on powerful computer- and information-science abstractions for common models that can straddle many disciplines. Framework projects tend to focus on social structures that bring together people from many backgrounds. Database projects tend to focus on building data stores of high-quality information that may be integrated with other database projects. This ‘interoperability driver’ is a major reason the three types of projects adopted different governance models. It is notable that for these projects, interoperability is an asset—not a problem to be solved.

For all of the studied organizations, partnerships and collaborations are essentially a structural component of their business model and not simply a nice-to-have addition. Despite the limited funding resources available for projects, projects must still maintain a generous collaborative mindset because science-based projects rely on partnerships. Open-source software development, in particular, pushes forward a culture of sharing and must live up to their message—despite essentially creating their own competition. These qualities might make these projects unique in the world of business. While many businesses engage in both, all of the studied projects *prioritize* and put considerable resources into collaboration and partnership endeavors. In fact, most projects are not structured to stand completely on their own, and are instead part of tightly coupled networks, relying on partners for critical parts of their infrastructure.

## 4.5 STRATEGIC DEVELOPMENT

Many of the studied projects attempted to do something that had never been done before, so there was high risk involved. All of the subsample group projects felt the need to stay focused on their missions. It is enticing for projects to veer away from their mission, but with their limited resources, such distractions can be detrimental to the project. While a couple of studied projects mentioned taking business risks to pursue different innovations, most mentions of failed paths and near failure were tied to leadership and interpersonal issues within the team. These failures were often due to misalignment between the personal goals of a leader and that of the overall mission of an organization. Such conflicts have also been noted in studies on organizational economics, where organizational failure has been attributed to leaders focusing too much on short-term problems or not knowing how to address issues, while lacking the communication skills to help gather the support that they need (Garicano & Rayo 2016).

Uncertainty played a significant role in the challenges these projects face. The main sources of uncertainty are rooted in funding, that is, the need to obtain sufficient funding in the immediate term and to develop ways to enable dependable long-term funding. Closely related to funding uncertainties are staffing issues. All of the interviewees reported staffing issues as among

the most difficult. For these small organizations, the loss of one staff member meant critical organizational knowledge was lost. These two types of uncertainty form a feedback loop that can intensify the risk that losing one can lead to the loss of the other—effectively compounding the magnitude of the loss. Periods of uncertainty also often occurred during the transition out of project initiation, which is often a very risky time for projects (Skinner & Drummond 2022). This is one aspect of the projects that is in complete alignment with challenges found in for-profit businesses.

Despite all of these challenges, it was notable that most interviewees embraced an abundance mindset, where the field of Earth Science data was seen as full of possibilities and opportunities for many. As opposed to viewing competition as a zero-sum game, an abundance mindset frames competition as a way to produce multiple winners. Such a framing fits well with the culture of sharing and cooperation that underlies most Earth Science endeavors, and particularly for those projects that are involved in open science frameworks, which most of our study projects were.

**Differences among project types.** Keeping a project relevant is an issue that is more important to Database and Middleware projects, since they rely more heavily on technology than Framework projects. Since technology evolves quickly, they are faced with the challenge of staying technologically relevant (usable) while also broadening their user group (interoperability) and innovating (creating new technologies). Middleware projects need to understand and anticipate technological evolution. Technology is expensive to develop and sustain. In comparison, Framework projects are fairly inexpensive to maintain and not typically limited by time budgets; resources can go a much longer way when devoted to community-oriented infrastructure. Database projects can depend on the relative stability of scientific disciplines and a panel of experts to ensure their data holdings have the quality, attributes, and accessibility needed by their users. Framework projects maintain relevancy primarily by incorporating members as major players in the governance process.

## 5. CONCLUSIONS

In this study, we examined the commonalities among long-lasting Earth Science data infrastructure projects. While many studies on the successes and failures of organizations do exist, our study provides a unique contribution to this field by providing a study on Earth Science data infrastructure that was primarily conducted by and for members of the Earth system science community. To that end, our study focused on questions that our community cares about, and we looked at organizations and talked to leaders who are highly regarded in our field.

To achieve sustainability, a project must typically transition from an initiation phase (project) to a more formalized operations phase (organization). This transition is often a difficult inflection point, and a flexible governance structure helps to support successful change. During the initiation phase, most data infrastructure projects are developed by individuals or small groups of individuals with a shared vision. Grant funding, in-kind support, and star-powered leadership is sufficient and ideal for the project stage. However, for the organizational phase, new leadership that focuses on execution and more stable funding for reaching long-term goals is needed. Almost all disciplinary scientists who were also project leaders are often forced to ‘learn on the job’ with respect to business models and community engagement.

Here, we summarize other key points of this study:

- **What is sustainability?** There are two parts to sustainability: sustainability of the project; and sustainability of the product/mission. Recognition of this distinction often proves to be critical in moving forward with long-term planning and supporting sustainability for projects.
- **Middleware, Framework, and Database project types.** There were significant structural differences among Middleware, Framework, and Database projects but they also faced similar obstacles. Their common struggles point toward ways that science funding agencies could support their growth and sustainability.
- **Leadership evolution.** For projects that are science-driven, practicing scientists play major leadership roles in the initial stages of the projects. As projects matured, different types of leadership were sometimes needed; for example, leaders skilled with building

communities were needed if a project depended on community engagement. The most successful projects were able to identify the right leadership at the right time in the project.

- **Flexible governance structure.** None of the studied projects began with a formal governance model. Instead, each project adopted a governance model over time. This approach worked because the identity, intentions, and community base were still unclear and evolving at initiation of the projects.
- **Projects that do not transition to organizations are most at-risk.** Part of the definition of a project is that it remains in a short-funding cycle with no long-term business model. One of our Database case studies effectively demonstrated the vulnerability of a project, since at the time of the study it had not made the transition into an organization. It is also likely, however, that this case study is not an outlier for Database projects in science. There is an inherent fragility associated with Database projects, which operate on unstable funding without the explicit backing of either major scientific societies or federal funding agencies.
- **A project's value is closely tied to their community of users.** Middleware and Framework projects tended toward a more diverse range of disciplines than the Database projects, where the focus on a specific field was more pronounced. Framework and Database projects spent significant resources on building community trust. Database projects in particular require engagement by trusted disciplinary scientists; their governance of database systems often included an advisory board made up of community volunteers. Framework projects were effectively inseparable from their community and delegated significant aspects of governance to that community. Community was defined in different ways depending on the type of project, ranging from end users, stakeholders, and collaborators. Community played varying roles in each project, depending on the goals and business model of the project. Roles for the community ranged from participating in governance, providing a financial base (e.g., membership fees, writing grants), and contributing in-kind support (e.g., adding data to a database). The most successful projects developed a growing group of motivated, engaged, and devoted participants, and had a clear value proposition for their community. All projects began with an innovative idea and/or critical development that fulfilled user needs.
- **Middleware projects are outliers in science.** Middleware projects' governance often resembled a for-profit corporation. They were often challenged to balance the contrasting needs, culture, and requirements of science and technological development; these two domains are often at odds with each other. While Middleware projects often began in the university/academic environment, they often transitioned away from that environment in order to overcome limitations such as staffing and financial reserve building.
- **Research funding is a poor fit for projects once they become organizations.** All of the projects faced existential issues with funding and developed various ways to sidestep the three-year research grant cycle, even though most of the projects were initiated using that funding mechanism. Despite addressing issues that are essential for science today, the projects often have long-term goals that the research grants were not originally designed to support. Funders take risks to help initiate these projects, but sustaining them through this same source of funding is usually not within their scope. At the same time, however, some projects in this study demonstrate that funders will sometimes sustain an essential project (creating a 'facility'). Most projects 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 agencies. Each project spent significant effort finding ways to fund their digital initiatives.

Data infrastructure is becoming increasingly more crucial for both research and knowledge sharing. Database and Framework projects are particularly at risk, because there is no established 'onramp' for emerging projects to become sustainable after development. In part, this situation is occurring because we are at a unique crossroads in science practice and reporting. It is clear that data and digital products are necessary for today's disciplinary science work. However, we have not yet developed pathways that allow disciplinary communities to create the essential digital infrastructure they need to support science, research, and technological development.

Due to the informed consent that human subjects of the research agreed to, we are unable to release the data associated with this study. However, the research tools (detailed methodology and interview template) are publicly available in this technical report: <https://doi.org/10.6075/JOJH3MBN>.

## ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix.** This document consists of quotes from project interviewees to help support the results presented in the paper. The quotes are organized to correlate with the different sub-sections of the results. DOI: <https://doi.org/10.5334/dsj-2024-014.s1>

## ETHICS AND CONSENT

The research included human subjects. The study protocol was reviewed by the Institutional Review Board (IRB) at the University of Wisconsin – Madison and considered ‘exempt’ from further review. In keeping with the intent of the IRB, we utilized informed consent with the interviewees.

## ACKNOWLEDGEMENTS

This project was an outgrowth of internal discussions from the Council of Funded Projects of the NSF EarthCube initiative. We thank all of the participants in the Council of Funded Projects. The group who developed the research approach, discussed the interviews, summarized the results, and contributed to an internal report for EarthCube (Virapongse et al. 2022) were Peter Cornillon, Brooks Hanson, Rebecca Koskela, Susan Shingledecker, Chad Trabant, and the authors. In addition, Mohan Ramamurthy and Suzanne Pierce helped to develop the research approach. We want to thank Leslie Hsu and anonymous reviewers who also contributed to improving the manuscript.

## FUNDING INFORMATION

This work was funded by two internal grants from the NSF-funded EarthCube national office (grants provided for the project titled ‘Sustainability models for integrated digital Earth Science beyond EarthCube’; NSF 1928208). Funding for open-access publication fees was provided by OPeNDAP. Tikoff was supported by NSF grant 1928273.

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTIONS

AV: Contributions: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review and Editing

JG: Contributions: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review and Editing

BT: Contributions: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Writing – Review and Editing



**Arika Virapongse**  [orcid.org/0000-0002-0194-5618](https://orcid.org/0000-0002-0194-5618)

Middle Path EcoSolutions, US

**James Gallagher**  [orcid.org/0000-0002-0311-202X](https://orcid.org/0000-0002-0311-202X)

OPeNDAP, US

**Basil Tikoff**  [orcid.org/0000-0001-6022-7002](https://orcid.org/0000-0001-6022-7002)

University of Wisconsin – Madison, US

## REFERENCES

- Ainsworth, J.** 2020. Feelings of ownership and volunteering: Examining psychological ownership as a volunteering motivation for nonprofit service organisations. *Journal of Retailing and Consumer Services*, 52: 101931. DOI: <https://doi.org/10.1016/j.jretconser.2019.101931>
- Altamimi, H and Liu, Q.** 2022. The nonprofit starvation cycle: Does overhead spending really impact program outcomes? *Nonprofit and Voluntary Sector Quarterly*, 51(6): 1324–1348. DOI: <https://doi.org/10.1177/08997640211057404>
- Arp, LG and Forbes, M.** 2018. *It takes a village: Open-source software sustainability: A Guidebook for programs serving cultural and scientific heritage*. Lyrasis, updated in 2022. <https://research.lyrasis.org/server/api/core/bitstreams/29629696-9abf-4144-8a79-62c6388ddb38/content>. Accessed on March 20, 2023.
- Bitzer, J and Schröder, PJ.** 2006. *The economics of open-source software development*. Elsevier Science, Inc. DOI: <https://doi.org/10.1016/B978-044452769-1/50001-9>
- Blue, G and Davidson, D.** 2020. Advancing a transformative social contract for the environmental sciences: From public engagement to justice. *Environmental Research Letters*, 15(11): 115008. DOI: <https://doi.org/10.1088/1748-9326/abbf14>
- Ceptureanu, SI, Ceptureanu, EG, Luchian, CE and Luchian, I.** 2018. Community-based programs sustainability. A multidimensional analysis of sustainability factors. *Sustainability*, 10(3): 870. DOI: <https://doi.org/10.3390/su10030870>
- Charan, R, Carey, D and Useem, M.** 2013. *Boards that lead: When to take charge, when to partner, and when to stay out of the way*. Harvard Business Review Press.
- Christensen, CM.** 1997. *The innovator's dilemma: When new technologies cause great firms to fail*. Harvard Business Review Press.
- Durst, S, Hinteregger, S and Zieba, M.** 2019. The linkage between knowledge risk management and organizational performance. *Journal of Business Research*, 105: 1–10. DOI: <https://doi.org/10.1016/j.jbusres.2019.08.002>
- EarthCube Office; EarthCube Leadership Council.** 2022. EarthCube: A retrospective 2011–2022. In: *EarthCube Organization Materials*. UC San Diego Library Digital Collections. DOI: <https://doi.org/10.6075/J04X5806>
- Ecological Society of America.** 2018. Workshop Report. Sustaining Data Repositories: A Workshop on Creating and Implementing Sustainability Plans. *Ecological Society of America*. January 25, 2018. [https://esa.org/sbi/wp-content/uploads/2019/09/Workshop-Report\\_Sustaining-Data-Repositories\\_FINAL.pdf](https://esa.org/sbi/wp-content/uploads/2019/09/Workshop-Report_Sustaining-Data-Repositories_FINAL.pdf).
- Erway, R.** 2023. Lasting Impact: Sustainability of Disciplinary Repositories. OCLC Research. Dublin, Ohio. <https://dfdf.dk/wp-content/uploads/2017/02/2012-03.pdf>.
- Ferraz, IN and Santos Júnior, CDD.** 2021. Organization of free and open source software projects: In-between the community and traditional governance. *BBR. Brazilian Business Review*, 18: 334–352. DOI: <https://doi.org/10.15728/bbr.2021.18.3.6>
- Fleming, PJ, Stone, LC, Creary, MS, Greene-Moton, E, Israel, BA, Key, KD, Reyes, AG, Wallerstein, N and Schulz, AJ.** 2023. Antiracism and community-based participatory research: Synergies, challenges, and opportunities. *American Journal of Public Health*, 113(1): 70–78. DOI: <https://doi.org/10.2105/AJPH.2022.307114>
- Garicano, L and Rayo, L.** 2016. Why organizations fail: Models and cases. *Journal of Economic Literature*, 54(1): 137–192. DOI: <https://doi.org/10.1257/jel.54.1.137>
- Goben, A and Sandusky, RJ.** 2020. Open data repositories: Current risks and opportunities. *College & Research Libraries News*, 81(2): 62–65. DOI: <https://doi.org/10.5860/crln.81.1.62>
- Hsu, L, Hutchison, VB and Langseth, ML.** 2019. Measuring sustainability of seed-funded earth science informatics projects. *PLoS ONE*, 14(10): e0222807. DOI: <https://doi.org/10.1371/journal.pone.0222807>
- Kaufman, YJ, et al.** 1997. The MODIS aerosol algorithm, products, and validation. *Journal of the Atmospheric Sciences*, 55(4): 569–579. DOI: [https://doi.org/10.1175/1520-0469\(1998\)055](https://doi.org/10.1175/1520-0469(1998)055)
- Kitchin, R, Collins, S and Frost, D.** 2015. Funding models for open access digital data repositories. *Online Information Review*, 39(5): 664–81. DOI: <https://doi.org/10.1108/OIR-01-2015-0031>



- Lin, D, et al. 2020. The TRUST Principles for Digital Repositories. *Scientific Data*, 7(1): 144. DOI: <https://doi.org/10.1038/s41597-020-0486-7>
- Maron, N. 2014. A guide to the best revenue models and funding sources for your digital resources. Ithaka S+R. [http://www.sr.ithaka.org/wp-content/uploads/2015/08/Jisc\\_Report\\_032614.pdf](http://www.sr.ithaka.org/wp-content/uploads/2015/08/Jisc_Report_032614.pdf). Accessed on March 20, 2024.
- Maron, N. 2018. Case Study: CSD and the CCDC. Ecological Society of America. [https://esa.org/sbi/wp-content/uploads/2018/05/Case-Study-1\\_CSD-and-the-CCDC.pdf](https://esa.org/sbi/wp-content/uploads/2018/05/Case-Study-1_CSD-and-the-CCDC.pdf). Accessed on March 20, 2024.
- McMillan, CJ and Overall, JS. 2017. Crossing the chasm and over the abyss: Perspectives on organizational failure. *Academy of Management Perspectives*, 31(4): 271–287. DOI: <https://doi.org/10.5465/amp.2017.0018>
- Minkler, M, Blackwell, AG, Thompson, M and Tamir, H. 2003. Community-based participatory research: implications for public health funding. *American Journal of Public Health*, 93(8): 1210–1213. DOI: <https://doi.org/10.2105/AJPH.93.8.1210>
- Molk, P and Sokol, DD. 2021. The challenges of nonprofit governance. *BCL Rev.*, 62: 1497.
- NASA. 2020. Science Investigator-led Processing Systems (SIPS). Earthdata: Open Access for Open Science. <https://www.earthdata.nasa.gov/eosdis/sips>.
- National Science Foundation (NSF). 2023. About Earth Sciences. National Science Foundation. Retrieved March 9, 2023, from <https://www.nsf.gov/geo/ear/about.jsp>.
- Paulson, JW, Succì, G and Eberlein, A. 2004. An empirical study of open-source and closed-source software products. *IEEE Transactions on Software Engineering*, 30(4), 246–256. DOI: <https://doi.org/10.1109/TSE.2004.1274044>
- Raymond, E. 1999. The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3): 23–49. DOI: <https://doi.org/10.1007/s12130-999-1026-0>
- Robinson, OC. 2014. Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology*, 11(1): 25–41. DOI: <https://doi.org/10.1080/14780887.2013.801543>
- Saadat, V and Saadat, Z. 2016. Organizational learning as a key role of organizational success. *Procedia-Social and Behavioral Sciences*, 230: 219–225. DOI: <https://doi.org/10.1016/j.sbspro.2016.09.028>
- Simpson, D. 2013. Looking forward through the rearview mirror: Reflections on 22 years at IRIS – and what the future might hold. *IRIS Membership Meeting*, December 9, 2013.
- Skinner, K and Christina, D. 2022. The Quest for Home: Transforming from Grant-funded Project to Sustainable Operation. [https://scholarlykitchen.sspnet.org/2022/05/19/the-quest-for-home-transforming-from-grant-funded-project-to-sustainable-operation/?informz=1&nbd=de50bbdb-968d-471a-ac98-0cfecba6bca0&nbd\\_source=informz](https://scholarlykitchen.sspnet.org/2022/05/19/the-quest-for-home-transforming-from-grant-funded-project-to-sustainable-operation/?informz=1&nbd=de50bbdb-968d-471a-ac98-0cfecba6bca0&nbd_source=informz).
- Stirman, SW, Kimberly, J, Cook, N, Calloway, A, Castro, F and Charns, M. 2012. The sustainability of new programs and innovations: A review of the empirical literature and recommendations for future research. *Implementation Science*, 7: 17. PMID:22417162. DOI: <https://doi.org/10.1186/1748-5908-7-17>
- Strauss, AL and Corbin, J. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. London: Sage.
- Virapongse, A, Gallagher, J, Tikoff, B, Cornillon, P, Koskela, R, Shingledecker, S, Trabant, C and Hanson, B. 2022. Sustainability models for integrated digital Earth Science. In: *EarthCube Organization Materials*. UC San Diego Library Digital Collections. DOI: <https://doi.org/10.6075/J0JH3MBN>.
- Volentine, R, Specht, A, Allard, S, Frame, M, Hu, R and Zolly, L. 2021. Accessibility of environmental data for sharing: The role of UX in large cyberinfrastructure projects. *Ecological Informatics*, 63: 101317. DOI: <https://doi.org/10.1016/j.ecoinf.2021.101317>

#### TO CITE THIS ARTICLE:

Virapongse, A, Gallagher, J and Tikoff, B. 2024. Insights on Sustainability of Earth Science Data Infrastructure Projects. *Data Science Journal*, 23: 14, pp. 1–27. DOI: <https://doi.org/10.5334/dsj-2024-014>

**Submitted:** 04 October 2023

**Accepted:** 28 December 2023

**Published:** 28 March 2024

#### COPYRIGHT:

© 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.