

Introduction

The FAIR Island Project's goal is to translate the broader FAIR principles into a set of specific requirements and implementable activities that demonstrate how good data management practices and policies accelerate research to the benefit of all stakeholders. The Project is a collaboration between the California Digital Library; Metadata Game Changers LLC; the University of California, Berkeley Gump South Pacific Research Station, located on the island of Moorea; and the Tetiaroa Society, which operates Tetiaroa's recently established field station. The Project started in early 2020 with generous support from the University of California and it is now funded by NSF Grant 2132549 ([full proposal text](#)). This first annual report covers initial work through December 2021.

The relationship between people, place, and data underpins some of the greatest challenges and opportunities of the 21st century. This project is place-based and comprises three 'places':

1. **FAIR Island Conceptual Model:** The FAIR Island concept model houses the generic, optimal data policy and technical infrastructure built on the FAIR Data Principles (Wilkinson et al., 2016). The conceptual model is realized at real field stations described in 2. It is iterated on with each implementation at a real site.
2. **Physical Site:** The Project will leverage the Tetiaroa Ecostation on the atoll of Tetiaroa in French Polynesia (where coPI Davies is Science Director) that supports NSF-funded and international research programs and is closely aligned with research stations in Moorea, Hawaii and California. Tetiaroa is part of the Society Islands in the South Pacific and lies about 35 miles from Papeete, French Polynesia, the closest international airport. The ecostation hosts research teams from all over the world to study reef environments, archeological sites, brackish lakes and deep ocean.
3. **Digital Twin of the Physical site:** Together with the actual site and the FAIR Island imaginary and building on the Island Digital Ecosystem Avatars (IDEA) Consortium (Davies et al., 2016), the third component of the FAIR Island is to establish a digital environment (digital twin) where all data and knowledge collected on Tetiaroa is curated and made openly available as quickly as possible, when it is appropriate to do so.

Based on the work of this EAGER and real-world experience at the Tetiaroa Ecostation, the project iterates to incorporate best practices with regard to ethical, legal, and social issues, such as application of the CARE principles for Indigenous data (Carroll et al., 2020) back into the Infrastructure Imaginary and expands to additional field stations.

Project Objectives & Accomplishments

Through a community-oriented, open approach, the FAIR Island Project advances social and technological infrastructure for place-based open science. In the section below, the project objectives, accomplishments through December 2021 and next steps are described:

- (1) 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.*

The FAIR Island data policy was developed based on Tetiaroa and generalized to be a place-based data policy for the FAIR Island Infrastructure Imaginary. Community input was solicited from the University of California Natural Reserve System (UCNRS), the 4Site Data Management Working Group, Tetiaroa researchers and at conferences including the Research Data Alliance plenaries and GO-FAIR-US. In addition, review was sought from legal experts, data experts and the policy was evaluated by FAIRsFAIR, as part of their [Policy Enhancement Support Programme](#).

Outputs:

- [Draft Tetiaroa Data Policy](#)
- [FAIRsFAIR evaluation of Tetiaroa Policy](#)
- [FAIR Island Data Policy v0.0.0](#) published to Zenodo

Next Steps: We hope that the Tetiaroa Society will adopt the draft Tetiaroa Data Policy. Beyond Tetiaroa, Santa Cruz Island Director, Jay Reti, has also expressed interest in using the data policy for Santa Cruz Island, a UCNRS field station. A second site using a version of the data policy would move our overall objective forward.

- (2) 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.*

The DMPTool is a critical component of the FAIR Island infrastructure. The DMPTool template for Tetiaroa was developed based on input from the FAIR Island Team; Tetiaroa Science Director, Neil Davies; and researchers coming to Tetiaroa. An initial DMP has been started for a few projects.

The Tetiaroa Ecostation uses the UCNRS Reserve Application Management System (RAMS). The CDL DMPTool Team completed the integration between RAMS and DMPTool. Upon submitting a research application to Tetiaroa, information from RAMS (such as the project abstract and collaboration details) is sent to the DMPTool via an API. This data transfer automatically creates a new DMP from the Tetiaroa

Society template. The applicant is then notified via email that they must complete their DMP before their application is approved. After testing this summer, this integration supported all new applications effective October 2021. After testing this summer, this integration supported all new applications effective October 2021.

Output:

- [Tetiaroa DMP Template](#) (Will need to login to DMPTool)
- [DMP Hub published page for Moorea Biocode project](#)

Next Steps: As we operationalize the Data Policy and DMP requirement at Tetiaroa Ecostation, applications will not be approved until the research team completes a DMP. We need to support researchers filling out the DMP and iterate the DMP based on feedback. DMPTool completed another integration with RSpace to link lab notebooks with DMPs and is planning to integrate [Protocols.io](#) into DMPTool. In 2022, DMPTool is planning to add functionality that will make it possible to upload existing DMPs.

- (3) *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.*

This year saw exciting steps forward for networked DMPs and built on a prior EAGER that CDL received. DMPTool includes space to identify ORCIDs for people and RORs. The Tetiaroa DMP Template also includes a reference to the ROR for the field station. The DMPTool now mints DOIs for published DMPs. Another responsive development from DMPTool, based on our experience, was to make it possible to link DOIs from outputs like articles, datasets and software to a published DMP.

Outputs: With the above functionality, we did a retrospective [DMP for the Moorea Biocode project](#). We were then able to search and identify research outputs from the project and link them to the DMP. Finally, we used DataCite's PIDGraph notebooks to draw our first PIDGraph. Figure 1 shows the evolution at the start with no outputs, and then the second and third images show articles and datasets as they are added.



Figure 1: PIDGraph for Moorea Biocode Project

Next Steps: The PIDGraph in Figure 1 centers a single project, the Moorea Biocode project, with the purple dot as a DMP DOI. Moving forward, we are working with DataCite to include RORs in DMP metadata. Utilizing RORs for field station identifiers, will allow us to draw an additional PIDGraph where the middle dot is a field station and the connections will be related projects (DMP DOIs). This view will provide one way to visualize the scientific impact a field station has. We are also working with DataCite to utilize the DataCite Commons page for facilities as a dashboard for field stations ([Gump Example](#))

- (4) *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.*

This Project has shared work at several meetings in the last two years including:

1. [RDA - Nov 2020](#)
2. [PIDapalooza 2021](#)
3. [GOFAIR US WIPS/CHIPS, Feb 2021](#)
4. RDA BoF session on Place-based research, April 2021
 - a. [Agenda](#)
 - b. [Slides](#)
5. [UCNRS Meeting, May 2021](#)
6. [RDA-US Webinar, August 2021](#)
7. [DataCite Community Call, September 2021](#)

8. [UC Libraries Forum, October 2021](#)
9. [Public Access Colloquim, December 2021 \(Agenda\)](#)

In December 2021, we invited three additional field station-related staff to participate in the FAIR Island Advisory Group. The advisory group includes: Dr. Neil Davies, Tetiaroa Society Science Director; Dr. Corinna Gries, co-PI of the Environmental Data Initiative; Dr. Jay Reti, Santa Cruz Island Reserve Director; and Dr. Nicholas Wolff, The Nature Conservancy.

Next Steps: Through this outreach, we expect to expand to other sites, including the Gump field station on Moorea, and to continue to engage with our Advisory group to both iterate on the infrastructure and grow the adoption across other field stations.

References

- Cousijn, H., Braukmann, R., Fenner, M., Ferguson, C., van Horik, R., Lammey, R., Meadows, A., & Lambert, S. (2021). Connected Research: The Potential of the PID Graph. *Patterns*, 2(1), 100180. <https://doi.org/10.1016/j.patter.2020.100180>
- Davies, N., Field, D., Gavaghan, D., Holbrook, S. J., Planes, S., Troyer, M., Bonsall, M., Claudet, J., Roderick, G., Schmitt, R. J., Zettler, L. A., Berteaux, V., Bossin, H. C., Cabasse, C., Collin, A., Deck, J., Dell, T., Dunne, J., ... Wood, S. (2016). Simulating social-ecological systems: The Island Digital Ecosystem Avatars (IDEA) consortium. *GigaScience*, 5(1), 14. <https://doi.org/10.1186/s13742-016-0118-5>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>