
Plan Overview

A Data Management Plan created using DMPTool

Title: EAGER: The FAIR Island Project for Open Science

Creator: Maria Praetzellis - **ORCID:** [0000-0001-5047-3090](https://orcid.org/0000-0001-5047-3090)

Principal Investigator: John Chodacki, Neil Davies, Erin Robinson

Data Manager: Maria Praetzellis, Brian Riley

Project Administrator: Catherine Nancarrow

Affiliation: DMPTool

DMP ID: <https://doi.org/10.48321/D17598>

Funder: National Science Foundation ([nsf.gov](https://www.nsf.gov))

Funding opportunity number: 51258

Template: NSF-GEN: Generic

Project abstract:

In response to the need for Findable, Accessible, Interoperable, and Reusable (FAIR) data for Open Science (Wilkinson et al. 2016), the FAIR Island Project is a multi-institutional initiative to deploy, test and iterate on optimal data policies and technical infrastructure. Our immediate goal is to build and showcase exemplar policies and technical infrastructure utilizing Digital Object Identifiers (DOI) for place-based research data that will allow all data, samples, and knowledge generated at field stations, marine laboratories, long-term ecological research (LTER) sites (and other networks spanning physical, biological and social sciences) to be made openly available as quickly as possible. The FAIR Island Project is coordinated through the University of California Gump South Pacific Research Station (host of NSF's Moorea Coral Reef LTER site) and the California Digital Library. Use cases include research programs addressing ocean acidification, ecosystem restoration, infectious disease, archeology, and community resilience, among others. Our initial focus is a new research station on the atoll of Tetiaroa in French Polynesia, where we can implement best practices in research data management from the ground up. In the next phase, we aim to include established research facilities across California (41 field sites in the University of California Natural Reserve System) and across the Pacific Islands through the '4Site Collaborative', a transect from Hawaii to French Polynesia involving University of Hawaii (HIMB marine lab; Oahu), The Nature Conservancy (Palmyra atoll research station; Palmyra), France's CNRS (CRIOBE marine station; Moorea), University of California (Gump Station; Moorea), and Tetiaroa Society (Tetiaroa atoll research station; Tetiaroa). Informed by on-the-ground implementations of data management policies and requirements at working field stations, the project will make federally funded research more accessible and usable, improving reproducibility and validation to enhance scientific rigor and maximize impact. Data management plans will be utilized as key documents for tracking provenance, attribution, compliance, deposit, and publication of all data collected at the research sites. The FAIR Island project will (i) build templated implementations and generalize the findings and outcomes from tracked, monitored implementation of FAIR principles utilizing machine-actionable data management plans (maDMP) and Digital Object Identifiers (DOI) for research data in a working environment, (ii) translate broader data management principles into a set of specific requirements and implementable activities for field stations, and (iii) demonstrate how best practices and policies accelerate research for the benefit of all stakeholders, and how they can be implemented at place-based research facilities across the United States and worldwide.

Start date: 07-30-2021

End date: 08-31-2023

Last modified: 11-29-2022

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text

does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

EAGER: The FAIR Island Project for Open Science

Types of data produced

The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

Three types of data will be generated during this project:

1. The project will produce persistent identifiers, utilizing the DataCite registry system to generate DOIs for research data.
2. Software produced or modified as part of the project will be maintained in the DMPRoadmap GitHub repository during and after the project (<https://github.com/DMPRoadmap/roadmap>), where it is available under an open source MIT License.
3. Miscellaneous research products include stakeholder feedback, project communications, and adoption tools and guidance. These products will be made available during the project via peer-reviewed articles, content on the FAIR Island project website and blog, presentations to conferences and webinars, and data repositories. The dissemination of resulting products will be coordinated by the project lead.

Data and metadata standards

The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

The project is designed explicitly to draw on global standards, with close collaboration with Research Data Alliance (RDA). For example, Specimen data will be digitized to conform to the Darwin Core (DwC), MixS, and Ecological Metadata Language (EML) standards. DwC was ratified in 2009 by Biodiversity Information Standards -Taxonomic Databases Working Group and has been internationally adopted and extended to specialized areas including gene bank data and georeferencing best practices. The MixS metadata standard is supported through the Genomic Standards Consortium and defines metadata for any type of sequence. EML is a metadata specification developed by ecologists and is used broadly to describe a variety of datasets.

Policies for access and sharing

Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

All data and products developed for the project will be made available for review and reuse during the course of the project and following its completion. Communications about products will take place on multiple channels (e.g., listservs, newsletters, social media) to reach multiple audiences. The dissemination of resulting products will be coordinated by the project lead.

DMPRoadmap is an active software product and will continue to be managed in the community GitHub repository. We will document current and future releases with release notes and the code for each version will be publicly available using Git history. Major versions of the software will be deposited in the Merritt Repository Service and findable with a persistent identifier.

All data produced by consuming partners during this research will be freely available to the public; we anticipate no sensitive or confidential data.

Policies for re-use, re-distribution, derivatives

Policies and provisions for re-use, re-distribution, and the production of derivatives.

All software products resulting from this project will be reusable and redistributable during the project and after its completion. The only restriction placed on redistribution of the software is that the copyright and license statement be kept intact as required by the MIT open source license. The software is expected to be of interest to national and international data infrastructure providers, data centers and repositories, institutional administrators, and individual researchers.

Plans for archiving and preservation

Plans for archiving data, samples, and other research products, and for preservation of access to them.

DMPRoadmap is an active software product and will continue to be managed in the community GitHub repository. We will document current and future releases with release notes and the code for each version will be publicly available using Git history. Major versions of the software will be deposited in the Merritt Repository Service and findable with a persistent identifier.

Planned Research Outputs

Software - "DMPRoadmap Codebase"

Software - "DMPHub Codebase"

The [FAIR Island project](#) will be using the [DMPTool](#) to allow researchers to create Data Management Plans. The DMPTool registers the DMP metadata once it has been approved. The DMPHub contacts EZID to register a new DMP ID (aka DOI) which then gets passed back to the DMPTool. When the user updates their DMP within the DMPTool, the updated metadata is sent to the DMPHub. The DMPHub will be able to communicate with systems like the PID Graph (via EventData) to find new related identifiers for a DMP. It could then update the original creator of the DMP (in this case, the DMPTool)

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
DMPRoadmap Codebase	Software	2023-08-31	Open	GitHub		MIT License	None specified	No	No
DMPHub Codebase	Software	2023-08-31	Open	GitHub		MIT License	None specified	No	No