

IMPLEMENTATION STORIES



FAIRSFair
Fostering Fair Data Practices in Europe



THEME
4

**SUPPORTING DATA MANAGEMENT
PLANNING FOR FAIR**

IMPLEMENTATION STORIES

THEME 4 SUPPORTING DATA MANAGEMENT PLANNING FOR FAIR

Tagging and tracking outputs with machine-actionable DMPs – the FAIR Island project

Authors

Patricia Herterich, Angus Whyte
(Digital Curation Centre, University of Edinburgh)

From interviews with

Erin Robinson (Metadata Game Changers), Maria Praetzelis (California Digital Library), Neil Davies (University of California Gump South Pacific Research Station, Berkeley Institute for Data Science (BIDS), Tetiaroa Society), Kristian Garza (DataCite)



■ Introduction

The FAIR Island Project for Place-based Open Science is an exploratory project that is comprehensively testing the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles in an Open Science context, aiming to apply these principles from the start of data gathering. This data gathering is initially centred on the island atoll of Tetiaroa, which is northwest of Tahiti in French Polynesia, yet offers general lessons applicable anywhere.

A key goal of the FAIR Island project is the development of an exemplar place-based data policy that provides the policy framework to require researchers at the field stations to not only generate Data Management Plans (DMPs) but to share research data upon completion of their projects. This story shows how new features from DataCite can be used to automatically produce an inventory of project outputs for a field station. With this aim, the FAIR Island project illustrates how to build into research practice the interoperability between DMPs and identifier systems, which includes DOIs, RORs and ORCIDs. By ensuring information contained within DMPs is machine-actionable, this can save researchers and field station managers time on project administration and allow systems to leverage DMPs as 'living documents' to automatically record project outputs. The FAIR Island Project is a collaboration between the California Digital Library (CDL), University of California Gump South Pacific Research Station, Berkeley Institute for Data Science (BIDS), Metadata Game Changers LLC, and DataCite. This implementation story highlights initial results and recommendations, based on a webinar for Research Data Alliance US (Robinson et al., 2021) and follow-up with the presenters.

FAIRsFAIR recommendation

"Formalise and support appropriate data management plans (DMPs) for FAIR data"

FAIRsFAIR Recommendations on practice to support FAIR principles

Approach taken

The goal of the work described in this story is to track all researchers, organisations, funders and research outputs generated or completed at field stations. The FAIR Island Project for Place-based Open Science builds on three major components to achieve their vision of developing “the optimal data policies and technical infrastructure necessary to create an environment where all data and knowledge collected at a place is curated and made openly available as quickly as possible.” These three components are:

- A concise, easily understood **data policy** that lays out expectations for researchers;
- **Data Management Plans (DMPs)** that translate the policy into actionable plans for researchers; and
- **Integrations between existing infrastructure** to allow for continuous analysis and improvement.

The Tetiaroa Ecostation based on the Tetiaroa atoll in the Pacific ocean and managed by the Tetiaroa Society, a US501(c)3 nonprofit is in an ideal position to trial this. The Tetiaroa Society acts as stewards of the place and the data allowing for experimentation with the local data policy and technical environments for research, to demonstrate advantages that can be gained from FAIR data.

The first focus area for the FAIR Island team was to work on a data policy, with the goal to agree on a draft that allows for the curation and sharing of data as early as possible¹. The draft policy’s key message can be summarised as follows: *“Participants in research projects agree to make their metadata and prepublication data available to each other through a “Tetiaroa Data Trust” (TDT) that facilitates data re-use according to mutually agreed terms and a code of conduct. The TDT is maintained by Tetiaroa Society to support decision-making for wise and equitable stewardship of the atoll. The TDT follows FAIR+CARE data principles and is distributed across repositories that meet global standards in research data stewardship.”* (FAIR Island Project, 2021)

■ Leveraging machine-actionable DMPs and the PID Graph

In addition to creating a data policy, the team started working on leveraging two innovations in the infrastructure for research data management: machine-actionable DMPs and the ‘PID Graph’, a term coined in the FREYA project referring to the connections that can be made between researchers, organisations, and outputs based on the persistent identifiers (PIDs) recorded in metadata about these research outputs².

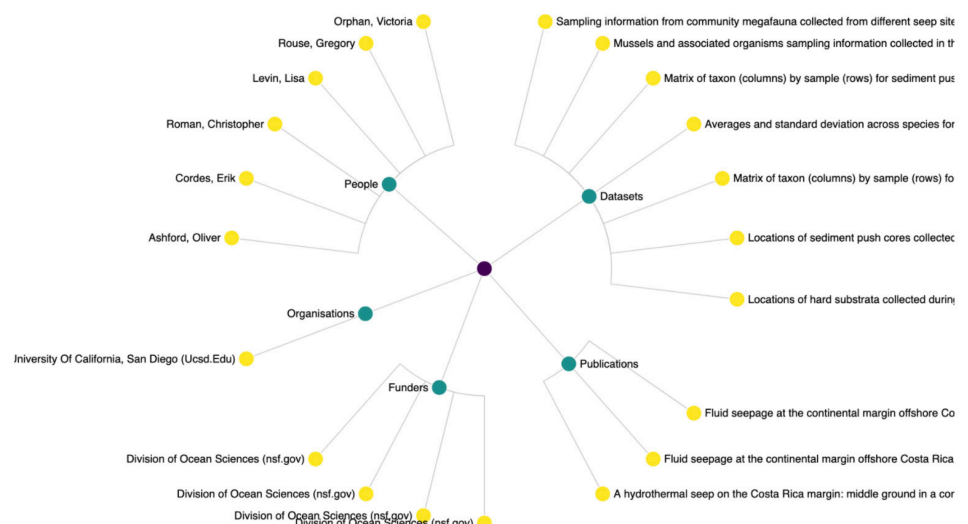


Figure 1: A visualisation of a DMP in the PID Graph (DataCite, 2021)

1. Davies, Neil, Chodacki, John, Praetzelis, Maria, Nancarrow, Catherine, & Robinson, Erin. (2021). Generic Place-Based Research Data Policy (V0.0.0). Zenodo. <https://zenodo.org/record/5735528#.Ybj1IH3MLtg>

2. <https://www.project-freya.eu/en/blogs/blogs/the-pid-graph>

IMPLEMENTATION STORIES

THEME 4 SUPPORTING DATA MANAGEMENT
PLANNING FOR FAIR

'Generating DMP IDs creates an unbreakable link between a data plan to the project outputs and allows access to DataCite's supporting services such as Event Data to facilitate connections via the PID Graph.'
(Garza & Buys, 2021)

Challenges encountered and addressed

FAIR Island also implements the RDA Common Standard for machine-actionable DMPs (maDMP) (Miksa, Walk & Neish, 2020) in the DMPTool, a tool for supporting DMPs (Praetzelis, 2019). Using this standard allows the DMPTool to exchange information commonly recorded in a DMP with other systems used by an organisation to manage research. The project team has taken that standard and coupled it with the DOI system to ensure straightforward ways of delivering maDMP information to machines and humans alike.

The FAIR Island project also integrated maDMPs into research project workflows through the field station reservation system: Reserve Application Management System (RAMS). As is common with most field stations, researchers at the Tetiaroa Ecystation need to use this to apply for the opportunity to conduct their research project at a field station. With the updates made by the FAIR Island team, once an application is approved, a DMP is created in DMPTool using information provided in RAMS to prefill the DMP with project information. The researchers are then asked to complete their DMP and once that is done, the field station director is notified.

Once the DMP is approved, a DMP ID is created using DataCite, which is the starting point for the mappings to project outputs that can be visualised from the PID Graph. The DMP ID couples a custom metadata schema that aligns with the maDMP Common standard and DOI infrastructure for delivery and exchange of applicable information. Any DataCite member can mint DMP IDs by utilising the "OutputsManagementPlan" resourceTypeGeneral³.

By using the DOI infrastructure, this approach allows for quick mapping to other identifiers and research outputs. This results in a 'research graph', a broader term for PID graphs that has resulted in services like Open Aire Research Graph⁴ and DataCite Commons⁵. These provide a search engine allowing users to find research outputs that are linked through their PIDs. DataCite Commons offers a search interface for navigating these connections, by filtering a search for published works based on various facets. These filters include the people contributing, organisations involved, and the repositories holding the works, as well as more typical variables like the publication year or type.

The FAIR Island team is focused on exploring and resolving downstream issues that arise when DMPs are not linked to research outputs. These links may be missing because, for example, metadata about the DMPs are not publicly available as DMP IDs are a recent development, the outputs have not been shared in a repository, or the metadata has not been completed to identify related works.

3. See DataCite documentation at: <https://support.datacite.org/docs/create-a-dmp-id>

4. <https://graph.openaire.eu>

5. <https://commons.datacite.org>

IMPLEMENTATION STORIES

THEME 4 SUPPORTING DATA MANAGEMENT
PLANNING FOR FAIR

Potential impacts

To show what can be done if the gap is filled, the team recently created the illustrative example shown in Figure 1, using a project that has ended (Moorea Biocode). As a proof of concept, they created a DMP in DMPTool and retroactively linked it to research outputs from the project. (Davies, 2021) This was done by a curator manually searching through literature databases and papers to create a list of publications. This list is incomplete and therefore shows the general challenges to linking research outputs with the projects or grants they came from. As next steps, the team will explore ways to show that this can be effective only when both the systems and tools are interlinked as well as highlight the metadata that flows across these links.

One major potential benefit of the FAIR Island Project will be the ability to avoid repeated effort to compile information about research projects. Additional value will potentially flow from the use of established infrastructure, benefiting other stakeholders, as DataCite Commons and the PID graph offer new ways to analyse connections between people, their organisations and their work. Neil Davies, Tetiaroa Society Vice President and Science Director, offers a vision for how the integrated setup can help research administrators to see all the products that are created as part of the research, including material samples (RDA US, 2021). And although not explicitly covered by the project, there are cost savings that might occur as part of the systems being integrated and thus researchers not having to re-enter the same information in several systems. (Brown et al., 2021)

Using maDMPs with the PID Graph also enables researchers to do 'integrative science' by developing more complex models that can explore connections across different projects more easily. In the context of place-based research, getting an overview of the research carried out in a specific location can be especially important. This information can be used as a starting point for research engagement with local communities. The new infrastructure should allow the researchers to feed back any insights gained and support the local community with any follow-up impact, putting the CARE principles (Carroll et al., 2020) into practice and supporting compliance with permit requirements and other regulations for field research.

To explore these wider potential impacts the FAIR Island Project also seeks to establish a RDA group on place-based Open Science⁶. In that context, project results and lessons learned will continue to be shared with others working in similar circumstances.

6. <https://www.rd-alliance.org/place-based-open-science-joint-session>

IMPLEMENTATION STORIES

THEME 4 SUPPORTING DATA MANAGEMENT
PLANNING FOR FAIR

Recommendations

The FAIR Island project is still ongoing, so the project team recommendations are still emerging from the pilot. So far they include the following points :

1. Field stations and research facilities should actively make use of DataCite's resource types and label DMPs as type "OutputsManagementPlan" when making them or their metadata available and thus, enable tracking the impact of a data management graph through the PID graph.
2. Researchers should be encouraged to provide links between various research outputs as well as outputs and funding information to enhance connections in the PID graph.
3. Where not provided by researchers, curators should use PIDs to create links between various research outputs as well as outputs and funding information to enhance connections in the PID graph.
4. Systems should use PIDs where possible to automatically fill and enrich information in the system and showcase the benefits of the PID graph.

Further information

Brown, J., Jones, P., Meadows, A., Murphy, F., & Clayton, P. (2021). UK PID Consortium: Cost-Benefit Analysis (1.0). Zenodo.

<https://doi.org/10.5281/zenodo.4772627>

Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J.D., Anderson, J. & Hudson, M., (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), p.43.

DOI: <http://doi.org/10.5334/dsj-2020-043>

Davies, N.. (2021). "Moorea Biocode" [Data Management Plan]. DMPHub.

<https://doi.org/10.48321/D1F88S>

FAIR Island Project (2021, March 17). Data Policy for FAIR Island Project nearing completion – Fair Island. Retrieved 2 February 2022, from <https://www.fairisland.org/2021/03/17/data-policy-for-fair-island-project-nearing-completion/>

Garza, K., & Buys, M. (2021). A Brave New PID: DMP-IDs.

<https://doi.org/10.5438/J22A-5D79>

Miksa, Tomasz, Walk, Paul, & Neish, Peter. (2020). RDA DMP Common Standard for Machine-actionable Data Management Plans.

<https://doi.org/10.15497/rda00039>

Praetzellis, M., Chodacki, J., Davies, N., Robinson, E., Buys, M., & Nancarrow, C. (2021). EAGER: The FAIR Island Project for Place-based Open Science. Zenodo.

<https://doi.org/10.5281/ZENODO.5117892>

Praetzellis, M. (2019, October 14). What's new with our machine actionable DMP work? DMPTool Blog.

<https://blog.dmptool.org/2019/10/14/whats-new-with-our-machine-actionable-dmp-work/>

RDA US Region. (2021, August 27). FAIR Island: Networked, Machine-Actionable DMPs for Open Science.

<https://www.youtube.com/watch?v=dua8a7FYDao>

Robinson, E., Praetzellis, M., Davies, N., & Garza, K. (2021, August 12). FAIR Island: Networked, Machine-Actionable DMPs for Open Science. RDA.

<https://www.rd-alliance.org/fair-island-webinar-Aug2021>

■ *About FAIRsFAIR Implementation Stories*

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report [Turning FAIR into Reality](#) and the [FAIRsFAIR Recommendations on practice to support FAIR principles](#). FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

■ *FAIRsFAIR - Fostering FAIR Data Practices in Europe*

European Commission Grant Agreement No. 831558
H2020-INFRAEOSC-2018-4
www.fairsfair.eu - support@fairsfair.eu

■ *Acknowledgements*

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

■ *Disclaimer*

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

■ *Copyright rests with the authors*



This work is released under a Creative Commons Attribution License, version 4.0.
For details please see <https://creativecommons.org/licenses/by/4.0/>

