

You shoulda put a PID on it: Leveraging the PID Graph for DMPs

Matt Buys, John Chodacki, Kristian Garza, Maria Praetzellis, Brian Riley

PIDapalooza
January 27, 2021



Agenda

Project background

PIDs for DMPs

DataCite work

Modeling connections via the PID Graph

Demo time!



Instructions for the DEMO section

Visit <https://rb.gy/8o3hba>

Keep the Window in the background





Platform for DMP creation and guidance with 43 templates for 17 US funders (NSF, NIH, DOE, DOT, etc.) and international funders

Goal to create a bridge for researchers to discuss research data with librarians and support staff

55k+ users with 52k+ data management plans at 294 participating institutions



DMPRoadmap & DMPTool

Platform for DMP creation and guidance with the goal of facilitating the creation of effective DMPs reflecting today's best practices

Harness community development efforts

Create next-generation, machine-actionable DMPs



What is a networked DMP?

*The goal of a machine-actionable DMP is to facilitate a research data management (RDM) system that **allows data and information about research to be communicated and shared across stakeholders**, linking metadata, repositories, and institutions, and allowing for notifications and verification, real-time reporting, and automated compliance. If leveraged appropriately, maDMPs have the ability to lessen the administrative burden on researchers and grant administrators.*

- Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support.
<https://doi.org/10.29242/report.effectivedatapactices2020>



NSF EAGER research outline

Explore the potential of machine-actionable DMPs as a means to transform the DMPs from a compliance exercise based on static text documents into a key component of a networked research data management.

- [NSF Funded EAGER Research Grant](#)
- Started based on the framework set out in RDA Working Groups
- Testing the hypothesis that we can connect DMPs to PID graph via the Common Standard
- Isolate gaps that we need to compensate for
 - New identifiers needed?
 - How to handle versioning of DMPs over time?



Features that network research activities

DMPTool (via DMPRoadmap) currently supports PIDs within a DMP:

- **DMP IDs** (in test environment only)
- **RORs** for research organizations
- **Funder Registry IDs** for funders
- **ORCiDs** for DMP creators and collaborators
- **Registry of Research Data Repositories** (re3data)

[API](#) exchanges information about DMPs that are compliant with RDA Common Standard Metadata Schema

Ability to export plans as RDA compliant JSON



Modeling connections via the PID Graph



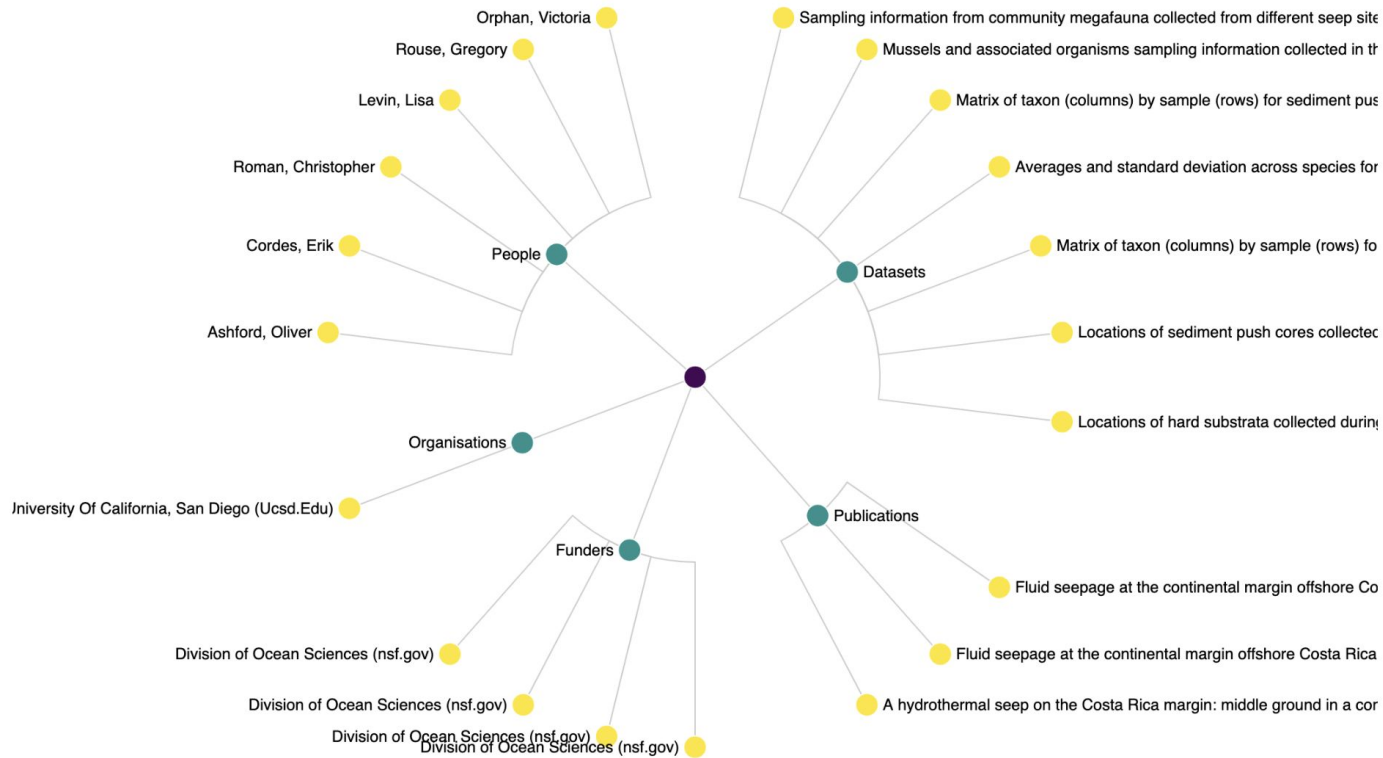
PIDs for DMPs

Generating identifiers for DMPs create 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.



[From Flickr by highwaysengland. CC BY 2.0](#)





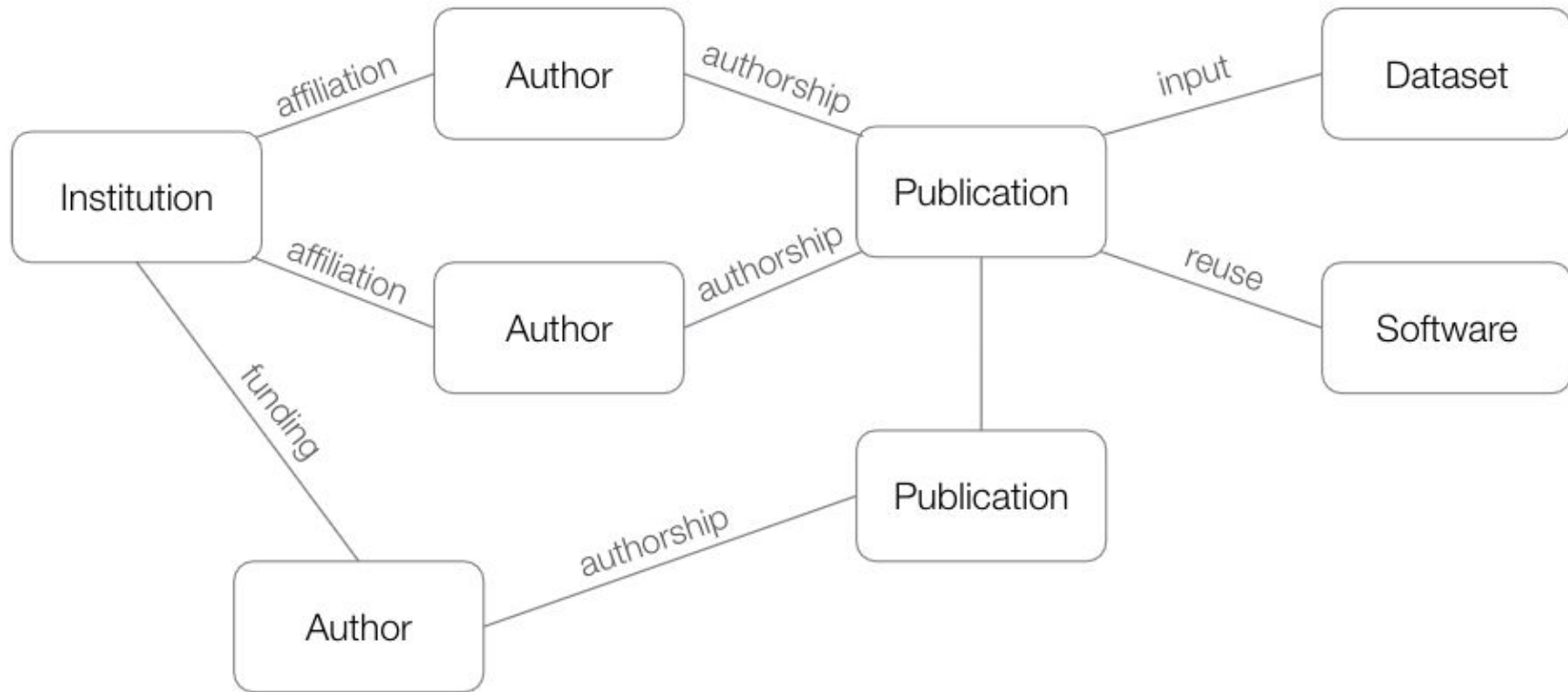
Leveraging the PID Graph

Having unique persistent identifiers for researchers and their outputs is crucial to connecting pieces of the research landscape together.

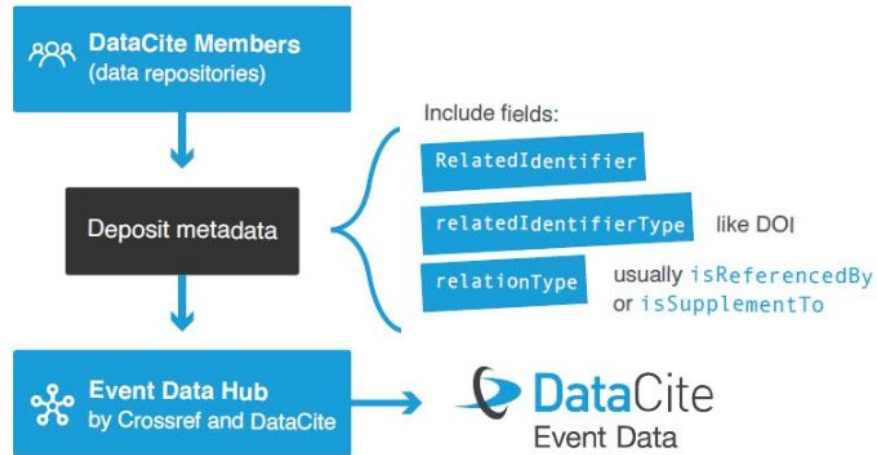
PIDs already have the potential to enable the connected research graph, but we're not yet taking full advantage of their connecting powers.

We can now clearly link PIDs together via relations in their metadata to enable the discovery of connections at least two "hops" away





Metadata



Interested in using this information? Find out more at: <https://support.datacite.org/docs/eventdata-guide>

BCO-DMO

- DMPs are often generated at the beginning of a research project. We however wanted to show the potential connections that could be made with a DMP over the lifetime of the entire research project.
- Our partners at BCO-DMO supplied us with historical data in a series of CSV files. The data included metadata for:
 - Research projects
 - DMPs
 - Contributors
 - Research outputs (e.g. Datasets and publications)
- We transformed the raw data into the new RDA metadata standard for DMPs: <https://rda-dmp-common.github.io/RDA-DMP-Common-Standard/>

```

{
  "dmp": {
    "schema": "https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master/examples/JSON/JSON-schema/1.0",
    "title": "Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic ...",
    "description": "The deep ocean supplies food, energy, novel drugs and materials, and plays essential roles in the ... ",
    "dmp_id": {
      "type": "URL", "identifier": "https://www.bco-dmo.org/project/648472/plan/1421"
    },
    "contributor": [{
      "name": "Cordes, Erik",
      "affiliation": {
        "name": "Temple University (temple.edu)",
        "affiliation_id": {
          "type": "ROR", "identifier": "https://ror.org/00kx1jb78"
        }
      },
      "contributor_id": {
        "type": "ORCID", "identifier": "https://orcid.org/0000-0002-6989-2348"
      },
      "roles": [
        "http://credit.niso.org/contributor-roles/investigation",
      ]
    }
  ],
  "project": {
    "title": "Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic ...",
    "start": "2016-10-01T00:00:00Z",
    "end": "2020-09-01T00:00:00Z",
    "funding": [
      {
        "name": "Division of Ocean Sciences (nsf.gov)",
        "funder_id": {
          "type": "ROR", "identifier": "https://ror.org/05wqqhv83"
        },
        "grant_id": {
          "type": "URL", "identifier": "https://www.nsf.gov/awardsearch/showAward?AWD_ID=1635219"
        },
        "funding_status": "granted"
      }
    ]
  }
}

```


DMPHub

- Many of the DMPs that BCO-DMO identified were not generated within the DMPTool.
- We needed a place to store the metadata and provide a persistent landing page for our new DMP IDs
- We developed the DMPHub as a repository for DMP metadata. The new system:
 - Allows RDA common standard metadata to be loaded via an API
 - Acquires a DMP ID from Datacite
 - Provides a landing page for the DMP with links back to the original document whether it be a PDF or resides in a system like the DMPTool

A data management plan for this project was submitted to the Division of Ocean Sciences (nsf.gov) with the information below.



This page represents key information from a data management plan available here.

Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic communities and the surrounding deep sea

Contributors to this project

Ashford, Oliver: Data-curation, Investigation, University of California, San Diego (ucsd.edu), <https://orcid.org/0000-0001-5473-7057>

Cordes, Erik: Data-curation, Investigation, Temple University (temple.edu), <https://orcid.org/0000-0002-6989-2348>

Levin, Lisa: Data-curation, Investigation, University of California, San Diego (ucsd.edu), <https://orcid.org/0000-0002-2858-8622>

Orphan, Victoria: Investigation, California Institute of Technology (caltech.edu), <https://orcid.org/0000-0002-5374-6178>

Roman, Christopher: Investigation, University of Rhode Island (ww2.uri.edu), <https://orcid.org/0000-0002-9185-4532>

Rouse, Gregory: Investigation, University of California, San Diego (ucsd.edu), <https://orcid.org/0000-0001-9036-9263>

Project details

Project Start: October 01, 2016

Project End: September 01, 2020

Created: October 20, 2020

Modified: October 20, 2020

Ethical issues related to data that this DMP describes?: no

For more information visit the project landing page: Unknown

Citation

When citing this DMP use:

Ashford, Oliver. (2020). "Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic communities and the surrounding deep sea" [Data Management Plan]. DMPHub. <https://doi.org/10.26008/1912/bco-dmo.805488.1>

When connecting to this DMP to related project outputs (such as datasets) use the ID:

<https://doi.org/10.48321/D17G67>



their transition to background systems at sea (and on land), engage with K-12 students through social media and in-class presentations and exhibits.

Planned outputs

Dataset for: Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic communities and the surrounding deep sea

Format: Dataset

Anticipated volume: unknown

Release timeline: unknown

Intended repository: unknown

License for reuse: unknown

Other works associated with this research project

Articles

Levin, L., Orphan, V., Rouse, G., Rathburn, A., Ussler, W., Cook, G., Goffredi, S., Perez, E., Waren, A., Grupe, B., Chadwick, G., Strickrott, B. (2012). "A hydrothermal seep on the Costa Rica margin: middle ground in a continuum of reducing ecosystems" [Article-journal]. In The Royal Society. <http://dx.doi.org/10.1098/rspb.2012.0205>

Sahling, H., Masson, D., Ranero, C., Hühnerbach, V., Weinrebe, W., Klauke, I., Bürk, D., Brückmann, W., Suess, E. (2008). "Fluid seepage at the continental margin offshore Costa Rica and southern Nicaragua" [Article-journal]. In American Geophysical Union (AGU). <http://dx.doi.org/10.1029/2008GC001978>

Datasets

Levin, L., Rouse, G. (2020). "Averages and standard deviation across species for all macrofauna found on each carbonate rock collected during R/V Atlantis cruise AT37-13 in the Pacific margin of Costa Rica from May to June 2017" [Dataset]. In Biological and Chemical Oceanography Data Management Office. <https://doi.org/10.1575/1912/bco-dmo.747575.1>

Levin, L., Rouse, G. (2020). "Locations of hard substrata collected during R/V Atlantis cruise AT37-13 in the Pacific Ocean off Costa Rica from May to June 2017" [Dataset]. In Biological and Chemical Oceanography Data Management Office. <https://doi.org/10.1575/1912/bco-dmo.750308.1>

Levin, L., Rouse, G. (2020). "Matrix of taxon (columns) by sample (rows) for sediment push cores collected by HOV Alvin during R/V Atlantis cruise AT37-13 in the Pacific margin of Costa Rica from May to June 2017" [Dataset]. In Biological and Chemical Oceanography Data Management Office.

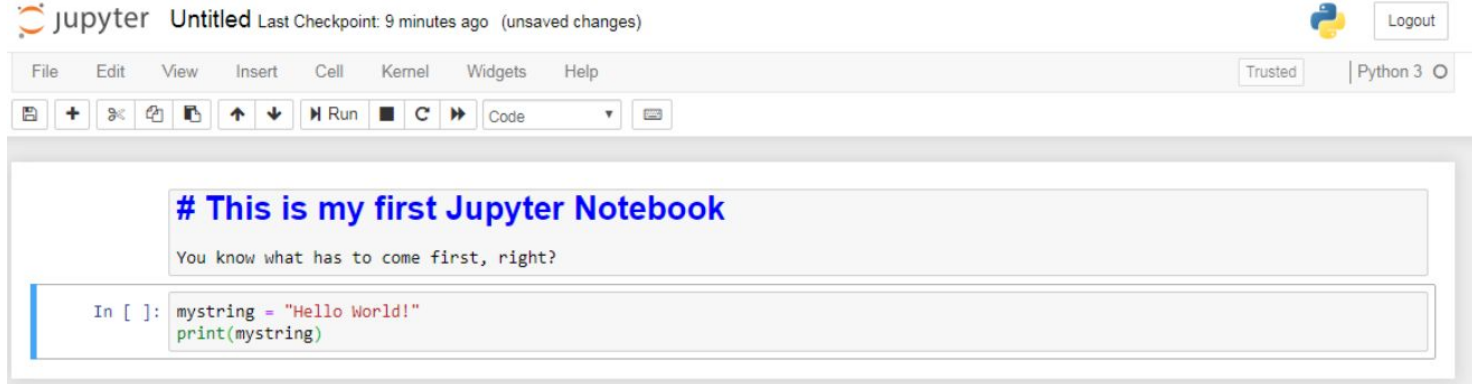
Demo time!

Visit <https://rb.gy/8o3hba>



Demos in Jupyter Notebooks

- Notebooks are documents that contain both code and rich text elements and different ways of visualising data via graphs, tables and figures.
- Jupyter Notebooks great at performing rapid visualisations that you can test out, change and share easily.



Demo 1 - Machine Actionable Data Management Plan connections

Choose DOI:

Fetch

We obtain

...

...

Data T

Simple tra

graph

- ✓ <https://doi.org/10.48321/D17G67>
- <https://doi.org/10.48321/D1H59R>
- <https://doi.org/10.1575/1912/bco-dmo.775500.1>
- <https://doi.org/10.48321/D1G59F>
- <https://doi.org/10.48321/D14S38>
- <https://doi.org/10.48321/D1101N>
- <https://doi.org/10.48321/D1W88T>
- <https://doi.org/10.48321/D1RG6W>
- <https://doi.org/10.48321/D1MS3M>
- <https://doi.org/10.48321/D1H010>
- <https://doi.org/10.48321/D1C885>
- <https://doi.org/10.48321/D17G67>
- <https://doi.org/10.48321/D13S3Z>
- <https://doi.org/10.48321/D1001B>
- <https://doi.org/10.48321/D1V88H>
- <https://doi.org/10.48321/D1QG6K>
- <https://doi.org/10.48321/D1KS39>
- <https://doi.org/10.48321/D1G01P>

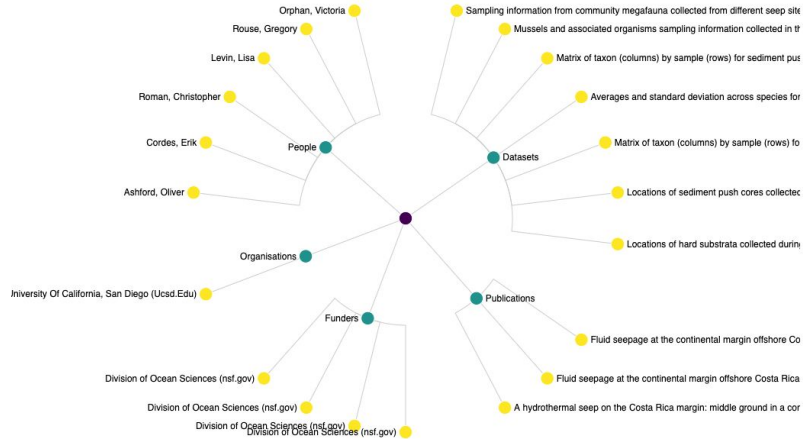


Chart Collaborative research: Quantifying the biological, chemical, and physical linkages between chemosynthetic communitie

labels

radius

extent

rotate

layout tidy cluster

links orthogonal

A series of sliders and option are included to interact with the visualisation is displayed. One can remove the labels, rotate the nodes, zoom in/out, and adjust the layout.



Demo 2 - Organisation/Funder/Repository Data Management Plans statistics

display(f)

European Commission - ror.org/00k4n6c32

Choose Or...

California Digital Library - ror.org/03yrm5c26

```
processTable("organization")
```

California Digital Library							
	DMP	Funder	Producer	NumDatasets	NumPublications	NumPeople	doi
0	DMPRoadmap: Making Data Management Plans Actionable	National Science Foundation (NSF)	University Of California System	0	0	4	https://doi.org/10.48321/d1mw28
1	LTREB: Drivers of temperate forest carbon storage from canopy closure through successional time	National Science Foundation (NSF)	University Of Michigan	1	3	5	https://doi.org/10.48321/d1h59r
2	Late Season Productivity, Carbon, and Nutrient Dynamics in a Changing Arctic	National Science Foundation (NSF)	Oregon State University	0	0	5	https://doi.org/10.48321/d17p4
3	REU Site: A Multidisciplinary Research Experience in Engineered Bioactive Interfaces and Devices	National Science Foundation (NSF)	University Of Kentucky	0	0	4	https://doi.org/10.48321/d1cc7t
4	Brown carbon characterization	National Science Foundation (NSF)	College, Harvey Mudd	0	2	3	https://doi.org/10.48321/d13w2m
5	A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy	National Science Foundation (NSF)	Western Washington University	0	2	3	https://doi.org/10.48321/d10593
6	Finding Levers for Privacy and Security by Design in Mobile Development	National Science Foundation (NSF)	University Of Maryland, College Park	0	6	4	https://doi.org/10.48321/d1vc75
7	Use of telemetry and the Acoustic Wave Glider to study southern flounder migrations	National Science Foundation (NSF)	East Carolina University	0	0	6	https://doi.org/10.48321/d1kw2z
8	The Virgin Islands Partnership to Increase Participation and Engagement through Linked, Informal, Nurturing Experiences in STEM (V.I. PIPELINES)	National Science Foundation (NSF)	University Of The Virgin Islands	0	0	7	https://doi.org/10.48321/d1qp4w
9	DMP for The Role of Temperature in Regulating Herbivory and Algal Biomass in Upwelling Systems	National Science Foundation (NSF)	University Of North Carolina, Chapel Hill	0	13	3	https://doi.org/10.48321/d1g59f



What's next?

- Talk to us about integrating with the DataCite APIs or using DataCite Fabrica to register your own DMP IDs
- Share your experiences and provide feedback
- Look for upcoming webinars, best practice documentation and blog posts for more details

Real-life pilot project
that allow us to test our
assumptions



FAIR Island Project

Develop **optimal data policies** and **technical infrastructure necessary** to create an environment where all data and knowledge collected by field stations managed by the University of California is curated and made **openly available as quickly as possible.**



FAIR Island Project Partners

BERKELEY

Institute for
Data Science



RESEARCH DATA ALLIANCE



DMPTool

Build your Data Management Plan

roadmap



TETIAROA
SOCIETY

SUSTAINABILITY IN ACTION



Natural Reserve System

UNIVERSITY OF CALIFORNIA

Reserve Application Management System



Where to Learn More

- [Networked DMPs on the DMPTool Blog](#)
- [*Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support*](#)
- [May 2019 Dear Colleague Letter on Effective Practices for Data](#)
- [Machine-actionable data management plans \(maDMPs\) \(2017 Rio Journal\)](#)
- [Ten principles for machine-actionable data management plans](#) - a 2019 paper published in PLOS Computational Biology
- [FAIR Island Project](#)

