A Programmatic and Scalable Approach to Making Data Management Machine-Actionable

COLLECTION:
DATA MANAGEMENT PLANNING ACROSS DISCIPLINES AND INFRASTRUCTURES

PRACTICE PAPER

]u[ubiquity press

MARIA PRAETZELLIS
MATTHEW BUYS
XIAOLI CHEN
JOHN CHODACKI
NEIL DAVIES
KRISTIAN GARZA
CATHERINE NANCARROW
BRIAN RILEY
ERIN ROBINSON

*Author affiliations can be found in the back matter of this article

ABSTRACT

The challenge of tracking research productivity and impact is compounded by the fragmented ecosystem in which research data outputs are managed, with stakeholders such as funders, research centers, government agencies, and academic institutions struggling to ensure compliance with mandates for data sharing and other regulatory requirements. While data management plans (DMPs) are crucial for effective research data management (RDM), their inadequate integration exacerbates fragmentation. This paper presents a robust, scalable, persistent identifier (PID)–enabled infrastructure that interlinks metadata and updates DOI records, effectively incorporating DMPs to bolster data discoverability, reusability, and compliance.

CORRESPONDING AUTHOR:
Maria Praetzellis
California Digital Library, University of California Office of the President, US
maria.praetzellis@ucop.edu

KEYWORDS:
data management plan; persistent identifier; DOI; research data management; open infrastructure

TO CITE THIS ARTICLE:
Praetzellis, M, Buys, M, Chen, X, Chodacki, J, Davies, N, Garza, K, Nancarrow, C, Riley, B and Robinson, E. 2023. A Programmatic and Scalable Approach to Making Data Management Machine-Actionable. *Data Science Journal*, 22: 26, pp. 1–9. DOI: https://doi.org/10.5334/dsj-2023-026

The data management plan (DMP), while seen by many as an ancillary document during a grant application, is a rich source of contextual information that is key to ensuring researchers, funders, and institutions follow the best possible and most appropriate research data management (RDM) practices. Unfortunately, the current practice is to transmit this information to the funder as a PDF or Word file through their web portals. As optimizing internal workflows and information sharing is a priority across the research space, retooling DMPs as machine-readable and machine-actionable will enable leveraging key information to build RDM strategies collectively. Similarly, there is a growing need to streamline workflows, reuse information, and reduce the burden on researchers.

Established in 2011 by the California Digital Library (CDL), the DMPTool began as a collaborative effort among eight institutions to address the growing need for data management plans required by funding agencies. Now, 13 years later, it continues as a free open-source platform for creating DMPs and is used by over 380 institutions worldwide.

For the past several years, technical development for the DMPTool has focused on transforming the DMP from a static textual document into a machine-actionable source of information about key aspects of research to be exchanged between systems. To support this work, in 2017, CDL was awarded a National Science Foundation EAGER innovation grant to explore ways of mapping research project outputs as described in a DMP to the broader ecosystem (National Science Foundation 2017). This initial work led the DMPTool team to explore how to best capture information about associated outputs (preprints, datasets, protocols, instruments, samples, etc.) and map these resources to other related research outputs.

This paper presents a robust, scalable, persistent identifier (PID)-enabled infrastructure that interlinks metadata and updates DOI records to bolster data discoverability, reusability, and compliance with data sharing mandates. This infrastructure streamlines collaboration by addressing core research questions, fostering transparency, and championing open data practices, and it significantly alleviates the challenges of tracking and promoting research outputs.

To explore this approach, the paper first highlights the importance of tracking research outputs linked to DMPs and contextualizes this need within emerging policies governing federally funded research projects. Then, the paper presents several ongoing projects employing machine-actionable DMPs (maDMPs) as illustrative use cases, demonstrating how maDMPs can facilitate the creation of scalable, findable, accessible, interoperable, and reusable (FAIR) workflows.

BACKGROUND

Multiple communities seek to understand the impact of research data outputs, but despite numerous tools and services, organizations struggle to systematically track research productivity and ensure compliance with data sharing mandates (Anger et al. 2022). This stems from fragmented information collection and dissemination, challenging data interoperability and reusability. This difficulty mainly arises from an ecosystem in which information about research is collected and disseminated through separate platforms and scattered across multiple sources and services, rendering it inaccessible for interoperability or reusability. As a result, each group has a piece of the puzzle, but no one has the complete picture.

Investments have been made in knowledge graphs and databases (Priem, Piwowar & Orr 2022), yet a community-driven, open infrastructure to share and consume connections between scholarly works is lacking. A scalable PID-enabled infrastructure could efficiently track research by connecting metadata across sources, providing transparency and facilitating open data practices.

Recent years have seen growing recommendations and requirements for data management and sharing practices. In the United States, NSF's 2019 recommendations for PIDs and maDMPs (Dear colleague letter 2019) have been expanded upon by various institutions and translated into legal requirements by the CHIPS and Science Act (H.R.7178). The NIH Data Management & Sharing Policy also supports open data and open science practices (National Institutes of Health 2020). Expanding on the NSF recommendations, the Association of Research Libraries (ARL), the California Digital Library (CDL), the Association of American Universities (AAU), and the Association of Public and Land-grant Universities (APLU) collaborated on an NSF-funded report, Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support, which collected key recommendations for effective data practices to support a more open research ecosystem, and this report also focused on using PIDs and maDMPs (Chodacki et al. 2020).

With increasing mandates, the focus has shifted toward implementation and compliance. Building a robust research data infrastructure to address these needs requires leveraging existing community-driven open systems and workflows to track research outputs from planning to long-term preservation. Allowing organizations to augment existing metadata and update DOI records with relationships to other scholarly outputs will bolster the ability to effectively track research in an automated and efficient manner. Tracking research outputs via PID-enabled infrastructure can answer many fundamental questions to research discovery and reach. For example:

- What are the downstream effects of research?
- Was a dataset reused or cited in subsequent research?
- Was a protocol reused or modified to support additional investigations?
- What was the cumulative impact of research conducted at a specific lab or field station?
- What were the broader impacts of research on society?

A PID-enabled infrastructure answers such questions by harvesting and connecting the rich metadata currently found in many disparate sources throughout the scholarly publishing ecosystem. Trusted partners can assert new connections that the wider community can leverage. This approach aims to provide transparency in the research process and facilitate optimal open data practices by collecting, recording, and redistributing information about research outputs.

APPROACH

Developing the technical capacities underpinning this approach to tracking research outputs has been in the works for many years, thanks to the active development of the Research Data Alliance (RDA) community and many international collaborations. Two principal technologies underpin this proposed system: the maDMP and the Event Data system. The benefits of coupling these two technologies became apparent as the DMPTool team worked to develop the maDMP and partnered with DataCite on the Event Data service. Since DMPs can be structured as machine-readable documents, it is now possible to track project updates over time and to record these activities in the metadata of the DMP record. For example, a maDMP can track where and when related datasets are deposited or how these outputs are reused or cited. Other related outputs can also be connected to a project, for example, protocols, samples, and instruments.

The Event Data system is a central clearinghouse for trusted sources to assert new information about research outputs. The aim of Event Data is to have a community-led system to act as that central component. Over the past five years, CDL has partnered with Crossref and DataCite to prototype approaches that collect, record, and redistribute additional information about research outputs through user dashboards (Dasler & Cousijn 2018). The response has been positive, with the endorsement of maDMPs coming from institutions and the RDA and funding support from the National Science Foundation. However, there is a need to strengthen the technology and the community of practice.

As currently architected, the Event Data system can monitor various sources and record an 'event' when the service identifies a mention of a research article with a registered DOI. Event Data facilitates connections between two DOIs or a DOI and a URL. Some examples include connecting a published manuscript to a related dataset(s) or a published protocol with a grant ID or related manuscript. These connections are made through metadata using the <relatedIdentifier> property, sent back to DataCite, and updated in the DOI metadata.

Although basic infrastructure is in place, expanding the architecture to include additional sources and 'events' from trusted sources can fill gaps and create new connections for the wider community to leverage. For instance, when a proposal is awarded, how can the award be connected to the related DMP? Or, if a dataset is published, how can the data publication be connected to the grant that sponsored the work or to the samples used to generate the dataset? Only organizations that mint the DOI (e.g., the Dryad data repository) or 'trusted sources' defined in the Event Data service can assert these connections in the current Event Data workflow. While an important security measure, it limits trusted sources from contributing new connections or correcting existing ones. Expanding the system aims to address these limitations in collecting and exposing relationships between research outputs, enriching metadata within the DOI record, and allowing the wider community to benefit from the collective information. For example, if achieved, a funding agency could easily assert connections between funded projects and their research outputs, or a publisher could utilize Event Data to record relationships between articles and related dataset(s).

Only with the open exchange of metadata will it be possible to track grant-funded research outputs effectively. This approach to tracking research data relies on open citations, openly available metadata, and a more robust Event Data system that allows trusted sources to augment information about the scholarly record.

USE CASES

To help illustrate this approach, the following section details three use cases derived from CDL's collaborations over the past four years involving university administrators, field station directors, and research teams.

USE CASE: RESEARCH UNIVERSITIES

The upcoming NIH requirements for data management and sharing have added a sense of urgency for universities struggling to fully understand and quantify the research their faculty and staff undertake. Universities are ultimately responsible for ensuring that their research complies with these new policies; they face potential funding loss if they fail to do so. This reality has led to a significant uptick in organizations utilizing the DMPTool. The DMPTool team has participated in many conversations with research and grant offices that seek tools to provide comprehensive insight into research currently being conducted, their associated planned outputs, and eventually published resources. University-based administrators require a system that creates an audit trail demonstrating compliance with award terms, including open data mandates and other regulatory requirements that might apply to different studies. These users are also interested in ensuring that all relevant data outputs are correctly associated with their affiliation, thus enabling citation counts—a crucial aspect of building incentive-reward systems that, along with obligations, will underpin broad adoption.

USE CASE: FIELD STATIONS

The National Science Foundation–funded FAIR Island Project collected user stories on data issues from an advisory board of field station directors and staff (Robinson et al. under review). These stories illustrate an all too common disconnect between research outputs and the places from which they are derived: field observations and samples are only sometimes traceable to/from papers and datasets or back to the stewards of the places to which they pertain (Davies et al. 2021). The failure to maintain linkages between places—at scales from small islands to entire continents—and derived data and samples has important impacts on the capacity to distribute benefits of the research to society (e.g., to natural resource managers), to accelerate other research related to a place (e.g., through reuse of samples and place-based data), and to ensure recognition of the value added by field-based institutions and infrastructure (Thomer 2022).

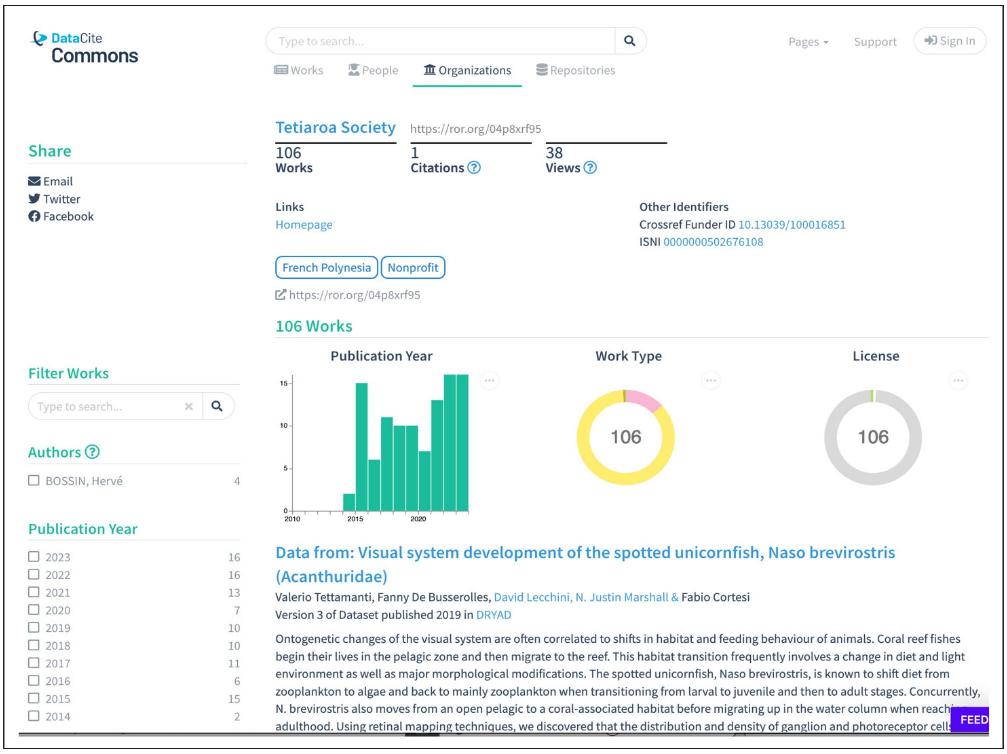
Questions field station directors and administrators would like to see automatically answered through research outcomes being linked back to 'places' (e.g., natural reserves, national parks, marine protected areas, Long-Term Ecological Research (LTER) sites, or international jurisdictions like Antarctica or the high seas) include the following:

- What was the impact of research conducted and supported by a station (or place) over time? For example, what was the number of publications, number of datasets, number of citations by year, and by discipline?
- What research outputs resulted from time spent on the station (e.g., data, samples, papers)?

- Where were these outputs published?
- Were they cited or reused in subsequent publications?
- Where was data gathered on the station or in nearby field sites?
- What types of research are being conducted?
- What permissions or authorizations were obtained, and what outputs resulted from them?

To answer these questions, and many others that emerge when these links are made, the FAIR Island Project has experimented with two working field stations on neighboring islands in French Polynesia: the Tetiaroa Society Ecostation (on Tetiaroa atoll) and the University of California Gump South Pacific Research Station (on Moorea). Based on this experience, the project developed a generic Place-Based Data Policy that requires teams accessing a site to use PIDs for organizations (RORs) and people (ORCIDs) and to have a DMP (Davies, Chodacki et al. 2021).

The team retrospectively created a DMP-ID (utilizing the DMPTool) for the Moorea Biocode project (Robinson et al., under review), which ended in 2012. Through the DMP-ID metadata for this project, the team added RORs, ORCIDs, and funder information. It updated the DMP metadata to include the existing research outputs to reconstruct how the project's research outputs grew over time. One novel use of ROR in this experiment was to include the field station ROR as a contributor-type 'sponsor' (DataCite Metadata Working Group 2021). Doing so enabled a key link between the project, the research outputs, and the station (see Figure 1).



Tetiaroa Society https://ror.org/04p8xrf95

106 Works | 1 Citations | 38 Views

Links
Homepage

Other Identifiers
Crossref Funder ID 10.13039/100016851
ISNI 0000000502676108

French Polynesia | Nonprofit

https://ror.org/04p8xrf95

106 Works

Authors

□ BOSSIN, Hervé	4

Publication Year

□ 2023	16
□ 2022	16
□ 2021	13
□ 2020	7
□ 2019	10
□ 2018	10
□ 2017	11
□ 2016	6
□ 2015	15
□ 2014	2

Data from: Visual system development of the spotted unicornfish, Naso brevirostris (Acanthuridae)

Valerio Tettamanti, Fanny De Busserolles, David Lecchini, N. Justin Marshall & Fabio Cortesi
Version 3 of Dataset published 2019 in DRYAD

Ontogenetic changes of the visual system are often correlated to shifts in habitat and feeding behaviour of animals. Coral reef fishes begin their lives in the pelagic zone and then migrate to the reef. This habitat transition frequently involves a change in diet and light environment as well as major morphological modifications. The spotted unicornfish, Naso brevirostris, is known to shift diet from zooplankton to algae and back to mainly zooplankton when transitioning from larval to juvenile and then to adult stages. Concurrently, N. brevirostris also moves from an open pelagic to a coral-associated habitat before migrating up in the water column when reach adulthood. Using retinal mapping techniques, we discovered that the distribution and density of ganglion and photoreceptor cells

Figure 1 DataCite Commons dashboard for the Tetiaroa Ecostation demonstrates an aggregated view of 106 research outputs connected to the Ecostation using ROR and DMP-IDs. https://commons.datacite.org/ror.org/04p8xrf95.

Through this experiment, the team showed initial answers to the questions above for a single project at the Tetiaroa Ecostation. By using DOIs, it is possible to share rich metadata via the existing DOI research infrastructures. In this project, the DMP-ID was a proxy identifier for the project. The team is now exploring adding a project ID and the relationship between project IDs and DMP-IDs. With that relationship established, research outputs will be connected through machine-actionable links.

USE CASE: RESEARCH TEAMS

Making research easy to find, access, integrate, and reuse (FAIR) requires collaboration among researchers, funders, administrators, tools, platforms, and the open scholarly infrastructure. DMPTool serves as a hub for these stakeholders, offering templates based on funder requirements to help researchers plan, document, and track their outputs. Adding user-generated metadata to the PID graph makes the DMP and related outputs more easily discoverable.

The FAIR Workflows project is developing an exemplar FAIR and open research workflow based on the complete research life cycle. With multiple partners, including DataCite, the Max Planck Institute for Empirical Aesthetics, ChronosHub, and the Australian Research Data Commons, CDL is working toward creating an end-to-end FAIR workflow for researchers that enables tracking of all different components of a research project through PIDs and their metadata. For this project, the DMP-ID has become the central identifier for the project, and all related works (datasets, protocols, manuscripts, etc.) are connected to the project through related identifiers and the Event Data service.

The FAIR Workflows project builds best practices for effective, automated, and open information exchange by facilitating metadata exchange and enrichment. When implemented, the workflows developed through the project realize the potential of PID infrastructure to allow stakeholders such as institutional repositories (IRs) and current research information systems (CRIS) to query, contribute, and augment publicly available metadata (Chen, Cousijn & Stathis 2022). This metadata increases the FAIRness of research outputs by expanding relationships to other related works and, importantly, provides a vehicle for the community to contribute insights regarding the usage and impact of research outputs (see Figure 2).

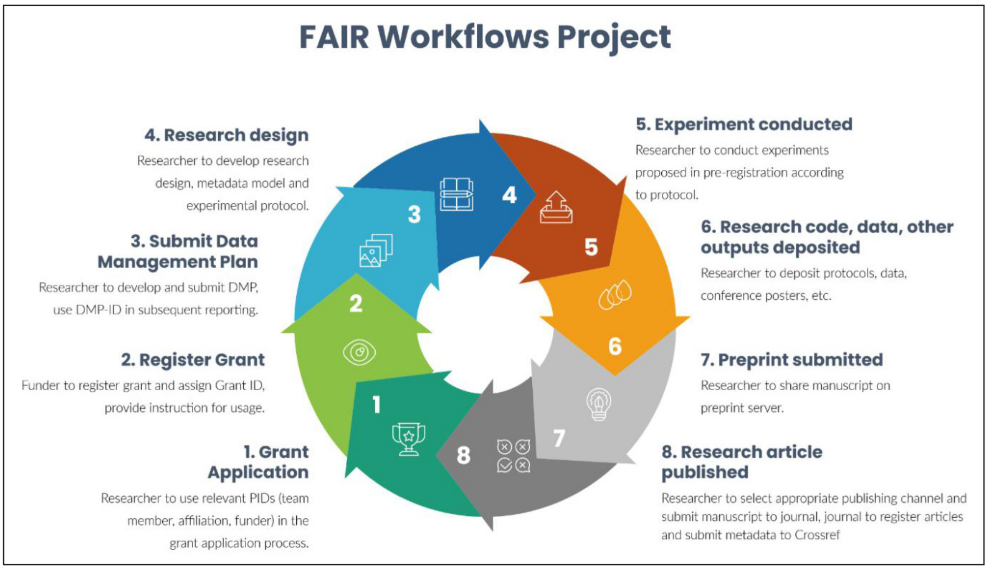

FAIR Workflows Project
4. Research design
Researcher to develop research design, metadata model and experimental protocol.
5. Experiment conducted
Researcher to conduct experiments proposed in pre-registration according to protocol.
3. Submit Data Management Plan
Researcher to develop and submit DMP, use DMP-ID in subsequent reporting.
6. Research code, data, other outputs deposited
Researcher to deposit protocols, data, conference posters, etc.
2. Register Grant
Funder to register grant and assign Grant ID, provide instruction for usage.
7. Preprint submitted
Researcher to share manuscript on preprint server.
1. Grant Application
Researcher to use relevant PIDs (team member, affiliation, funder) in the grant application process.
8. Research article published
Researcher to select appropriate publishing channel and submit manuscript to journal, journal to register articles and submit metadata to Crossref
1
2
3
4
5
6
7
8

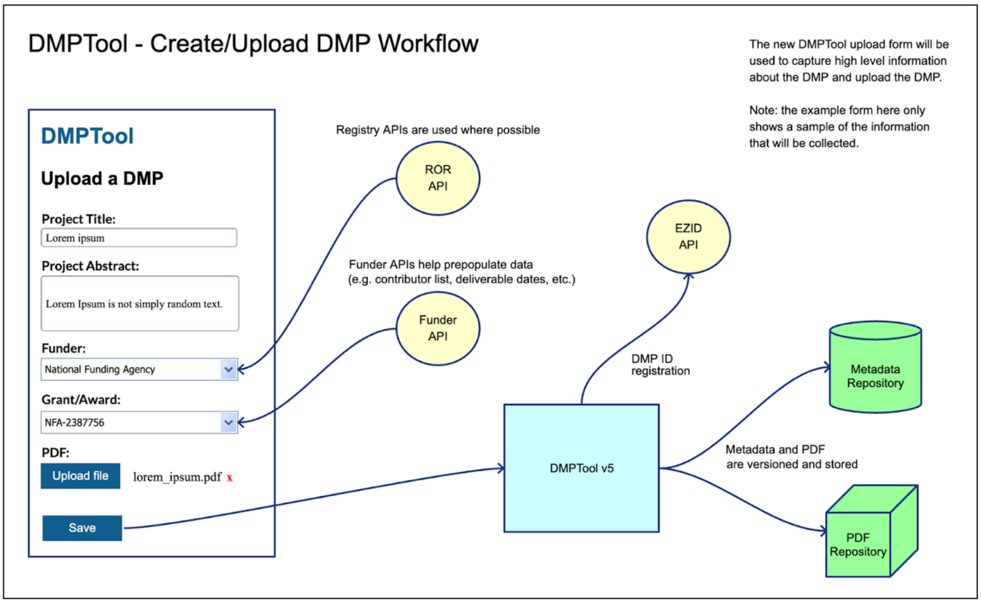
Figure 2 The FAIR Workflows implementation cycle illustrates the specific points along the research life cycle where the project builds PID-enabled system integrations.

OUTLOOK

The DMPTool team continues to expand this new, interconnected, and interoperable approach within the application. The team is currently building a new workflow to ingest DMPs created outside of the DMPTool to expose this information in a machine-readable format and generate a PID for the DMP (the DMP-ID) (see Figure 3). Connections to funder APIs automatically transfer relevant metadata to avoid duplication of researcher effort and augment information found in a DMP (see Figure 4). With these DMP-IDs, information connected to the project can be shared with external sources such as repositories, publishers, funders, and university administrators, facilitating automated actions such as notifications and compliance reporting.

Additional information found in external funder APIs and other research systems is added to the DMP-ID record throughout a project. For example, integrating the NSF and NIH awards APIs links funding and project information. Facilitating these updates allows for tracking of research projects as they progress through the research life cycle. It also provides administrators, librarians, and other stakeholders insight into research currently being conducted and the full spectrum of associated outputs, deliverables, and forthcoming publications.

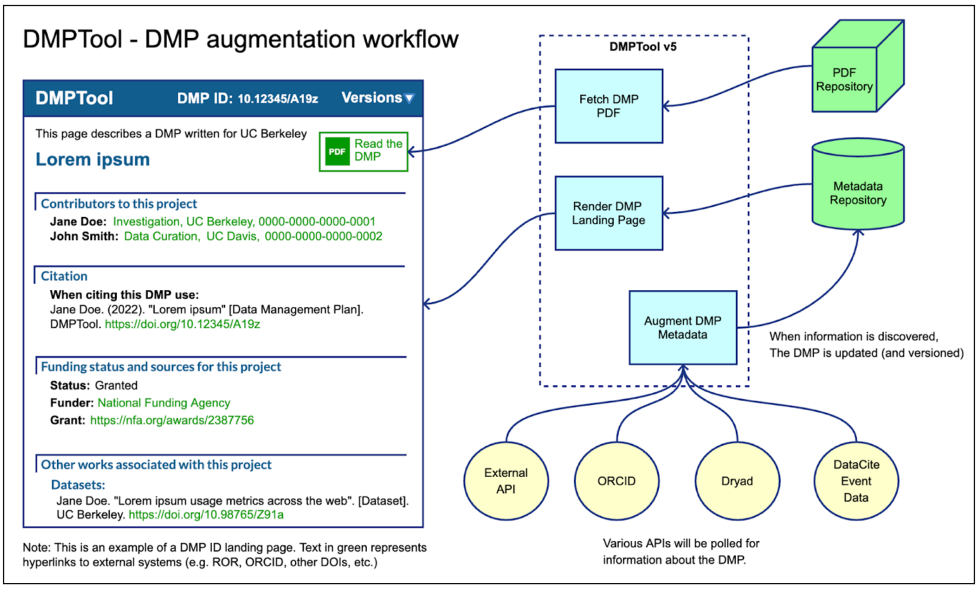
Event Data serves as a core piece of this workflow. All related outputs associated with a project are passed back to DataCite and recorded in the metadata of the DMP-ID as a related identifier. By submitting updated metadata to DataCite, this workflow facilitates tracking scholarly outputs, bolsters the larger scholarly record, and is openly available for consumption and use by other systems. This coupling of maDMPs with the PID-enabled infrastructure of the Event Data service enables the effective tracking of projects over time. Furthermore, it allows these relationships to be shared with the larger research data ecosystem in an open and interoperable manner.



The new DMPTool upload form will be used to capture high level information about the DMP and upload the DMP.

Note: the example form here only shows a sample of the information that will be collected.

Figure 3 DMPTool architecture for structuring and generating DMP-IDs for externally created plans. This approach enhances the scalability of PID generation for DMPs by simplifying DMP-ID creation and offering PID-enabled metadata for research projects.



This page describes a DMP written for UC Berkeley

Lorem ipsum

Contributors to this project

Jane Doe: Investigation, UC Berkeley, 0000-0000-0000-0001
John Smith: Data Curation, UC Davis, 0000-0000-0000-0002

Citation

When citing this DMP use:
Jane Doe. (2022). "Lorem ipsum" [Data Management Plan]. DMPTool. https://doi.org/10.12345/A19z

Funding status and sources for this project

Status: Granted
Funder: National Funding Agency
Grant: https://nfa.org/awards/2387756

Other works associated with this project

Datasets:
Jane Doe. "Lorem ipsum usage metrics across the web". [Dataset]. UC Berkeley. https://doi.org/10.98765/Z91a

When information is discovered, The DMP is updated (and versioned)

Note: This is an example of a DMP ID landing page. Text in green represents hyperlinks to external systems (e.g. ROR, ORCID, other DOIs, etc.)

Various APIs will be polled for information about the DMP.

Figure 4 Workflow illustrating the DMPTool integration of external funder APIs and research systems to enrich DMP-ID records throughout a project's life cycle. For example, it links funding and project information from NSF and NIH awards APIs, enabling tracking of research progress and providing stakeholders with insights into ongoing research, associated outputs, deliverables, and future publications.

With the impending NIH requirements for data sharing and the newly passed CHIPS Act, it is an opportune moment to build on the existing open infrastructure for a sustainable suite of tools that support open data practices. The scholarly community has spent years advocating for effective open data policies and invested in building tools and infrastructure to facilitate open science practices. Now that significant open data policies are in place, there is a need to ensure that the systems developed in response to these policies are based on the same principles of openness.

FUNDING INFORMATION

The work presented in this paper was supported by the National Science Foundation (Davies: Internet of Samples: Toward an Interdisciplinary Cyberinfrastructure for Material Samples NSF#2004642; RCN: Sampling Nature: A Network to Enhance the Natural History Value Chain for Sustainability Science NSF#2129268, Chodacki: EAGER: The FAIR Island Project for Place-based Open Science NSF #2132549; Waibel: EAGER: DMP Roadmap: Making Data Management Plans Actionable #1745675) and the Templeton World Charity Foundation (Cousijn: Implementing FAIR Workflows: A Proof of Concept Study in the Field of Consciousness #0568).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation; and the Templeton World Charity Foundation.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Maria Praetzellis orcid.org/0000-0001-5047-3090
California Digital Library, University of California Office of the President, US

Matthew Buys orcid.org/0000-0001-7234-3684
DataCite, NL

Xiaoli Chen orcid.org/0000-0003-0207-2705
DataCite, DE

John Chodacki orcid.org/0000-0002-7378-2408
California Digital Library, University of California Office of the President, US

Neil Davies orcid.org/0000-0001-8085-5014
Gump South Pacific Research Station, University of California Berkeley, Moorea, French Polynesia

Kristian Garza orcid.org/0000-0003-3484-6875
DataCite, DE

Catherine Nancarrow orcid.org/0000-0001-8659-3115
California Digital Library, University of California Office of the President, US

Brian Riley orcid.org/0000-0001-9870-5882
California Digital Library, University of California Office of the President, US

Erin Robinson orcid.org/0000-0001-9998-0114
Metadata Game Changers, US

REFERENCES

Anger, M, Wendelborn, C, Winkler, EC and **Schickhardt, C.** 2022. Neither carrots nor sticks? Challenges surrounding data sharing from the perspective of research funding agencies—a qualitative expert interview study. *PLOS ONE*, 17(9): e0273259. DOI: https://doi.org/10.1371/journal.pone.0273259

Chen, X, Cousijn, H and **Stathis, K.** 2022. Implementing FAIR workflows D1.1 workflows specification. DOI: https://doi.org/10.5281/ZENODO.7382642

CHIPS and Science Act of 2022, H.R.7178. Available at https://www.commerce.senate.gov/2022/8/view-the-chips-legislation [Last accessed 12 December 2022].

Chodacki, J, Hudson-Vitale, C, Meyers, N, Muilenburg, J, Praetzellis, M, Redd, K, Ruttenberg, J, Steen, K, Cutcher-Gershenfeld, J and **Gould, M.** 2020. *Implementing effective data practices: Stakeholder recommendations for collaborative research support*. Washington, DC: Association of Research Libraries. DOI: https://doi.org/10.29242/report.effectivedatapractices2020

Dasler, R and **Cousijn, H.** 2018. Are your data being used? Event Data has the answer! *DataCite Blog*, 8 October 2018. DOI: https://doi.org/10.5438/s6d3-k860 [Last accessed 12 December 2022]).

DataCite Metadata Working Group. 2021. DataCite metadata schema documentation for the publication and citation of research data and other research outputs v4.4 [Application/pdf]. DOI: https://doi.org/10.14454/3W3Z-SA82

Davies, N, Chodacki, J, Praetzellis, M, Nancarrow, C and **Robinson, E.** 2021. Generic place-based research data policy. DOI: https://doi.org/10.5281/zenodo.5781442

Davies, N, Deck, J, Kansa, EC, Kansa, SW, Kunze, J, Meyer, C, Orrell, T, Ramdeen, S, Snyder, R, Vieglais, D, Walls, RL and **Lehnert, K.** 2021. Internet of samples (iSamples): Toward an interdisciplinary cyberinfrastructure for material samples. *GigaScience*, 10(5). DOI: https://doi.org/10.1093/gigascience/giab028

Dear colleague letter: Effective practices for data. 2019. National Science Foundation. Available at https://www.nsf.gov/pubs/2019/nsf19069/nsf19069.jsp [Last accessed 15 November 2022].

National Institutes of Health. 2020. Final NIH policy for data management and sharing. Available at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [Last accessed 12 December 2022].

National Science Foundation. 2017. EAGER: DMP roadmap: Making data management plans actionable. Award number 1745675. Available at https://www.nsf.gov/awardsearch/showAward?AWD_ID=1745675 [Last accessed 1 November 2022].

Priem, J, Piwowar, H and **Orr, R.** 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv:2205.01833 [Cs]. Available at http://arxiv.org/abs/2205.01833 [Last accessed 15 December].

Robinson, E, Chodacki, J, Praetzellis, M, Nancarrow, C and **Davies, N.** under review Generic place-based research data policy. DOI: https://doi.org/10.5281/zenodo.5781442

Thomer, AK. 2022. Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits. *Journal of the Association for Information Science and Technology*, 73(8): 1155–1170. DOI: https://doi.org/10.1002/asi.24620

TO CITE THIS ARTICLE:
Praetzellis, M, Buys, M, Chen, X, Chodacki, J, Davies, N, Garza, K, Nancarrow, C, Riley, B and Robinson, E. 2023. A Programmatic and Scalable Approach to Making Data Management Machine-Actionable. *Data Science Journal*, 22: 26, pp. 1–9. DOI: https://doi.org/10.5334/dsj-2023-026

Submitted: 16 December 2022
Accepted: 25 May 2023
Published: 25 August 2023