EAGER: The FAIR Island Project for Place-based Open Science
California Digital Library (CDL), the University of California Natural Reserve System (UCNRS), the UC Berkeley Gump Research Station, Metadata Game Changers, and DataCite requests $296,234 to support our EAGER Project, FAIR Island.

Project Summary (1-page)

In response to the need for real-world implementations of Findable, Accessible, Interoperable, and Reusable (FAIR) data for Open Science (Wilkinson et al. 2016), the FAIR Island Project offers a working field station as a controlled environment in which to test the implementation of optimal FAIR data policies and workflows. The Project is appropriate for an EAGER proposal because there is a unique and timely opportunity to leverage the new field station on the atoll of Tetiaroa (where coPI Davies is Science Director) that has close ties to U.S. institutions through the University of California, including a nearby NSF Long Term Ecological Research (LTER) site, as well as international research programs. The Project does not readily fit into existing NSF programs because it involves a focus on an innovative application and integration of research data infrastructures and policies at a scale and pace that do not correspond to most NSF program requirements.The elements of this work align directly with the aim to explore and grow community readiness across all disciplinary areas served by NSF for the advancement of open science.

Intellectual Merit: The relationship between people, place, and data underpins some of the greatest challenges, and opportunities, of the 21st century. Place-based data have a unique quality in that they span the sciences and humanities with time and space (geolocation) acting as foundational metadata used to assign data to "place" (or nested "places"). The FAIR Island Project will build interoperability between pieces of critical research infrastructure -- Data Management Plans (DMPs), research practice, DOIs, and publications contributing to the advancement and adoption of Open Science. The California Digital Library (CDL)'s leading work on transforming static data management plans into continuously-updated, networked DMPs transforms the DMP into a living document that can guide the course of research by integrating data management activities and plans with related systems and workflows in the research lifecycle (NSF Award #1745675). Based on this proposed work, the Project's goal is to make federally funded research more accessible and usable, improving reproducibility and validation to enhance scientific rigor, and maximize impact by expanding across field stations within the U.S. Greater accessibility and utility will allow for automated monitoring and reporting of data-related activities associated with individual grants and will track outputs that are regularly included in NSF's Publicly Accessible Repository (PAR). It will also give individual researchers, research Projects and institutes, funders, and others the opportunity to measure the downstream effects of these practices on improving scientific rigor beyond PAR.

Broader Impacts: The impacts of FAIR data stewardship promoted by the FAIR Island Project are likely to be far reaching: problems with the reuse of data in Europe, for example, are estimated to "cost the EU at least €10 billion each year in the academic sector alone, and €16 billion in lost innovation opportunities" (Mons 2020). Through an open, community-oriented approach, the FAIR Island Project will develop generalized place-based tools and methods for FAIR data management and will cultivate a Place-based Data Management community of practice through the Research Data Alliance (RDA) to support adoption and reuse of these methods across other sites. The Project will promote increased public scientific literacy and direct engagement with investigational research and scientific data by considering the relationships, power differentials, and the historical conditions in the collection of data; these affect how ethical and socially responsible data use can occur in artificial intelligence (AI), particularly in the sciences. The Project includes regions of the United States and the Pacific Islands that include traditionally underrepresented groups in STEM. Through these connections, the FAIR Island Project will contribute to synergizing open science (FAIR) principles with emerging principles on Indigenous Data Governance (Carroll 2020).

I. Project Overview

In response to the need for real-world implementations of Findable, Accessible, Interoperable, and Reusable (FAIR) data for Open Science (Wilkinson et al. 2016), the FAIR Island Project offers a working field station as a controlled environment in which to test the implementation of optimal FAIR data policies and workflows. Place-based data have a unique quality in that they span the sciences and humanities with time and space (geolocation) acting as foundational metadata used to assign data to "place" (or nested "places"). Furthermore, all place-based data are tightly bound up with personal and social identity, as place is an inherently human construct - even wilderness being designated to some degree by policy-makers. While the issues for Research Data Management are thus as diverse as the diversity of people and places, they share common underlying themes and issues (scientific, ethical, legal, and social) and much can be learned from shared experience informing common standards, useful tools, and best practices, which in turn allows comparisons among and between places enabling mutual learning networks to emerge to achieve common goals, such as those laid out in the UN 2030 Agenda for Sustainable Development and associated "Decades".

The Project will leverage the UC-administered field station on the atoll of Tetiaroa in French Polynesia (where coPI Davies is Science Director) that hosts NSF-funded and international research programs and is closely aligned with research stations in Hawaii and California. Through a community-oriented, open approach, the FAIR Island Project will advance social and technological infrastructure for place-based open science through these objectives:

(1) **SOCIAL (COMPLIANCE): Develop and iterate on an exemplar place-based data policy.** In Year 1, create and adopt a place-based data policy for Tetiaroa, iterate on it, and have 90% of research groups agree to abide by it. Then, create a generalized data policy for other field stations to adapt and adopt. The Project will support adoption by additional sites by the end of this EAGER. **(Module 1)**

(2) **TECHNOLOGICAL (USER-DRIVEN INTEGRATION): Adapt**, **test, and prove the capabilities of networked data management plans (DMPs) for field stations:** Networked DMPs integrate and leverage a range of existing tools and platforms. Adapted for place-based research, they will be utilized for research projects, with Tetiaroa as the use case, as a key way to track provenance, attribution, compliance, deposit, and publication of all data collected at research sites. In Year 2, the Project will expand the use of networked DMPs to additional field sites. **(Module 2)**

(3) **SOCIAL (INCENTIVES): Demonstrate the benefits of FAIR data policies and workflows to all stakeholders for the field station.** Through the use of persistent identifiers, networked DMPs will lessen the administrative burden on field stations, as well as researchers and grant administrators. The Project will visualize a PID Graph (Cousijn et al. 2021) specific to Tetiaroa research to quantify the connectivity of researchers, organizations and research outputs and to show this change over time. **(Module 3)**

(4) **SOCIAL (GLOBAL EXPANSION): Cultivate a place-based data management community through the Research Data Alliance.** As indicated above, the FAIR Island Project is not unique to one site, but has broad applicability in place-based research at field stations, marine labs, LTER sites, and beyond. The Project workflows and tools will be available for use by all research facilities, initially expanding (beta phase) in California and the Pacific Islands (see below). In addition, the Project will share, request feedback and co-create this work openly, using RDA to increase the reach and adoption of these products. **(Module 3)**

The elements of this proposed work align directly with the aim to explore and increase community readiness across all disciplinary areas served by NSF for the advancement of open science. The project team aims to apply what is learned about optimal data practices to established research facilities to established research facilities, including hosts of NSF LTER sites, across California though the UC Natural Reserve System (UCNRS) and across the Pacific Islands through the '4Site Collaborative'. Based on this proposed work, informed by on-the-ground implementations of data management policies and requirements at these initial working field stations, the Project's goal is to expand use of the data policy and DMPTool template across field stations within the U.S. to make federally funded research more accessible and usable, improving reproducibility and validation to enhance scientific rigor and maximize impact.

II. Project Design

The Project is designed around three modules over the course of two years. The modules represent work packages that are carried out in parallel and that have important feedback among them. Below, we outline briefly the specific Tasks under each Module.

Module 1: Testing Data Policies and Protocols

Task 1.1 Development of Data Policies for Place-based Research: The FAIR Island Project has collected existing resources, developed a draft Data Policy, and started to solicit feedback from the 4Site network. A reusable place-based data policy template will be shared and potentially implemented across various field stations and marine labs (4Site and UCNRS). As the template is reused, it will continue to be refined. The development process is being done in an open, transparent way so that others will also be able to use this method for other research applications. The Project will deliver results via website, whitepaper, and conferences like the Research Data Alliance (RDA) Plenaries. Once the FAIR Island Data Policy has been finalized, FAIR Island Project personnel will incorporate these policies into a templated data management plan within the DMPTool application (see Module 2).

Task 1.2 Protocols for Measuring Efficacy and Compliance: All researchers accessing Tetiaroa Field Station will be required to accept the FAIR Island Data Policy and provide a machine-actionable FAIR Island DMP, which will utilize PIDs and ontologies designed to track provenance, attribution, compliance, deposit, and publication of all research data collected on the island (see Module 2). In this Task, we will work with The UC community, the 4site group, and the broader data community (through Research Data Alliance) to establish exemplar protocols for assessing compliance and evaluating the efficacy of the data policy and associated procedures and tools. To quantify efficacy, we will explore metrics and indicators for assessing and tracking how well FAIR data principles are being applied to pre-publication datasets (those not yet linked to a paper in a journal) as well as to datasets that are associated with published research products. The protocols will help to ensure that the impact of FAIR Island Project policies, procedures, and tools will be applicable in multiple scenarios. Our initial success metrics will based on these use cases: (i) making research products more easily accessible and usable, (ii) directly linking research data and software to their associated publications, (iii) making it easier for others to confirm the validity of scientific results through transparent provenance of data and software used.

Module 2: Adapting User-Driven Systems Integrations for networked DMPs

Task 2.1 User Stories and Technical Design: Unique to the FAIR Island Project, Tetiaroa is home to a set of highly interdisciplinary scientific Projects including:

- Rat Eradication: The invasive, non-native rats present on Tetiaroa provide an initial focal point for our Project. Researchers on the island are currently collecting data in advance of a massive rat eradication Project and will follow up with additional studies once the species has been removed. These data sets taken together offer the opportunity to examine the ecological responses, across the coupled marine and terrestrial environments, that occur when a keystone species is removed from an ecosystem. The FAIR Island Project will utilize the rat eradication Project as a real-life example from which to map out the data management policies and processes that are required to accelerate the rate at which multidisciplinary data are being collected, released, and made available for reuse.
- Tetiaroa Island Digital Ecosystem Avatar (IDEA): All data collected on Tetiaroa is contributed to the Tetiaroa IDEA Project. Tetiaroa IDEA is a place-based data science infrastructure and computational platform for scenario-based planning. The Tetiaroa IDEA will model the complex feedback between climate change, management actions, public policy, and ecosystem services across the island's coupled marine-terrestrial landscape.

Both of these scientific initiatives will test the data policy and its implementation on Tetiaroa to help validate approaches and to fine tune tools, based on interaction with the various research teams working through the field station. We will develop protocols for garnering participant input and feedback from researchers and other stakeholders; this process will include interviews with researchers to assess the burden of new requirements. The feedback received will inform improvements and identify functional requirements for future development and technical designs.

Task 2.2 Preliminary Development and Testing: The Project will implement FAIR DMP templates and workflows across the systems and register PIDs (with relational metadata) through DataCite to ensure that the relations can be mapped in the PID Graph. This work will extend the work of machine-actionable DMPs to include developing external tool integrations and connecting PIDs and their metadata in order to describe the connections and answer new questions about the relationships between the different entities. Initial integrations will include finalizing work between three key systems utilized by the FAIR Island Project: the Reserve Application Management System of the UC Natural Reserve System, the DMPTool of California Digital Library, and DataCite.

RAMS - the Reserve Application Management System of the UCNRS is already in use by 56 field stations primarily in California; 41 UCNRS stations, the UC Gump Station on Moorea and several other UC stations; four stations managed by the California State University; and several privately run stations. By connecting the RAMS system with the DMPTool, we will create a unified workflow wherein researchers are required to submit a DMP that is then reviewed by the field station director/staff and a research data management specialist on the FAIR Island Project team. Upon approval of the DMP, a DOI is minted and can be connected with the DOI(s) assigned for each associated output.

DMPTool - The cornerstone of support for DMPs in the U.S. is the DMPTool, developed in 2011 by the California Digital Library (CDL) and founding collaborators to provide an open, central clearinghouse of information about funder requirements, data management standards, and customized guidance. The tool now serves over 57,000 users from 294 affiliated institutions and resulting in over 54,000 plans. The Digital Curation Center (DCC), based in the UK, and CDL have a formal partnership to co-develop and maintain a single, open-source platform, for providing DMP guidance, this shared software (DMPRoadmap) underpins both the DMPTool and the DCC service, DMPOnline.

DataCite is a leading global non-profit membership organization that provides persistent identifiers (DOIs) for research data and other research outputs, including DMPs, software, and instruments. For the FAIR

Island Project, all outputs that have DOIs assigned to their DMPs will become discoverable and associated metadata is made available through DataCite's PID Graph. DataCite has developed additional services to make it easier for the community to connect and share their DOIs with the broader research ecosystem and to assess the use of their DOIs within that ecosystem. DataCite will offer the support of a software developer to this Project in order to link the work done with the FAIR Island grant Project to the wider research community.

Task 2.3 Design and build a dashboard for researchers and field stations to track research outputs: Project partners will produce a dashboard interface that visualizes all research assets (protocols, instruments, outputs, investigators, and organizations) and relationships between these research assets, utilizing connections made possible through the use of identifiers. The different entities within the DMP will be connected through their Persistent Identifiers (PIDs) including ORCIDs for people, DOIs for datasets and papers, and RORs for organizations. This dashboard will allow users to obtain information such as investigators who contributed to a research output, how to access the output, funding details, licensing information, and other important contextual metadata surrounding the output. The user interface developed in this Project will be web-based, providing consolidated access to information about investigators, articles, datasets, data management plans, grants/awards, organizations, instruments, software, and protocols.

Task 2.4 Implementing FAIR Island system integration: Reaping the benefits of Open Science requires collaboration and interoperability across multiple systems. We will integrate with systems that have implemented best practice in research data management. Three examples with which FAIR Island Project personnel have close ties are provided below, but more will be included as the Project proceeds.

GEOME - Genomic Observatories MetaDatabase (Deck et al. 2017) is a web-based database that captures the who, what, where, and when of biological samples (biosample) and associated genetic sequences, and represents all sample metadata as material samples. GEOME helps users ensure that biological samples are FAIR, improves the quality of user data and compliance with global standards (Yilmaz et al. 2011), and helps integration with downstream systems, such as GenBank (Benson et al. 2013). Integration between GEOME and the DMPTool will significantly enhance data stewardship (including DOIs) pertaining to biosamples generated from field stations.

ENRICH - Equity in Indigenous Research and Innovation Co-ordinating Hub aims to fast-track and scale out the development and implementation of Indigenous approaches to data ethics, collective privacy, data governance, digital infrastructure, and responsive policy. ENRICH promotes two data products that are of particular interest for the FAIR Island Project: Traditional Knowledge (TK) and BioCultural Labels. The FAIR Island Project will explore integration of TK and/or BC Labels into networked DMPs and how this might help operationalize both FAIR and the recent CARE Principles for Indigenous Data Governance (Alliance. 2019). Adding the Traditional Knowledge and Bio Cultural labels to the FAIR Island DMP template will enable automated notification of appropriate stakeholders and adds provenance information that will help to maintain trust with local communities regarding data availability and future use.

Data repositories are cornerstones to the collection and preservation of research data. Initially, we are targeting repositories across a representative range of those used by scientists, including generic, institutional, and specialised repositories. Examples include generalist repositories such as the Dryad Digital Repository and CERN's Zenodo. We also plan to explore similar work with NSF's Public Access Repository (NSF-PAR). Examples of subject-specific repositories include: NSF's Biological and Chemical Oceanography Data Management Office (BCO-DMO) and GenBank, the NIH genetic sequence database (Benson et al. 2013). Integration with additional repositories will likely be included in the Project as

feedback and requirements are gathered from researchers. DOIs generated for datasets via the DMP will directly link the products of research collected on the field station to their associated published datasets.

Task 2.5 Testing systems integration and iterative engineering improvements with ongoing user testing: We will include formal interviews with researchers to monitor the impact and burden of new requirements. Based on the results, we will make adjustments in the policies and/or procedures, including future technology developments/integrations. Improvements will include ongoing user-driven feature development, first from researchers at the Tetiaroa field station and then from researchers at other sites as soon as it is feasible (see Module 3).

Task 2.6 Production Release and Additional Interoperability: Production releases will include user documentation and training materials (online and live webinars). Additional technical integrations with affiliated systems (beyond those mentioned in Task 2.3) will be explored and implemented during the Project where possible.

Module 3: Driving Adoption, Generalizing, and Scaling

Task 3.1 Generalizing and Expanding: This task involves the implementation of Project-developed guidelines at other stations as well as further vetting and validating with key stakeholders (e.g., repositories) and early adopters. Lessons learned from the initial work on Tetiaroa will be incorporated into existing training and other materials and will inform adoption across the other partner field stations in the 4Site Network (4 Pacific Island stations) and UCNRS (41 California stations). Once the policies, procedures, and technical infrastructure are in place, we will continue to iterate on our systems to adjust to the realities on the ground at key stakeholder research sites. This practical application of the FAIR data principles allows us to experiment with varying workflows in order to identify the optimal system for achieving Project goals.

Task 3.2 Outreach and Scaling: We will develop training and educational resources for best practice in research data management. As the foundational FAIR Island infrastructure is developed and new procedures are rolled out to researchers, we will produce a series of open educational materials including user documentation and offer webinars to facilitate the onboarding of researchers in regards to data policies and procedures. The FAIR Island Research Data Management Advisor will coordinate the onboarding of researchers while experimenting with varying workflows in order to identify the optimal system for achieving Project goals. The Project will host a virtual workshop that will serve to share information about the FAIR Island Project with key stakeholders (invited Participants). Members of the Steering Committee (CDL) are supported to participate in Research Data Alliance (RDA) conferences as available to communicate the advances made in the FAIR Island Project as well as to garner feedback from the broader RDA community. We are leading a Place-Based Data Management Practices Birds of a Feather at the Virtual Plenary 17 and anticipate it growing into an interest group to coordinate place-based data management at field stations and beyond.

Task 3.3 Quantification and Analysis: The intent of FAIR practices is to make reuse of data and collaboration of data more efficient. In order to assess the downstream impacts of the FAIR Island initiative data management policies and procedures, the team will track the compliance of datasets generated from the Project. We will also assess how the Project has contributed to the assessment of physical and financial resources needed at field stations, marine labs, and place-based research networks to facilitate syndication of FAIR Island infrastructure and policies.

Initially, there will be extra effort needed from researchers as they may need to adopt new practices as described in 3.2. We will evaluate the amount of data wrangling needed over time and the amount of data reuse occurring with data collected from Tetiaroa utilizing methods developed by the Ocean Health Index team and described in Lowndes, 2017. Because this Project will promote adoption and use of identifiers, it will feed the PID Graph, developed by DataCite, which will allow us to use the PID Graph to visualize and quantify the extent of collaborations formed and connectivity of resources over time.

III. Management Plan & Personnel

The Project is led by the PI (Chodacki) at California Digital Library and co-PI Neil Davies at the UC Gump Research Station. The management of the Project is assured by a **Steering Committee** composed of co-PIs and key representatives of UCNRS, CDL, DataCite, Metadata Game Changers, and UC Berkeley (full list in Appendix >>>). Coordination of technical aspects of system integration among key systems at CDL and UCNRS will be led by Praetzellis. Robinson will serve as a bridge between the FAIR Island Project and researchers, a role that is especially important in Module 2, as we monitor how new policies and processes are being adopted. Robinson will also support the community-building described in Module 3. A Gannt Chart is provided in Appendix >>> to further detail the work plan. An **Advisory Board** will be established to provide input from relevant experts and stakeholders, such as representatives of data repositories (e.g., Dryad, BCO-DMO, Zenodo), tools and data infrastructures (e.g., GEOME, Protocols.io, DataONE, DataCite, San Diego Supercomputing Center, ENRICH, NSF Big Data Hubs), standards organizations (e.g., Genomic Standards Consortium, TDWG Biodiversity Standards), place-based research networks (e.g., LTER Network, Pacific Transect Collaborative | 4Site Network, European Marine Biological Resource Center, Smithsonian Global Earth Observatory and Marine GEO) and associations (e.g., National Association of Marine Laboratories, Organization of Biological Field Stations).

IV. Rationale and Outcomes

The Project is appropriate for an EAGER proposal because we have a unique and timely opportunity to leverage the establishment of a new international field station that has close ties to U.S. institutions through the University of California, including a nearby NSF LTER site. This association provides linkages to global standards while implementations include those designed and used at the UCNRS in California. The Project does not readily fit into existing NSF programs because it involves a focus on an innovative application and integration of research data infrastructures at a scale and pace that do not correspond to most NSF program requirements. Expected outcomes include: (i) **a dashboard and API building on connections made possible via the networked DMP that allows users to track connections** between DMPs, investigators, outputs, organizations, research methods and protocols and to display citations throughout the research lifecycle. All software development will be conducted in an open, transparent manner, with the code and documentation available on GitHub under an open-source MIT license. (ii) **Multiple research reports, white papers, and other published findings**, communications, and presentations that provide tangible, quantifiable metrics and statistics documenting Project findings and providing recommendations for implementations in other research scenarios. (iii) **Creation of Open Educational Resources** (OER), training materials, and instructional websites, ensuring good data management practices and policies accelerating research. (iv) **Contributions to ongoing FAIR policy and practice development** through collaborations with FAIR initiatives (San Diego Supercomputer Center hosted GO FAIR US Office, CODATA, RDA, FORCE11, etc.) and integration of Project findings, results, and outputs into ongoing international policy and guidance development.

V. Broader Impacts

The impacts of FAIR data stewardship promoted by the FAIR Island Project are likely to be far reaching: problems with the reuse of data in Europe, for example, are estimated to "cost the EU at least €10 billion each year in the academic sector alone, and €16 billion in lost innovation opportunities" (Mons. 2020). Aside from direct economic impacts, addressing the greatest challenges facing human society, such as pandemics and environmental change, requires access to well-described, high-quality data. The FAIR Island Project addresses some of the core cyberinfrastructure needed to address these issues. One consequence will be to facilitate the quantification of productivity of the nation's Field Stations and Marine Laboratories, something that has proven difficult despite qualitative assessments of the immense value of this national infrastructure (Council and Others. 2014). The Project includes regions of the United States and the Pacific Islands that include traditionally underrepresented groups in STEM. The Project will promote increased public scientific literacy and direct engagement with science and scientific data by considering the relationships, power differentials, and historical conditions in the collection of data, all of which affect how ethical and socially responsible data use can occur in artificial intelligence (AI), particularly in the sciences. The FAIR Island Project will contribute to synergizing open science (FAIR) principles with emerging principles (CARE) on Indigenous Data Governance that will build confidence between scientists and local people, encouraging communities to help enrich scientific data with local and traditional knowledge, as well as permitting researchers to access digital and environmental resources, and accelerating the translation of research knowledge to solutions that benefit society, promoting environmental and social justice.

VI. Intellectual Merit

The relationship between people, place, and data underpins some of the greatest challenges, and opportunities, of the 21st century. Relationships between human communities and their natural and built environments are increasingly mediated through digital data. These data feed models and algorithms, including applications of Machine Learning and Artificial Intelligence, that impact decision-making in a range of contexts and at nested scales of governance from the stewardship of smart cities and Indigenous lands to international agreements over global commons such as the High Seas, Antarctica, or the Earth's atmosphere. Digital representations of complex systems (digital twins or avatars) are emerging as technology platforms that harness the predictive power of scientific understanding (e.g., the consequences of climate change), while raising vital ethical, legal, and social issues, including who should control these capabilities and how.

There has been general enthusiasm for, and work towards, the FAIR principles in the research data and Open Science community. Examples include documents such as the founding Force11 FAIR Principles (Hagstrom. 2014), The Beijing Declaration on Research Data, and the recent Sorbonne Declaration of Research Data Rights (CODATA, Committee on Data of the International Science Council et al. 2019). Few of these recommendations or their resulting Projects, however, have evaluated the downstream effects of implementing FAIR data policies and procedures. Guidance and best practices for FAIR principles can be considered mature and established, yet outcome-based evaluation of the impact of these recommendations remains under-examined. This proposal outlines a Project to conduct such an examination, focusing on the outcomes and benefits of field-tested research work operating under FAIR principles. In focusing on the effects of FAIR data policies and practices via study of mid-lifecycle Open Science research, this work will have reciprocal benefits both to the further development of these principles and to document cases of their post-implementation value to researchers. The Project aims to promote broader adoption of FAIR practices by working towards a quantifiable understanding of the value and effectiveness of their practical application in research work. A recent report from the UK's Jisc

explored FAIR implementations within the UK academic research community and found that use of the FAIR principles was primarily limited to "discussion at a fairly conceptual level" (Allen and Hartland. 2018). In the United States, support for the FAIR principles has been widespread, but primarily at a similarly abstract level of consensus and agreement, not measured practice and outcome (Praetzellis. 2020). Mandated implementation of the principles at a procedural or workflow level has yet to materialize. **A key recommendation of the aforementioned Jisc report is to create "exemplars demonstrating FAIR in practice."** It is precisely this gap in evidentiary knowledge across this complete sequence of agreed-upon principles, rigorous, on-the-ground implementation, and tested, measured results that the FAIR Island Project will address.

Prior studies indicate that adherence to actions promised in DMPs for federally funded Projects is inconsistent. One study of twenty-five NSF-funded Projects found that 76% of researchers had not shared their data as described in the originally submitted DMP. Other studies have found that meeting the basic requirement of composing a data management plan has "very little bearing on whether or how datasets are shared" (Van Tuyl and Whitmire. 2016). CDL's leading work on transforming static data management plans into continuously-updated, networked DMPs transforms the DMP into a living document that can guide the course of research by integrating data management activities and plans with related systems and workflows in the research lifecycle (NSF Award #1745675). This integration or networking will allow for automated monitoring and reporting of data-related activities associated with individual grants. The result of this work is that it will allow individual researchers, research Projects and institutes, funders, and others the opportunity to measure the downstream effects of these practices on improving scientific rigor. By building interoperability between pieces of critical research infrastructure -- DMPs, research practice, DOIs, and published research outputs -- the FAIR Island Project is positioned to contribute to the advancement and community readiness for the adoption of Open Science.

VII. Prior NSF Support

In 2017, the California Digital Library was awarded a 2-year NSF EAGER grant to support machine-actionable data management plans (DMPs). The grant explored converting DMPs from a compliance exercise based on static text documents into a key component of a networked research data management ecosystem that not only facilitates, but also improves the research process for all stakeholders. As part of this EAGER grant, CDL contributed towards the development of a common data model for the creation of machine-actionable DMPs, produced by the RDA working group on DMP Common Standards. This model was recently released for community feedback and a version 1 of the scheme was approved in 2020. CDL implemented this metadata model into the shared, open-source DMPRoadmap codebase. This fundamental step toward machine-actionable DMPs form the foundation to enable information flow between DMPs and affiliated external systems in a standardized manner. Additionally, a grant-supported partnership with DataCite afforded updates to the DataCite metadata schema to better support DMPs and to optimize a workflow for generating DOIs for DMPs. By relying on the DOI infrastructure, the Project was able to utilize the Event Data service from DataCite to record when assertions have been made on the DOI. In 2019, CDL received an extension from NSF for this grant and work continues on developing the APIs needed to further explore the potential of machine-actionable DMPs. The FAIR Island Project is a natural extension of this work, providing a real-world test case to prove the capabilities of implementing DOIs for datasets utilizing the machine-actionable DMP and to analyse the downstream effects of its use in the resulting release of data.

References

Allen, Robert, and David Hartland. 2018. *FAIR in Practice - Jisc Report on the Findable Accessible Interoperable and Reusable Data Principles*. https://doi.org/10.5281/zenodo.1245568.

Alliance, Global Indigenous Data. 2019. "CARE Principles for Indigenous Data Governance." *GIDA Https://www. Gida-Global. Org/care*.

Anderson, Jane, and Kimberly Christen. 2013. "'Chuck a Copyright on It': Dilemmas of Digital Return and the Possibilities for Traditional Knowledge Licenses and Labels." *Museum Anthropology Review* 7 (1-2): 105–26.

Benson, Dennis A., Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. 2013. "GenBank." *Nucleic Acids Research* 41 (Database issue): D36–42.

Billick, Ian, and Mary V. Price. 2010. *The Ecology of Place: Contributions of Place-Based Research to Ecological Understanding*. University of Chicago Press.

Check, Erika. 2006. "Treasure Island: Pinning down a Model Ecosystem." *Nature* 439 (7075): 378–79.

CODATA, Committee on Data of the International Science Council, CODATA International Data Policy Committee, CODATA and CODATA China High-level International Meeting on Open Research Data Policy and Practice, Simon Hodson, Barend Mons, Paul Uhlir, and Lili Zhang. 2019. *The Beijing Declaration on Research Data*. https://doi.org/10.5281/zenodo.3552330.

Cousijn, Helena, Ricarda Braukmann, Martin Fenner, Christine Ferguson, René van Horik, Rachael Lammey, Alice Meadows, Simon Lambert, "Connected Research: The Potential of the PID Graph", *Patterns*, Volume 2, Issue 1, 2021, https://doi.org/10.1016/j.patter.2020.100180.

Council, National Research, and Others. 2014. *Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century*. National Academies Press.

Deck, John, Katharine Barker, Reed Beaman, Pier Luigi Buttigieg, Gabriele Dröge, Robert Guralnick, Chuck Miller, et al. 2013. "Clarifying Concepts and Terms in Biodiversity Informatics." *Standards in Genomic Sciences* 8 (2): 352–59.

Deck, John, Michelle R. Gaither, Rodney Ewing, Christopher E. Bird, Neil Davies, Christopher Meyer, Cynthia Riginos, Robert J. Toonen, and Eric D. Crandall. 2017. "The Genomic Observatories Metadatabase (GeOMe): A New Repository for Field and Sampling Event Metadata Associated with Genetic Samples." *PLoS Biology* 15 (8): e2002925.

Hagstrom, Stephanie. 2014. "The FAIR Data Principles." FORCE11. September 3, 2014. https://www.force11.org/group/fairgroup/fairprinciples.

Lowndes, J., Best, B., Scarborough, C. et al. Our path to better science in less time using open data science tools. Nat Ecol Evol 1, 0160 (2017). https://doi.org/10.1038/s41559-017-0160

Mons, Barend. 2020. "Invest 5% of Research Funds in Ensuring Data Are Reusable." *Nature* 578 (7796): 491.

National Academies of Sciences, Engineering, and Medicine, Policy and Global Affairs, Board on Research Data and Information, and Committee on Toward an Open Science Enterprise. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. National Academies Press.

Praetzellis, Maria. 2020. "Advancing FAIR and GO FAIR in the U.S.: Notes from the Workshop." California Digital Library: UC3. https://uc3.cdlib.org/2020/04/17/advancing-fair-and-go-fair-in-the-u-s-notes-from-the-workshop/.

Stucky, Brian J., John Deck, Tom Conlin, Lukasz Ziemba, Nico Cellinese, and Robert Guralnick. 2014. "The BiSciCol Triplifier: Bringing Biodiversity Data to the Semantic Web." *BMC Bioinformatics* 15 (July): 257.

Van Tuyl, Steven, and Amanda L. Whitmire. 2016. "Water, Water, Everywhere: Defining and Assessing Data Sharing in Academia." *PloS One* 11 (2): e0147942.

Walls, Ramona L., John Deck, Robert Guralnick, Steve Baskauf, Reed Beaman, Stanley Blum, Shawn Bowers, et al. 2014. "Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to

the Biological Collections Ontology and Related Ontologies." *PloS One* 9 (3): e89606.

Wilkinson, Mark D., Michel Dumontier, I. Jsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (March): 160018.
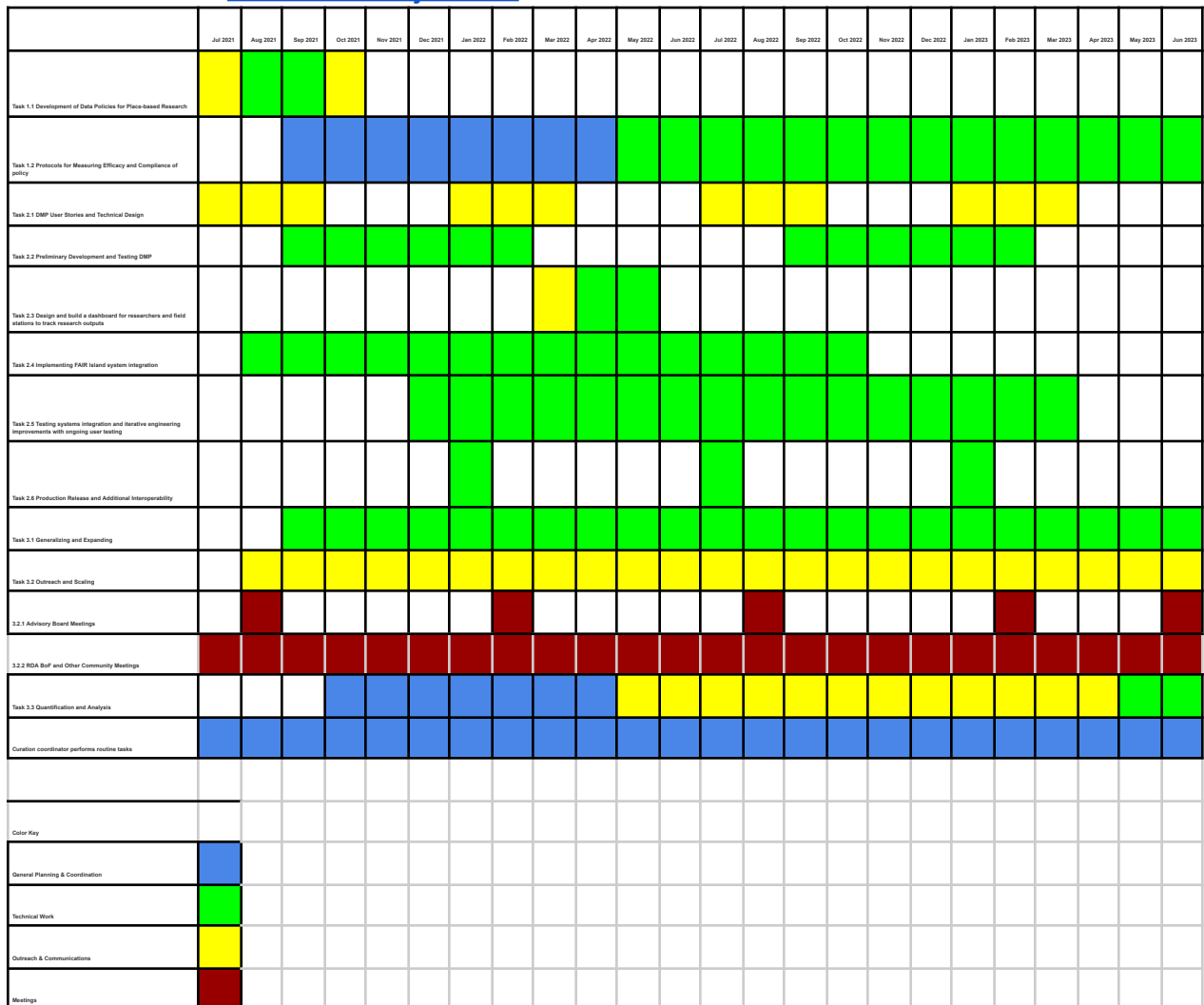
Yilmaz, Pelin, Renzo Kottmann, Dawn Field, Rob Knight, James R. Cole, Linda Amaral-Zettler, Jack A. Gilbert, et al. 2011. "Minimum Information about a Marker Gene Sequence (MIMARKS) and Minimum Information about Any (x) Sequence (MIxS) Specifications." *Nature Biotechnology* 29 (5): 415–20.

Appendix A: Steering Committee and Work Plan Gantt Chart

Steering Committee:
1. John Chodacki (California Digital Library) Principal Investigator
2. Neil Davies (UC Natural Reserve System) co- Principal Investigator
3. Matt Buys (DataCite)
4. Erin Robinson (Metadata Game Changers)
5. *Peggy Fiedler (UC Natural Reserve System)*
6. Karthik Ram (UC Berkeley)
7. Maria Praetzellis (California Digital Library)
8. Catherine Nancarrow (California Digital Library)
9. David Ackerly (UC Berkeley)

Draft Gantt Chart: FAIR Island Project Plan

	Jul 2021	Aug 2021	Sep 2021	Oct 2021	Nov 2021	Dec 2021	Jan 2022	Feb 2022	Mar 2022	Apr 2022	May 2022	Jun 2022	Jul 2022	Aug 2022	Sep 2022	Oct 2022	Nov 2022	Dec 2022	Jan 2023	Feb 2023	Mar 2023	Apr 2023	May 2023	Jun 2023
Task 1.1 Development of Data Policies for Place-based Research	🟨	🟩	🟩	🟨																				
Task 1.2 Protocols for Measuring Efficacy and Compliance of policy			🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 2.1 DMP User Stories and Technical Design	🟨	🟨	🟨				🟨	🟨	🟨		🟨		🟨	🟨					🟨	🟨	🟨			
Task 2.2 Preliminary Development and Testing DMP			🟩	🟩	🟩	🟩	🟩							🟩	🟩									
Task 2.3 Design and build a dashboard for researchers and field stations to track research outputs									🟨	🟩	🟩													
Task 2.4 Implementing FAIR Island system integration		🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩									
Task 2.5 Testing systems integration and iterative engineering improvements with ongoing user testing							🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩		
Task 2.6 Production Release and Additional Interoperability							🟩	🟩						🟩					🟩					
Task 3.1 Generalizing and Expanding			🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
Task 3.2 Outreach and Scaling		🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
3.2.1 Advisory Board Meetings		🟥						🟥						🟥					🟥					🟥
3.2.2 RDA BoF and Other Community Meetings	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥	🟥
Task 3.3 Quantification and Analysis			🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟩	🟩
Curation coordinator performs routine tasks		🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦

Color Key	
General Planning & Coordination	🟦
Technical Work	🟩
Outreach & Communications	🟨
Meetings	🟥

The Dear Colleague Letter (ref) and the report "Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support" (ref) emphasize the need for identifiers and data management plans put into practice.

It is precisely this gap in evidentiary knowledge across this complete sequence of agreed-upon principles, rigorous, on-the-ground implementation, and tested, measured results that the FAIR Island Project will address.