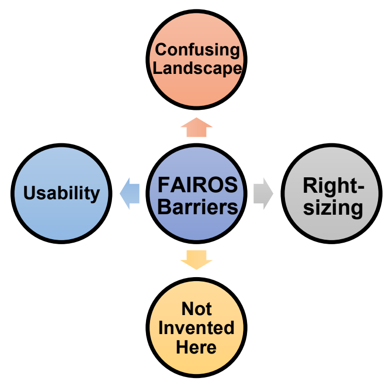
Project Description

Introduction

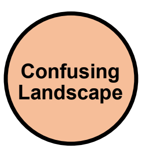
We propose to develop a Community FAIR research coordination network (RCN) that establishes discipline-based, community protocols for Open Science (OS) and FAIR-er data and software. Working directly with scientific disciplines, starting with the Earth, space, and environmental science, we will use a design patterns-approach to create reusable templates analogous to laboratory or experiment protocols that will guide researchers in best practices for implementing FAIR data, software, and sample metadata in their work (we group these collectively as “FAIR data” or data+). The lack of actionable, community accepted, FAIR data practices within specific disciplines has been the fundamental challenge in realizing a vision of interoperability and reusability. Our envisaged process has four parts: 1) creating the design pattern protocols; 2) testing them in several pilot disciplines; 3) extending this template widely; and 4) facilitating adoption and use of them. The design patterns are both proven and novel for this application that will address, as appropriate, elements of FAIR and OS, providing guidance in areas such as repositories and metadata for disciplines to enable interoperability, and eventually will be made machine readable. The design patterns will be developed with and by the discipline-specific communities. Similarly, we will work with the same communities to identify effective ways to facilitate adoption and use of the protocols. Central to the success of the project will be leveraging researcher relationships with society-organized domain communities together with repositories, research services and infrastructure providers supporting OS, and sharing data+. Letters of support for this effort are included by 50 of these groups, spanning most of the natural and biological sciences (representing a large RCN). These relationships are critical to the creation, evaluation, and ongoing maintenance of a design pattern and protocol library that ultimately will be a resource for many disciplines. This effort will thus go a long way to addressing the key barrier for widespread adoption of FAIR data practices.

Problem Statement

Barriers to FAIR-er Scientific Objects



The list of barriers that inhibit FAIR-er data and more OS includes technology, culture, law, policies, societal norms, and more (Borgman, 2015). We will focus on addressing four connected core problems that have slowed the development and uptake of FAIR data and software practices and consequently inhibits OS and data reuse: 1) overwhelming landscape of guidance, tools, and policies, of which a subset may be relevant for a researcher to solve a FAIROS data+ problem; 2) right-sizing the problem with a solution that will continue to evolve with technology and science, 3) the “not-invented-here problem”, and 4) usability.



Scientists seeking to enhance the “FAIR-ness” of their data and/or software face a confusing array of FAIR-related tools, standards, guides, and more. There are dozens of data formats ranging from proprietary to open. There are at least as many metadata standards and domain keyword vocabularies, not to mention ontologies and other data-related schema. There are a variety of repository options from which to choose for publishing data including institutional repositories, libraries, domain repositories, and non-commercial general-purpose repositories; some of which only accept data from certain users. Different institutions from funders to publishers to universities have policy directives that overlap (Borgman, 2015, p. 38). How is an individual scientist or project team supposed to figure out which among these choices will help them to make digital research objects more FAIR and how much effort will it take? Though guides, recommendations, how-to’s, certifications, and even the framework suggested by FAIR itself are available for scientists, what is missing is succinct guidance developed and vetted by the community for which the researcher belongs. This type of focused guidance could also help researchers implement FAIR-er data throughout the research lifecycle, from proposal to publication.

Right-sizing

There are additional challenges related to determining how to increase FAIR-ness versus finding the “right-size”. There is the implication that making digital research objects FAIR is a binary choice, either the data+ is FAIR or not, and the just-in-time nature of research—including trying to comply with data+ requirements at the time of publication—make it hard to allocate scarce resources to a task. Given that choice, researchers retain resources to support their science out of a concern that the requirement is too hard and the effort does not give an equal or greater benefit. It is not clear how FAIR data+ needs to be, or what the “right-size” should be to get the most benefit.

Not Invented Here

Scientific domains face another challenge in the context of enabling FAIR digital research objects known as the “not-invented-here problem”. This problem highlights a scientist’s perception that solutions developed for another domain will not work for their research objects. Externalizing the expert knowledge relevant to their data in ways that would facilitate things like discoverability and interoperability, becomes over-complicated with over-engineered solutions not providing the expected levels of interoperability, or uptake. Again, this contributes to the perception that the threshold for improving FAIR-ness is too high to invest the requisite time and resources. Having guidance that researchers recognize as relevant to their work, and developed by leading authorities in their discipline engenders trust in the process.

Usability

Another factor that limits researchers adopting and deploying tools and practices to enhance FAIR-ness is usability. Usability in this context includes providing the right level of information related to improving FAIR-ness at the right time in the research process in the right form. Ultimately, this means reducing information and cognitive overload, presenting the relevant information in such a way to facilitate uptake and use for a FAIR-related task, and improving the researcher experience in implementing FAIR practices. In other words, keeping it simple is key. Usability has many dimensions including communications and adoption. How do you capture community consensus in a way that facilitates re-use and how do you inculcate adoption? For example, focusing on developing checklists has been shown to produce better procedural outcomes as opposed to detailed manuals and how-to’s. The idea is that operators/users need to know or be reminded what to do, not how to do it (Gawande, 2010).

FAIR-ness, Open Science, and Cross-Cutting Challenges

FAIR-ness and Open Science are linked. The lack of broadly operationalized data, software, and sample management protocols within domains inhibits cross-domain transparency. FAIR-er digital research objects are building blocks for promoting OS as well as trust in science. Without meaningful sharing throughout the scientific process, the process will remain opaque, and the potential benefits will remain unrealized. The challenges related to FAIR-er digital research objects also apply to developing discipline consensus on what OS means and how to operationalize ways to achieve a more open approach to science. For example, in addition to multiple descriptions of OS, basic concepts often associated with OS, such as replicability and reproducibility, are incredibly difficult to achieve, vary from one domain to another, and in the case of some domains efforts to replicate well known scientific papers have failed (Bavel, 2016; Van Bavel et al., 2016).

Two aspects of the “not-invented-here problem” are particularly relevant as impediments to cross-cutting uses of FAIROS tools: the “my-data-are-special problem” and the ongoing stove-piped nature of the science enterprise. The first is the inherent tension between the desire to capture detailed domain-specific expert knowledge, which makes curation difficult, versus a goal of promoting adoption, in which there is agreement of a “good enough” outcome, not the best outcome (Schwartz et al., 2002). In spite of the importance of interdisciplinary, transdisciplinary, convergent, and team science, there is no coordinated effort within or across the disciplines to create or manage the data+ protocols (re: best practices) that enable the objectives of FAIR and OS. The lack of these types of protocols allows the data and software that underpins research to be kept private, inaccessible, and poorly documented.

In spite of the challenges above, scientists consistently report that sharing data and re-using data are valuable and of high priority. Researchers in many surveys have relayed that it is difficult to find data they need and that they often do not trust the quality or provenance when they are able to find the data. Even when data are deposited in repositories, the absence of quality metadata and even minimal documentation, especially in general repositories, can reduce their trustworthiness (Cutcher-Gershenfeld et al., 2016; Tenopir et al., 2018; Nusser et al., 2021).

Proposed Solution and How it Addresses the Challenges

The project will be led by an experienced team at the American Geophysical Union (AGU) and partners at the Renaissance Computing Institute (RENCI) and Earth Science Information Partners (ESIP), supported by a large number of relevant stakeholders across the sciences and data infrastructure, forming the RCN. These stakeholders have indicated their support for and intended engagement with this work (Table 4).

The developers of the FAIR data principles described several salient features of FAIR, two of which inform our vision for this RCN. First, there cannot just be a commitment to FAIR data, but that the follow-through must be part of that commitment. The second is that even some standardized progress may go a long way towards achieving FAIR goals broadly (Mons et al., 2017; Wilkinson et al., 2016). We argue that design patterns applied and developed for each major discipline provide the means for researchers to follow-through on their commitment to FAIR data+ and do so in a way that provides efficiencies to maximize the return. Additionally, we will also ensure that the design patterns are kept up to date through assessment, continuous improvement, and engagement, so that they remain applicable and used.

The work is structured with the following themes:

1. **Design Pattern Creation:** Develop, implement, and refine the design pattern creation process. The creation process includes developing the initial version, as well as built-in steps for refining and testing. Each design pattern represents the leading data+ practices of the discipline-specific research life cycle. This includes practices such as repository guidance for data types and sources, information on relevant metadata, standards for data and data product archiving (for example for streaming data), and more.
2. **Communications, Adoption, Dissemination, Engagement, and Education (CADEE):** This set of activities includes efforts to promote adoption, dissemination, and uptake of design patterns, as well as providing engagement strategies to recruit additional domains through our scientific society partners and provide for train-the-trainer (TTT) efforts.
3. **Diversity, Equity, Inclusion, and Accessibility (DEIA):** FAIROS requires data+ be accessible and as open as possible supporting equity and data democratization. In this theme we ensure our RCN is a safe space following AGU's Scientific Integrity and Professional Ethics guide as well as AGU's Meetings Code of Conduct for our workshops and working sessions. We include a DEIA design pattern iteration to inform our community about data+ elements to consider in their research work. For all design patterns we include guidance on designing equitable participation into the protocol and use of open-source tools where possible.
4. **Maintain and Sustain**: The ultimate outcome is for our RCN to use community-vetted practices and protocols for data+ that minimize the complexity of the wide variety of guidance currently available. We will incorporate the recommendations developed from the recent research in data sharing and reuse by connecting these practices to the user experience. Further, through partnership with ESIP, design patterns will be hosted on the Community FAIR website, extending ESIP's Data Help Desk to include a knowledge base and provide tutorials and materials for stakeholders to deploy the methodology independently.
5. **Assessment and Testing:** Assessment includes routine evaluation of the project progress, assessment of the efficacy of the design patterns, as well as direct user testing of initial design pattern versions.

What is a Design Pattern and Why is it Relevant?

Using "design patterns" is core to our approach. The concept originated in software engineering and development in the 1980s (Beck et al., 1996), and has been widely deployed in a variety of contexts. A design pattern is a reusable template created to address a common problem—in this case guidance for practicing FAIR data where the overall types of questions and approaches are common across disciplines but the specific information, like which repositories and metadata to use, is different and unique. They incorporate elements of processes (like flow charts) and decision trees. We will use design patterns as a means to guide researchers efficiently through their discipline-specific decisions in enabling FAIROS. Having an overall common backbone (the design patterns) enables scalability and transdisciplinary usage, and thus interoperability.

Design patterns are not best practices, standards, or documentation; instead, they incorporate, encode, or refer to these things in a common and replicable way. The advantages of design patterns are well established and include: facilitating communicating complex problems, promoting re-use, capturing community practices, and improving maintainability (Aversano et al., 2007; Beck et al., 1996). Design patterns have already made their way into some geosciences research: They are mentioned in 20 AGU abstracts since 2004, and the terms “design patterns and geosciences” are referenced in about 1400 Google Scholar citations. Examples include applications of design patterns to ontology development (Krisnadhi et al., 2015), geoscience education research (Kastens & Krumhansl, 2017), and earth process modeling (Tubini & Rigon, 2022). We believe that this application of design patterns as described is novel and provides an innovative way to meet scientists where they are as researchers.

A key challenge is to be sure that the design pattern is sufficiently detailed to guide researchers in FAIR data practices for the types of outputs they are generating, but not so detailed as to be cumbersome. The design pattern approach is useful because it can help researchers efficiently navigate a complex space with many types of data. Some types of information that are likely to be included are: the steps and key inputs needed for adding a permanent identifier to your dataset, what the NSF metadata and repository expectations are (as outlined in places such as the NSF Access Initiative (2015)), how to estimate and budget data management costs for your project, and more.

Our goal is to leverage the unique position occupied by the concept of a design pattern. A design pattern is not a blueprint for a house. A design pattern for building a house would, however, outline key steps, systems, and dependencies that need to be understood to create a blueprint that can be used to build a house. The design pattern can be used to create any number of different types of houses or could be used to produce a digital version of a blueprint instead of a hardcopy, i.e., a CAD design, that can be used to leverage automation to build a specific house. In this example, a design pattern enables an architect or engineer to create the blueprint for the general contractor to build the house. In this example, elements such as standards are relevant, but a design pattern is not a standard in the sense of a metadata standard. The design pattern could incorporate the need to follow building codes and other requirements, but it would not necessarily reference the exact version and section of code to follow. The design pattern would be style-independent, i.e., you could use it to create a design for an ultra-modern house, or a traditional two-story colonial. The design pattern functions as an artifact supporting the development of a boundary object that has meaning to each domain involved in building a house.

In a similar way, the design patterns we seek to develop as part of this RCN will address common problems associated with FAIROS, providing scientists with practical guidance analogous to research methodologies, experimental protocols, and laboratory procedures. To promote FAIROS, we do not need to make researchers into metadata specialists, ontology engineers, or data librarians. However, researchers can benefit from community-based design patterns that address difficulties common to all researchers and integrate the requisite knowledge from multiple domains (e.g., topical domain, information technology, library and information science, and computer science). For example, creating a design pattern such as outlining the steps necessary to develop a data management plan for hydrologic research will need to incorporate not only referencing domain specific standards, but also basic data management protocols, knowledge of what policies may apply, and an understanding of the availability of relevant cyberinfrastructure tools, and how and where to get further help.

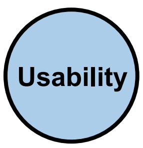
Confusing Landscape

A design pattern distills and synthesizes across information sources to address essential elements of a problem thus helping the scientific community to overcome the confusing landscape challenge. A design pattern that results from the process outlined in this RCN will represent a consensus artifact that enables scientists who seek to advance FAIROS to focus on only those elements most directly related to the activity they wish to address, e.g., picking a repository for storing data. **Design patterns reduce the cognitive load on a researcher related to operationalizing FAIROS.** The researcher doesn’t have to figure out what to do, only decide the way in which they would like to implement the design pattern.

Right-sizing

The intent of design patterns to address common problems encountered by groups pursuing similar goals scopes the solution to a particular problem and allows for a level of modularity among patterns associated with higher level activities, i.e., creating a software application. The research lifecycle provides an overall framework for identifying a set of design patterns to address elements of the lifecycle and thereby right-sizing the needed FAIROS activities. Because the overall life cycle is common across disciplines, a design pattern approach is ideal. A design pattern for handling early versions of data developed in the initial phases of a research project might differ from a design pattern that outlined a checklist for repository submission as part of a journal publication created to summarize findings but both would form a common cross-disciplinary framework. Following a lifecycle approach also allows researchers to start from a point, orient themselves to FAIROS goals, and identify particular pain points in their workflows for which they can leverage the design patterns. Points along the lifecycle where an appropriate FAIROS design pattern might be relevant includes experiment design, device calibration, data capture, cleaning, and analysis, publication and preservation (Pepe, 2011). We can also use the concept of a researcher 'data story', to elicit either repeated patterns that are already in place, but not documented, or gaps that need to be addressed (Pepe et al., 2014).

In their most basic form, design patterns will be reusable across domains, and inherently cross cutting since they will avoid encoding specific solutions. In the language of software engineering, a design pattern can be applied to any code development language, e.g., not limited to Java or Python. However, we do expect that the reality of the FAIROS design patterns to be developed is that they will likely include domain-specific supporting material, e.g., domain-specific repository options. However, those specific elements could easily be swapped out depending on the domain without affecting the overall cross-cutting utility of the design pattern. **This approach transcends disciplinary parochialisms associated with the 'not invented here' problem.**



Design patterns in and of themselves should have a high degree of utility. However, in order to make sure that the design patterns have utility and to promote uptake, we will include an intentional step to transform the design pattern from its initial raw, technical description into a refined version designed to facilitate comprehension and use. The usability of the refined version will be assessed with findings used to improve the final version. Further, as explained below, the community-generated design patterns will not be static. As technology and science evolve, design patterns may be updated to maintain their utility.

Extending the Approach

As a complement to our proposed activities designed to support maintaining and sustaining the Community FAIR activity, we will implement a three-part strategy for expanding the reach of this approach within and beyond the Earth, space, and environmental sciences. First, this RCN will implement multiple iterations of design pattern development for selected domains. We plan our first RCN iteration with the cryosphere community, and then engage additional domains in subsequent iterations. Second, we will encapsulate the approach to provide a franchise methodology whereby additional groups with the Earth, space, and environmental sciences can gain the skills to run their own design pattern iterations, i.e., a Community FAIR Franchise. In this scenario the Community FAIR RCN acts as coach or mentor guiding implementation by a new franchise. Third, to promote uptake across other disciplines, the Community FAIR RCN will leverage AGU's participation in a collaborative community of scientific societies and federations (WeShareData.org) that are actively working to support their researchers in developing skills in FAIR Principles and OS. For this proposal we will continue to develop that collaboration and share our work, provide train-the-trainer workshops, and actively support their efforts. These societies and federations cut across broad disciplines and we are open to adding addition collaborators:

- American Meteorological Society (AMS) Board on Data Stewardship,
- Federation of American Societies for Experimental Biology (FASEB), representing 30 societies
- American Astronomical Society (AAS),
- Federation of Associations in Behavioral & Brain Sciences (FABBS), representing 29 societies

How the RCN will work

Our proposed RCN will include a large number of linked stakeholders committed to improving FAIR and OS (See Table 1) representing the breadth of knowledge needed to develop usable design patterns, across most major science disciplines. Stakeholders include researchers from disciplinary sections within a professional society; scientific societies working to develop data+ protocols for their research communities; repositories, research services and infrastructure providers supporting data sharing and OS across communities. The research coordination network and its activities will be structured around the themes described earlier which address the challenge areas described above. These themes will also help frame outcomes and provide a management framework.

Project Organization, Management, and Timeline

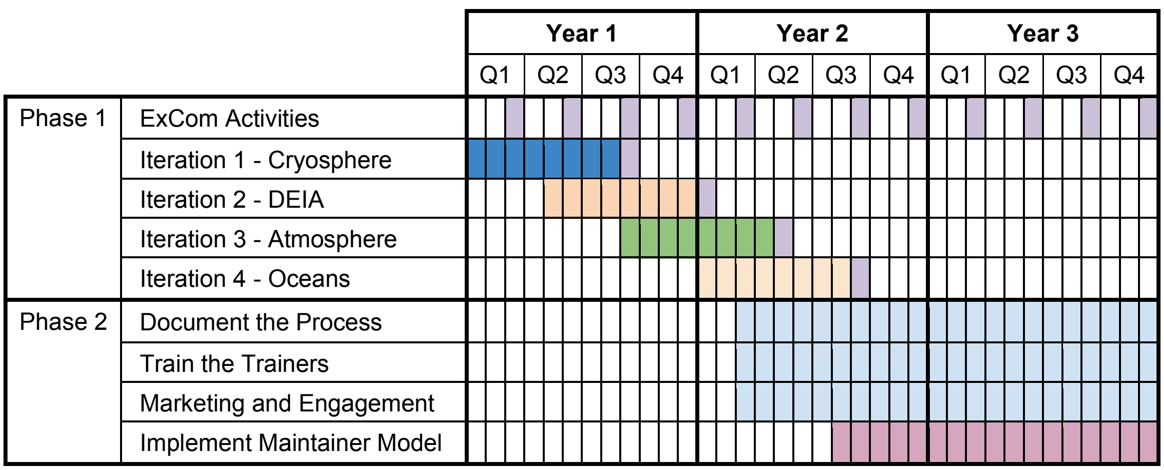
Project Overview

Work Breakdown Structure

The activities of this RCN will be aligned with the thematic elements described above, Design Pattern Creation, Assessment and Testing, Communications, Adoption, Dissemination, Engagement, and Education (CADEE), Diversity, Equity, Inclusion, and Accessibility (DEIA), and Maintain and Sustain.

During the course of the RCN, our goal is to oversee the implementation of four Iterations of the Design Pattern Process (See Table 1) and to explicitly address operationalizing and institutionalizing sustainability for the project objectives. Each Design Pattern Process iteration will run over the course of approximately eight months, at the end of which we will have discipline-specific protocols based on the design patterns. These will be developed, assessed, and deployed. Subsequent iterations will begin roughly one-third to one-half of the way through the previous iteration. Overlapping iterations will enable faster progress, faster learning across iterations, and cross-cutting synergies. At the start of the second year we will shift our focus to prepare others on how to create data+ design pattern practices, building on the lessons learned during the prior iterations to provide consultative guidance to domains and professional organizations and scientific societies, providing materials, train-the-trainer events, and coaching.

Table 1: Proposed Project Timeline Roll-up



Management and Decision-making

The RCN will be managed by an Executive Committee, composed of project PIs. The Executive Committee coordinates project activities and has primary responsibility for monitoring project progress. The Executive Committee will meet monthly to coordinate RCN activities and planning. In addition, every three months the Executive Committee meeting will be devoted to formal ‘check and adjust’ activities, linked to assessment efforts to ensure continual improvement. We anticipate that check and adjust meetings will require a longer session either co-located at a relevant meeting or conducted virtually.

The Executive Committee includes Shelley Stall (AGU), Chris Lenhardt (RENCI), Brooks Hanson (AGU), Chris Erdmann (AGU), Susan Shingledecker (ESIP), Danie Kinkade (WHOI), Margaret O’Brien (UCSB), Lauren Parr (AGU), Karl Benedict (UNM), Jane Wyngaard (USAfrica). The committee will be assisted by a full-time Project Coordinator, Laura Lyon. She will monitor deadlines, reporting requirements, help to manage the project schedule and timeline, and assist with preparing and managing project workshops.

Management and Decision-making Deliverables/Outcomes:

1. Project schedule and updates
2. Agenda and minutes for each executive committee meeting.
3. Documented lessons learned, and changes made to the process.

Theme Activities

Design Pattern Creation (Leads: Lenhardt/Stall)

Creating design patterns will be conducted via a six-to-eight-month iterations divided into two phases. See Table 2 for overview of the iteration schedule. The goal of Phase I is to develop the initial version of a design pattern. In Phase II the goal is to test and refine the initial version of a design pattern developed in Phase I and ready the polished design pattern for dissemination, adoption, and maintenance. Each phase will lead off with a workshop which will be organized to achieve the goals specific to each phase. See Table # for an example of draft agendas for each workshop. Workshop 1 will socialize the design pattern concept, work with the stakeholders to identify specific pain points to address that could benefit from the creation of a design pattern and which if addressed, would increase the FAIR-ness of community digital objects and/or contribute to enabling more open science. During the workshop, smaller working groups will be formed that will have the charge to develop one of the proposed design patterns, creating a v0. The v0 design pattern will be further developed by the Project Coordinator, with support from AGU staff, to create a testable version for Workshop II. Workshop II will focus on assessment and feedback on the initial version. Feedback from Workshop II will be incorporated into a revised version, v1, of the design pattern and will be prepared for dissemination to the wider community. At the end of the iteration, the design pattern may be revised further. Once reviewed and accepted, the design pattern will be known as the protocol for the topic it addresses. It will be preserved and added to the virtual online library for easy discovery and access. The domain time, AGU, ESIP and partners will promote its use and adoption.

Table 2: Design Pattern Creation Iteration

Month	1	2	3	4	5	6	7	8
Iteration Science domain(s)	1st workshop (two virtual sessions)	Working groups (6-8 weeks)	Evolve v0 → v1	2nd workshop (two virtual sessions)	Incorporate feedback	Comms Prep	Section Reveal	Adoption and Maintenance

Table 3: Example Workshop Agendas

DRAFT Agenda - Workshop 1	**DRAFT Agenda - Workshop 2**
<u>Session 1</u> 1. Orient the participants on process -- What is a design pattern and RCN application 2. Data Story review [ways hydrologists use data and software] - “a new story might be a new design pattern” -- the lifecycle: creation, use, storage, reuse 3. Provide examples 4. Consensus development on current landscape (goals, barriers, metrics, etc.) 5. Homework: What are the gaps? <u>Session 2</u> 1. Summary of previous session before 2. Gap discussion 3. Identify potential design patterns to create 4. Establish working groups (5-7 people), one per proposed design pattern 5. Determine working group schedule, other next steps 6. Workshop assessment	<u>Session 1</u> 1. Summary of Workshop 1 outcomes and work between Workshop 1 and 2 2. Present draft design patterns 3. Focus group evaluation 4. Outbrief, discussion, synthesis 5. Proposed revisions <u>Session 2</u> 1. Summary of previous session 2. Session 1 reactions (things to continue, drop, modify) 3. Release planning (e.g. schedule, comms, evaluation, training) 4. Discuss initial maintenance plan 5. Determine other next steps 6. Workshop assessment

Design Pattern Creation Deliverables/Outcomes:

1. Discipline Specific Design Patterns, accessible in the Community FAIR library, preserved in Zenodo.
2. Workshop materials for each iteration.

Communications, Adoption, Dissemination, Engagement, and Education (CADEE) (Leads: Shingledecker/Lyon)

Communications: This set of activities includes efforts to promote adoption, dissemination, and uptake of completed design patterns in coordination with the engaged geoscience section at AGU and international sister societies the European Geosciences Union (EGU) and the Japan Geoscience Union (JpGU). Also included is providing engagement strategies to recruit additional geoscience disciplines to work with Community FAIR by running their own workshop following our methodology or working together. This extends to other scientific societies along with providing train-the-trainer (TTT) efforts. Communications also includes working with a geosciences section, identifying their data+ leaders and those interested in developing design pattern protocols. <u>Tools:</u> Community FAIR webpage, AGU/EGU/JpGU communication channels.

Adoption: Adoption is closely aligned to communications. The Community FAIR team, in collaboration with the researchers who helped to develop their discipline-specific design pattern protocols, will promote the use and value of their design patterns through section communications, events and personal use cases

Dissemination: As new or updated design patterns are completed, we want to invite all researchers in the discipline to learn about them, along with the international informatics community, librarians, and other stakeholders. <u>Tools:</u> Community FAIR webpage, AGU/EGU/JpGU communication channels, the Data Help Desk and knowledgebase, International Partner channels, the Data Curation Network.

Engagement: Tightly connected to communications, engagement provides the value statements and motivations for researchers and their communities to participate in developing and using design pattern protocols. On the Community FAIR webpage we will provide means to receive communications and participate in existing and upcoming design pattern development iterations and events. We work closely with discipline leaders in data+ to provide clear value statements and examples to encourage participation. We also recruit additional domains through our scientific society partners and provide for TTT efforts.

Education: During section events, discipline conferences (e.g., Frontiers in Hydrology), we offer workshops on design patterns in collaboration with the relevant section. For those interested in developing their own design patterns we create materials and provide TTT events.

CADEE Deliverables/Outcomes:

1. Communications plan with AGU, EGU, and JpGU communications channels
2. Schedule of engagements and education.
3. Community FAIR website updates with future events promoted, and recordings and materials from previous events.

Diversity, Equity, Inclusion, and Accessibility (DEIA) (Leads: Stall/Williams)

Equity and data democratization are important elements in OS. In this project we seek to tangibly address DEIA in the following ways:

1. **Process:** We will use AGU's Scientific Integrity and Professional Ethics (2017) which includes AGU's Meetings Code of Conduct for our workshops and working sessions to ensure safe spaces.
2. **Participation.** We will use AGU's internal diversity criteria guide for forming volunteer groups to better reflect the breadth of perspectives in our community: gender, geography, career stage, employer type, low- and middle-income countries, and science.
3. **Outcomes:** We include in a design pattern iteration specific to DEIA to inform our community about data+ elements to consider in their research work. We are deliberately putting this focus early in the RCN project so that results will inform follow-on work as the RCN moves forward.

Following **data democratization recommendations** made by Sabina Leonelli and Louise Bezuidenhout (CITE), we will ensure “equitable participation in the creation of knowledge, through data stewardship that is transparent, subject to scrutiny and grounded on commitment to justice and fairness” (Leonelli et al, 2021). In specific, Leoneli and Bezuidenhout state:

a) Make data scrutinizable and re-usable, while remaining mindful of their social and political value. Data are not neutral facts.
b) Guide participants to design protocols with both high-tech and low-tech options.
c) Encourage design patterns to include use of Open Science tools such as GitHub and the Center for Open Science.

1. **Extended Engagement:** We also draw on the DEI expertise and experience from the greater project team and partners involved in OS efforts ranging from The Carpentries/Library Carpentry to scholarly communications efforts such as the Coalition for Diversity and Inclusion in Scholarly Communications (C4DISC). Libraries are central to our efforts, the expertise and experience from the project team, partners to create welcoming spaces, and adding expertise on Accessibility and the Usability work identified in this project. AGU/ESIP work with community members on adopting and integrating the CARE principles for Indigenous Data Sovereignty, for instance, in data+ sharing guidance and repository specifications, which will also be integrated into our project work.

DEIA Deliverables/Outcomes:

1. DEIA Design Pattern to inform the community about data+ elements to consider.
2. Recommendations for design pattern development that include guidance on data democratization practices in OS.

Maintain and Sustain (Leads: Parr/Shingledecker)

Discoverable design pattern protocols vetted, maintained and governed through a collaboration of the geoscience sections (AGU, EGU, and JpGU) and our partner societies facilitated by the Community FAIR RCN are key to adoption and continued trust in the Community FAIR materials. This work will be ever-evolving following advances in science and technology. The protocols identified through the design patterns work will require cultivation, curation, review and dissemination to remain relevant and sustainable over time. The extensibility strategy for the RCN described earlier leverages specific mechanisms that we will use to support maintaining the RCN and for creating the momentum to sustain the RCN. These mechanisms include: Knowledge Management, a Franchise / Train-the-Trainer Capability, Collaborative Infrastructure, and implementing a Maintainer Model. The combination of the geoscience sections across all unions and ESIP’s collaborative infrastructure bring ready-made, well integrated, and on-going mechanisms for revision, refinement and engagement beyond the period of this project in the Earth and space sciences. Additional mechanisms, like the Data Help Desk, are a vehicle for broad dissemination of this knowledge base to the wider research community.

1. **Knowledge management.** Providing an up-to-date source of information related to this effort and platform for dissemination of the results is essential for developing the type of adoption we are promoting. AGU is committed to supporting the community management of these practices via a Community FAIR website, with strong visibility and attribution to all partners. The website will host the design pattern protocols with the ability to give feedback and direct questions to designated domain champions. In addition, the ESIP Data Help Desk will expand to include links to the Community FAIR website as well as a new 24/7 knowledgebase, hosted on Zendesk, to help researchers any time day or night. The knowledge base is intended to provide domain specific data+ management questions and answers to researchers no matter their location or time of day.
2. **Franchise / Train-the-Trainer Capability.** We will provide Design Pattern Creation process, TTT workshops, and coaching support as part of a franchise package to broadly enable engagement across 25 AGU sections and partner societies in the brain and behavioral science, experimental biology, and meteorology. For communities in the Earth, space, and environmental sciences we will host their final design patterns on the Community FAIR webpage along with the franchise package material openly accessible, version controlled, and using the interactive feedback functions in GitHub.
3. **Collaboration Infrastructure.** Much of science is conducted via distributed teams leveraging collaboration technologies and routines. This project is no exception, and our plan is to utilize the collaboration infrastructure and meeting opportunities provided by ESIP. For example, having a

working group hosted at ESIP allows the effort to tap additional resources such as online collaboration tools and also provides a means to enculturate new groups and researchers to the ways the group works. We do not have to create those processes anew.

4. **Maintainer model.** Our design pattern protocols maintenance approach builds from The Carpentries Handbook material on Lesson Maintenance which includes guidelines, onboarding, communication, organization, troubleshooting, and templates (The Carpentries, 2022). We will streamline and adapt the material to our approach, which will help volunteer maintainers streamline workflows on how they discuss and approve community contributions. In addition, the maintainer work can be tied to section/disciplinary work where recognition can be extended (which is not currently available/possible via The Carpentries). AGU is committed to managing the design pattern protocols even after the conclusion of the RCN.

Maintain and Sustain Deliverables/Outcomes:

1. Community FAIR Website, hosted on AGU's GitHub repo.
2. ESIP and AGU's knowledge base using Zendesk, supporting the Data Help Desk with links from/to the Community FAIR Website
3. Franchise Package materials hosted on the Community FAIR website, preserved in Zenodo.
4. Train-the-trainer materials hosted on the Community FAIR website, preserved in Zenodo.
5. Maintenance plan for the design pattern protocols.

Assessment and Testing (Leads: Parr/Stall)

The approach to the Community FAIR RCN will include ongoing activities to monitor and improve the functioning of the RCN, as well as incorporating testing the design patterns and associated training products for efficacy.

1. Project and Operational Metrics
   a) The Executive Committee will conduct quarterly 'check and adjust' activities. The purpose of the check and adjust activity is to monitor how well the RCN is functioning, identify barriers and successes, and evaluate progress on RCN goals. Based on the findings, adjustments will be made to future activities in an effort to improve outcomes. Results will be positioned for sharing across networks and in training materials as appropriate.
   b) We will track the development of the design pattern protocols and practices within the Earth, space and environmental sciences and support the adoption by our partner societies.
   c) Other metrics include website analytics and workshop participant surveys.
2. Outcomes Assessments
   a) As part of the design pattern development, we will conduct evaluations with stakeholders, for example convening a focus group at a workshop to review draft design patterns.
   b) We will implement a pre- and post- evaluation of the intervention to directly assess the viability of the design pattern and to determine stakeholder's level of knowledge, understanding, and self-described ability to implement what they have learned (competence) in their practice as well as any barriers they see to adoption of the design patterns (performance).
   c) If a given design pattern is linked to a specific activity, like a repository submission, or linked to a publication, we will work with that organization to track, if possible, submissions that leverage a design pattern as part of the process and leverage this information for feedback to the RCN. Metrics related to the adoption of a design pattern will be established with the RCN and partner organizations (metric framework) and may include design pattern references, access to the design pattern library, citations, journal data, etc. We can also track adoption of design patterns by following stakeholders who explicitly reference the design pattern or design pattern library, for example including a reference to the design pattern in submission instructions.
   d) Train-the-trainer materials and activities are key to the adoption of the design patterns and support the outcomes of the RCN. These materials will be developed to teach the application of design patterns, protocols, and approaches, as well as to support discipline-specific communities to develop their own design patterns. Additionally, materials will follow a Carpentries structure to ensure consistent training across RCNs and disciplines. Use of open education and co-created resources will also provide consistent feedback on the

perceptions and barriers to adoption of design patterns. Consolidated evaluations and assessments for use by trainers will ensure the ability to demonstrate outcomes across disciplines and the RCN providing continuous improvement loops. Trainers can also be assessed and connected via this model to share best practices and case studies.

Assessment and Testing Deliverables/Outcomes:

1. Project and Operational Metric Reports
2. Outcome Assessment Reports

Why this team and set of partners?

Leadership Experience

The project leadership brings a broad network of partners and extensive expertise in areas related to scientific digital object management, design patterns, community engagement, cyberinfrastructure, OS, cross-domain boundary spanning, and thought leadership. In all, we have received commitments for participation for 50 organizations across many disciplines. We thus have confidence that this network can deliver on the templates and ensure wide adoption. The project will be managed by staff at AGU, ESIP, and RENCI; these organizations have deep experience in implementing community engagement around FAIR data implementations. The co-leads for this RCN include:

Shelley Stall, Senior Director for Data Leadership at AGU, is a certified Project Management Professional (1341027) and has an MBA. She has successfully delivered her projects for AGU within budget. Prior to AGU she has supported both Defense and Federal Civilian US federal government agencies as well as working as a system integrator and project manager for large commercial companies.

Chris Lenhardt, a Senior Research Scientist in the Earth Data Science Group at RENCI, University of North Carolina at Chapel Hill, brings extensive experience supporting operational earth science data repositories, working with domain scientists to create cyberinfrastructure, as well as expertise in community-building and expertise in sociotechnical approaches to understanding the development of cyberinfrastructure.

To prepare for this RCN opportunity, AGU, ESIP, and RENCI are conducting a pilot event using our design pattern approach with the hydrology community during Summer 2022. We are partnered with the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) and will invite the international hydrology community to participate with details provided in May. AGU, ESIP, RENCI, and CUAHSI have been working together on developing this pilot and the structure of Community FAIR for over a year and discussing the approach with the research informatics community and AGU Leadership Council starting in Fall of 2021. The keen interest and encouragement from AGU volunteer leadership as well as EGU and JpGU leadership are a strong indicator that the community understands the value of the work and is willing to participate. We will invite the larger hydrology community to participate during a town hall the June 2022 Frontiers in Hydrology Meeting.

Results from Prior Work

Shelley Stall: (a) NSF Award number: 1929464, $499,953, 07/15/2019 – 06/30/2023, (b) Building New Tools for Data Sharing and Re-use through a Transnational Investigation of the Socioeconomic Impacts of Protected Areas (PARSEC) (c) The portion of work relevant to this proposal are the objectives specific to developing leading practices, toolkits and workshops to support data sharing.
Intellectual Merit: We are building guidance for PIs and researchers that establish FAIR and OS data management practices from the time of the proposal development through to publication and the end of the grant. We are creating checklists that have an accompanying 10-minute tutorial (recording) to assist the researcher in what to do. Completed topics include: Digital Presence (ORCIDs and linking), Software Citation, and Data Management Plan Workbook. In the next reporting period we will create versions of these topics in Spanish, Portuguese, French, and Japanese. We have been very active in international meetings bringing awareness of the project to the broader community. The completed meetings include large events in the US, Europe, Brazil, and Australia. To highlight just one event, the Brazil Round Table about Reproducibility and Workflow held March 2021, organized by the PARSEC team and hosted by FAPESP, had over 200 student participants. **Broader Impacts:** The materials we are building are specifically designed to support Belmont Forum’s Data and Digital Output Management Plan and help their researchers and network of funding agencies to be better prepared to make data and software FAIR

and as open as possible. Our current Data Management Plan has over 500 views, and 372 downloads. Our international outreach continues in 2022 with a workshop at SRI2022 with an introduction to “Your Journey to Open Science”. (d) See Reference Section, “Shelley Stall Results from Prior Work”. (e) All research products from this grant are in compliance with our Data and Digital Output Management Plan.

Chris Lenhardt: (a) Award #1902537, $41,722, 07/1/2019-06/30/2021 (b) RAPID: Collaborative Research: Building Digital Infrastructure and Communities to Assess Risk of Drinking Water Hazards Caused by Hurricanes (c) This RAPID research project worked with water quality researchers collecting field samples and other data analysis post-disaster to ascertain their data management, cyberinfrastructure, and information needs. The context for identifying the water quality community research data needs was research experiences learned from two hurricanes. Due to trends related to hurricanes and floods, such as increasing occurrence, duration, and magnitude, there is greater need to understand post-disaster drinking water contaminants in order to protect and educate communities against such water hazards. Additionally, at-risk populations suffer disproportionately from disaster impacts. In this RAPID, we focused on the post-disaster recovery period as it represents a timely opportunity to engage the necessary stakeholders for assessing drinking water related health risks and to improve preparations for future water security particularly among the most vulnerable populations. The specific goals of this research are to 1) advance cyberinfrastructure that will make RAPID hazard data discoverable and usable in advance of the next hurricane season; and 2) develop a strategy to accelerate research through network-to-network collaborations. **Intellectual Merit:** Integrating hurricane, environmental, water quality, and health data in one cyberinfrastructure system will allow researchers to examine all aspects of natural-human coupled systems response to extreme weather events. Our research also contributed to the emerging field of geohealth informatics. Our research enhanced sharing information and increasing the abilities of communities who may use digital infrastructure to foster self-resiliency by identifying and improving the linkages between cyberinfrastructure tools and data sharing. We also found that networks, both newly created and existing such as among education and community networks, are essential for enabling hurricane research and benefit from efforts to sustain the networks. The project underscores the sociotechnical connections between individuals, communities, education and data sharing, and are as necessary for achieving the desired outcomes as obtaining the data itself. **Broader Impacts:** Hurricanes are high profile events that have catastrophic and lingering societal impacts particularly in under-represented communities. We developed community-led interactive learning activities coordinated through hydrology, hazard, and data science research community organizations and university teams to ensure that the data and network resources were broadly accessible to the research community from hurricane impacted regions. (e) Datasets derived or developed for this project are stored in HydroShare, providing access for research findings, The various resources developed as part of this project are intended to be reusable and are accessible via a project website. Workshop materials include recordings, resources, and educational materials. Additional details can be accessed via the Disasters & Water Quality Research Network group on LinkedIn.

Team members for the Community FAIR RCN

Laura Lyon, Program Manager at AGU providing support and coordination for the Data Leadership initiatives, workshop design and management, training, and document development. Laura is an experienced designer and coordinator of both virtual and in-person working sessions for large groups. Laura’s previous work at AGU has been essential to our executive office supporting working meetings and communications for our AGU Board and Leadership Council with nearly 200 members engaged in these two bodies. Laura has a Masters in Earth Science.

Lauren Parr, Senior Vice President, Meetings and Learning at AGU, oversees all learning initiatives and partnerships for the organization as well as a robust portfolio of global meetings and events. Her background is centered in instructional design, continuing education, curricula development, and online learning. She’s developed and implemented programs with positive outcomes for the American Society for Microbiology and the Regulatory Affairs Professionals Society, in addition to her work at AGU. Lauren has a Masters in Management with an emphasis in non-profit organizations.

Susan Shingledecker, the Executive Director of ESIP (Foundation for Earth Science Information Partners), bringing over 15 years of nonprofit and federal grant management experience. Her experience centers in community management and behavior change to lead broad adoption of new approaches. She has led efforts on a diverse range of issues including: sustainable tourism, renewable energy, marine debris, invasive species, expanded public lands and access and data driven ecosystem restoration. At ESIP she leads a professional volunteer community of over 180 partner organizations with more than 30 collaboration area groups that work to solve some of our greatest Earth science data challenges across sectors and domains.

Billy M. Williams is the Executive Vice President for Ethics, Diversity, and Inclusion at AGU. Williams was the Principal Investigator and lead organizer for the September 2016, NSF-funded workshop, Sexual Harassment in the Sciences: A Call to Respond and serves as a co-Principal Investigator on the 2017 NSF Grant, ADVANCE Partnership: From the Classroom to the Field: Improving the Workplace in the Geosciences. He currently serves as the Principal Investigator on Catalyzing Cultural Change in the Sciences with New Resources and Tracking Tools, a project funded by the Alfred P. Sloan Foundation, and is also the principal investigator on the 2020 NSF Grant for AGU LANDInG, a project to establish a DEI community of practice and new cohorts of DEI Leaders in the Geosciences.

ESIP Community Manager, a new position created by this RCN, will be focused on the Data Help Desk, knowledge base development, supporting the discipline repository community, many of which are members of ESIP, to help ensure their participation is efficient and give a place for ideas of improvement and providing common services to be discussed and addressed. ESIP is also a community where researchers and those developing tools can collaborate to determine better and more efficient ways to manage and analyze data+.

Consultants:

- Margaret O'Brien, University of California, Santa Barbara, Environmental Data Initiative, Data Management Expert
- Danie Kinkade, Woods Hole Institute of Oceanography, BCO-DMO, Data Management Expert
- Jane Wyngaard, University of Cape Town, Design Pattern Expert
- Karl Benedict, University of New Mexico, Design Pattern Expert

Partner Roles

AGU will administer and coordinate logistics for the project, including meetings support. Conceptual work and framing of the project will be led by RENCI and facilitation activities will be supported by AGU, RENCI, and ESIP. Partner organizations represent Research Infrastructure, National Academy Board, US Research Community Organizations, Libraries, Professional Societies, Research/Development Centers, Individuals and Consultants, International Funders and Informatics Communities. All organizations will have a role in outreach to their various networks as well as reporting, while AGU will manage the project website and related communications channels.

Table 4: Partner Summary, letters of support are provided

Type	Organization/Individual Name
US Research Community Organization	GO FAIR US, Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), ESIP Data Readiness Cluster, Ocean Leadership, Materials Research Data Alliance (MaRDA), Science Gateways Community Institute (SGCI), U.S. Geological Survey Community for Data Integration
Research Infrastructure	4TU.ResearchData, Neotoma Paleoecology Database, Environmental Data Initiative (EDI), Digital Rocks Portal (Maša Prodanović), CLIVAR and Carbon Hydrographic Data Office (CCHDO), US Antarctic Data Center (USAP-DC), Oak Ridge National Laboratory Distributed Active Archive Center, Incorporated Research Institutions for Seismology (IRIS) Data Management Center, DataONE, OpenTopography, FIESTA, EarthRef, National Ecological Observatory Network (NEON)
National Academy Board	Board on Research Data and Information of the National Academies of Sciences, Engineering, and Medicine
Library	University of New Mexico Research Data Services, Association of Research Libraries (ARL), Data Curation Network (DCN, Members - Cornell, Dryad, Duke, Johns Hopkins, NYU, Penn State, Princeton, U of Illinois, UC Santa Barbara, U of Michigan, U of Minnesota, U of Nebraska Lincoln, Virginia Tech, Washington U at St. Louis, Michael J. Fox Foundation – Data Exploration Network), UCLA Library

Type	Organization/Individual Name
Professional Society	Association of American Universities (AAU), American Meteorological Society (AMS) Board on Data Stewardship, Federation of American Societies for Experimental Biology (FASEB), American Astronomical Society (AAS), European Geosciences Union (EGU), Federation of Associations in Behavioral & Brain Sciences (FABBS)
Research/Development Center	Drexel University Metadata Research Center (MRC), National Center for Atmospheric Research (NCAR), Center for International Earth Science Information Network (CIESIN), Midwest Big Data Innovation Hub (National Center for Supercomputing Applications), Cooperative Institute for Satellite Earth System Studies at North Carolina State University, Unidata/UCAR, San Diego Supercomputer Center (SDSC), Science for Life Laboratory (SciLifeLab)
Consultants	Margaret O'Brien (University of California, Santa Barbara, Environmental Data Initiative), Laurel L. Haak (Mighty Red Barn), David Moroni (JPL, Caltech), Danie Kinkade (Woods Hole Institute of Oceanography, BCO-DMO), Jane Wyngaard (University of Cape Town, Design Pattern Expert), Karl Benedict (University of New Mexico, Design Pattern Expert)
International Funders and Informatics Communities	Research Data Alliance (RDA), CODATA (the Committee on Data of the International Science Council), Australia Research Data Commons (ARDC), OpenAire, Digital Research Alliance of Canada, DANS (Data Archiving and Networked Services, Netherlands), FAIRsFAIR / FAIR-IMPACT, CERN IT, World Data System

The American Geophysical Union (AGU) supports a community of over 130,000 experts to enthusiasts worldwide who explore the Earth and outer space in their roles at universities, research institutes, governments, corporations, and other non-profit organizations. AGU has a leadership role in adopting FAIR data and software in the Earth, space, and environmental sciences and in successfully organizing diverse stakeholders to meaningfully improve data practices. This includes: AGU's Position Statement on data (2019), Data+ guidance, Data Leadership advancing data+ sharing, and OS.

AGU draws on its inclusive vision and interconnected DEI/gender-related efforts and programming towards greater inclusion and participation in meetings, honors, reports, and publications. This extends to the proactive advocacy for DEI-related legislation, the expansion of community science impacts through the Thriving Earth Exchange, and increased resources towards creating a more diverse talent pipeline.

AGU also works with the following groups and initiatives that will be invited to participate in the RCN: The Coalition on Publishing Data in the Earth and Space Sciences (COPDESS), the authors and community advancing the TRUST Principles for digital repositories, the authors and community advancing the CARE Principles for Indigenous data, DataCite, GeoNet, the Council for Data Facilities, the Thriving Earth Exchange, and AGU Data Management Advisory Board.

Earth Science Information Partners (ESIP) is a global community of earth science and environmental informatics professionals. Founded in 1998, the ESIP Community includes over 180 partner organizations spanning federal and state governments, international organizations, leading academic research institutions and private sector technology firms. While ESIP's members are organizations, it is the people who make the community. Each comes to ESIP, to work collaboratively across sectors and domains to solve some of the greatest informatics challenges in Earth and environmental science. ESIP excels in community building and convenes over 30 working groups or clusters monthly. These groups can be technical in nature (Information Quality, Data Readiness, Information Technology and Interoperability) or domain specific (Biological Data Standards, Marine Data, or Air Quality). In addition to the work of these groups that happens throughout the year, ESIP has a reputation for hosting interactive workshops and conferences where the Earth science informatic community comes together to solve real challenges such as preparing and disseminating dataset quality information (Peng et al., 2022). Relevant to the proposed work, ESIP includes repository partners (institutional, federal and international, general and domain-specific). We see their role in the data lifecycle and this work as essential. They are a critical component in AGU's and ESIP's Data Help Desk which serves as a bridge between the informatics community and researchers in disseminating leading practices and the work that will be established through the design patterns workshops. The Data Helpdesk started at AGU unused space at the back of the AGU exhibit hall. Today the Data Helpdesk is an anticipated part of leading scientific society meetings like AGU, EGU, AMS, and GSA. Over the last two years in a virtual world the Data HelpDesk has expanded in scope to house a growing catalog of resources both cross-cutting and domain specific.

ESIP, through its community participation guidelines, works towards creating a welcoming environment that recognizes the diversity in its global community. It is vital that the community space is welcoming and supportive to all to which aim ESIP's new Equity, Diversity, Inclusion and Justice Advisory Committee is working to define and action priorities for the ESIP Community to infuse these principles in the community and work. The recognition of diversity extends to ESIP fellowship and scholarship opportunities, where they work with underrepresented groups through fellowship and scholarship programs to cultivate more participation from historically underrepresented groups in their community.

The Renaissance Computing Institute (RENCI) at the University of North Carolina at Chapel Hill, is a data science research institute with a mission to develop and use advanced cyberinfrastructure to advance domain-driven data science. RENCI's work includes multiple areas of collaborative research including their Earth Data Science Group, a Data Analytics Group, and a Network Research & Infrastructure group and RENCI's work synthesizes across domains. In addition to expertise in cloud computing, analytics, Earth system modeling, and bioinformatics, RENCI has a strong presence in the area of facilitating team science and complicated research collaboratories. RENCI also maintains significant on premise compute, storage, and network infrastructure. RENCI co-leads the South Big Data Hub, is the founder of the National Consortium for Data Science, and works with researchers from a range of schools, institutes, and centers at UNC such as the Institute for the Environment, School of Library and Information Science, and the School of Public Health.

RENCI, based out of UNC-Chapel Hill, incorporates the work out of the Diversity and Inclusion (D&I) Office ranging from fellowship and ambassador programs to training in topics such as anti-racism. More broadly, in research computing, RENCI is involved in programs to elevate D&I across the Big Data Hubs and drive career and training programs that look to improve D&I across higher education.

RENCI actively works to promote diversity, equity, and inclusion efforts as part of its work and mission. RENCI works with various units at UNC who focus on DEI and RENCI has its own internal group who champions DEI efforts. Several RENCI staff were part of UNC's inaugural cohort for a new certificate program launched in 2021, "Diversity, Equity, and Inclusion in Research (DEIR)". RENCI has developed or manages programs such as fellowship and ambassador programs with DEI focus. RENCI also seeks to provide leadership in these areas at a higher level through its projects.

Broader Impacts

We stake our description of the broader impacts of this project on the following core elements: 1) We explicitly include AGU's policy and guidance related to DEIA in developing an equitable and inclusive RCN; 2) We directly frame elements of the project to focus on linking FAIROS design patterns with DEIA outcomes; 3) Our approach provides outcomes that directly promote FAIROS and therefore facilitate the types of broader impacts associated with the science that can be enabled as a result. Accordingly, we anticipate outcomes that promote the research, pedagogical, and potential policy uses of data+ in under-represented communities. Usable design patterns that address barriers encountered by under-represented groups need to be created. Implementing FAIR data+ and OS improves trust and transparency and provides direct benefits to research communities for example accelerating convergent science and the potential to address grand challenges. This effort will also provide an approach to address other complementary problems such as reducing the misalignment between researcher experience, their workflows, and infrastructure solutions. As such it will also help address a broader culture change across the sciences in support of FAIROS. Easy access to data will significantly help the ability of researchers to work with communities (and thus the broader impacts of their grants) and the ability of community leaders to make decisions based on science data.

Collaboration Letters

Type	Organization/Individual Name
Research Infrastructure	• 4TU.ResearchData • Neotoma Paleoecology Database • Environmental Data Initiative (EDI) • Digital Rocks Portal (Maša Prodanović) • CLIVAR and Carbon Hydrographic Data Office (CCHDO) • U.S. Antarctic Data Center (USAP-DC) • Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) • Incorporated Research Institutions for Seismology (IRIS) Data Management Center • DataONE • OpenTopography • Framework for Integrated Earth Science and Technology Applications (FIESTA) / EarthRef • National Ecological Observatory Network (NEON)
Academy Board	• Board on Research Data and Information of the National Academies of Sciences, Engineering, and Medicine (NASEM)
U.S. Research Community Organization	• GOFAIR US • Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) • ESIP Data Readiness Cluster • Materials Research Data Alliance (MaRDA) • Ocean Leadership • Science Gateways Community Institute (SGCI) • U.S. Geological Survey Community for Data Integration (USGS CDI)
Library	• University of New Mexico Research Data Services • Association of Research Libraries (ARL) • Data Curation Network (DCN), Members - Cornell, Dryad, Duke, Johns Hopkins, NYU, Penn State, Princeton, U of Illinois, UC Santa Barbara, U of Michigan, U of Minnesota, U of Nebraska Lincoln, Virginia Tech, Washington U at St. Louis, Michael J. Fox Foundation – Data Exploration Network • University of California, Los Angeles Library
Professional Society or Federations	• Association of American Universities (AAU) • American Meteorological Society (AMS) Board on Data Stewardship • Federation of American Societies for Experimental Biology (FASEB) • American Astronomical Society (AAS) • European Geosciences Union (EGU) • Federation of Associations in Behavioral & Brain Sciences (FABBS)
Research/Development Center	• Drexel University Metadata Research Center (MRC) • National Center for Atmospheric Research (NCAR) • Center for International Earth Science Information Network (CIESIN) • Midwest Big Data Innovation Hub (National Center for Supercomputing Applications) • Cooperative Institute for Satellite Earth System Studies at North Carolina State University [See ESIP Data Readiness Cluster letter] • Unidata/UCAR • San Diego Supercomputer Center (SDSC) [See GOFAIR US] • Science for Life Laboratory (SciLifeLab)

Consultants	• Margaret O'Brien (University of California, Santa Barbara, Environmental Data Initiative) • Laurel L. Haak (Mighty Red Barn) • David Moroni (JPL, Caltech) • Danie Kinkade (Woods Hole Institute of Oceanography, BCO-DMO) • Jane Wyngaard (University of Cape Town, Design Pattern Expert) • Karl Benedict (University of New Mexico, Design Pattern Expert) [See University of New Mexico Research Data Services letter]
International Funders and Informatics Communities	• Research Data Alliance (RDA) • The Committee on Data of the International Science Council (CODATA) • Australia Research Data Commons (ARDC) • OpenAIRE • Digital Research Alliance of Canada • Data Archiving and Networked Services (DANS, the Netherlands) / FAIRsFAIR / FAIR-IMPACT • European Council for Nuclear Research (CERN) IT • World Data System (WDS)