Harnessing the Power of machine actionable DMSPs: Where Are We Now and Building the Future

University of California
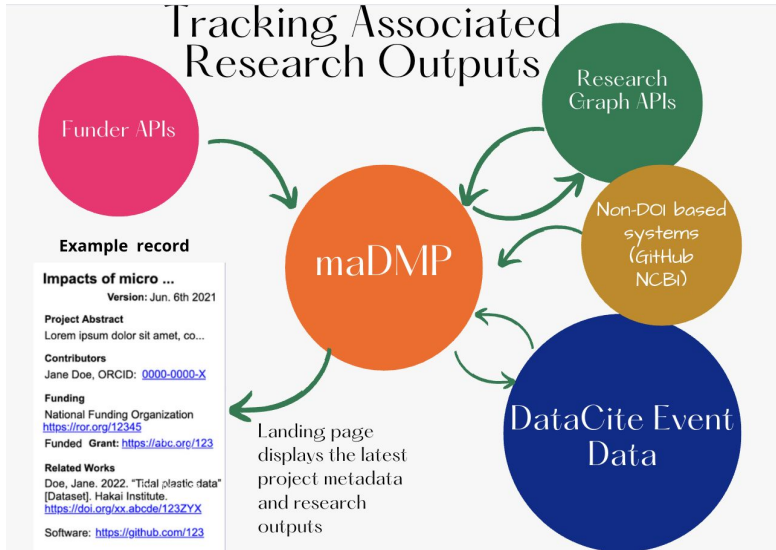**CDL**
California Digital Library

UC3

Project Overview

SMART-DMSP: Streamlining Metadata, Automation, and Research Tracking for Data Management and Sharing Plans
- Pilot and experiment
- Demonstrate value to our communities
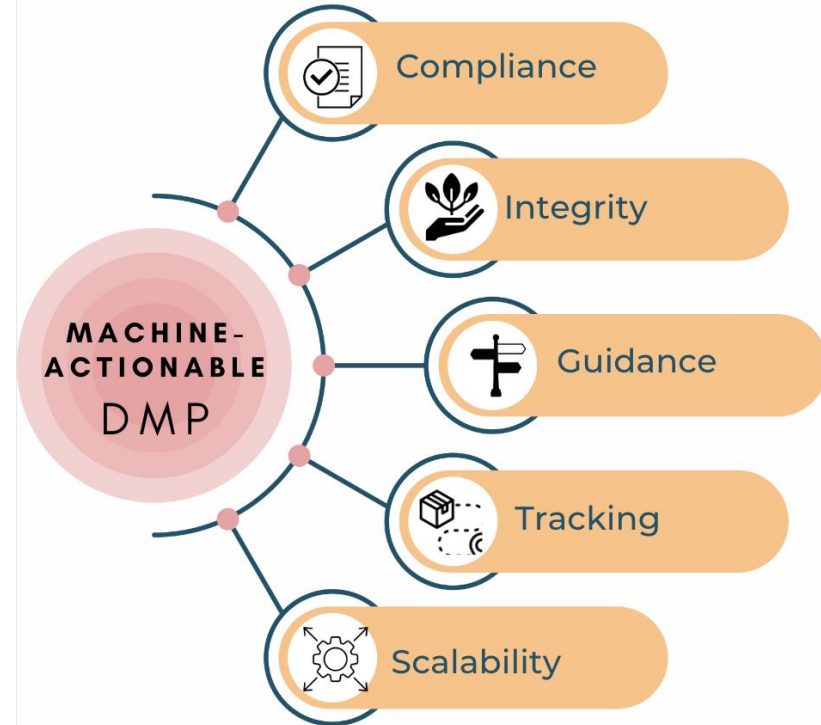- PI Maria Praetzellis - UC3 Associate Director

What is a maDMSP (or maDMP)?

Tracking Associated Research Outputs

Funder APIs

Research Graph APIs

maDMP

Non-DOI based systems (GitHub NCBI)

DataCite Event Data

Example record

Impacts of micro ...
Version: Jun. 6th 2021

Project Abstract
Lorem ipsum dolor sit amet, co...

Contributors
Jane Doe, ORCID: 0000-0000-X

Funding
National Funding Organization
https://ror.org/12345
Funded Grant: https://abc.org/123

Related Works
Doe, Jane. 2022. "Tidal plastic data" [Dataset]. Hakai Institute.
https://doi.org/xx.abcde/123ZYX

Software: https://github.com/123

Landing page displays the latest project metadata and research outputs

- **Structured**, machine-readable versions of traditional narrative based DMSPs.

- Enables **automation & integration** with administrative & research workflows.

- Living, versioned documents that are **updatable** over time.

Why are we talking about maDMSPs now?

- With increasing mandates conversations are now turning toward implementation and compliance.

- How to demonstrate and track institutional compliance with these mandates?

- In response we are building robust research data infrastructure using open, community-driven systems and workflows for tracking research outputs.

University of California
CDL
California Digital Library

MACHINE-ACTIONABLE
DMP

Compliance
Integrity
Guidance
Tracking
Scalability

What can you do with a maDMSP?

The goal of a machine actionable DMSP is to support the creation and stewardship of FAIR data

- Allow data and information about research to be communicated and shared across stakeholders

- Facilitating
  - easy retrieval of info about research
  - notifications and verification
  - automated compliance checks

- maDMSPs should lessen the administrative burden on researchers and grant administrators

Funder   Ethics review   Legal expert   Researcher   Publisher
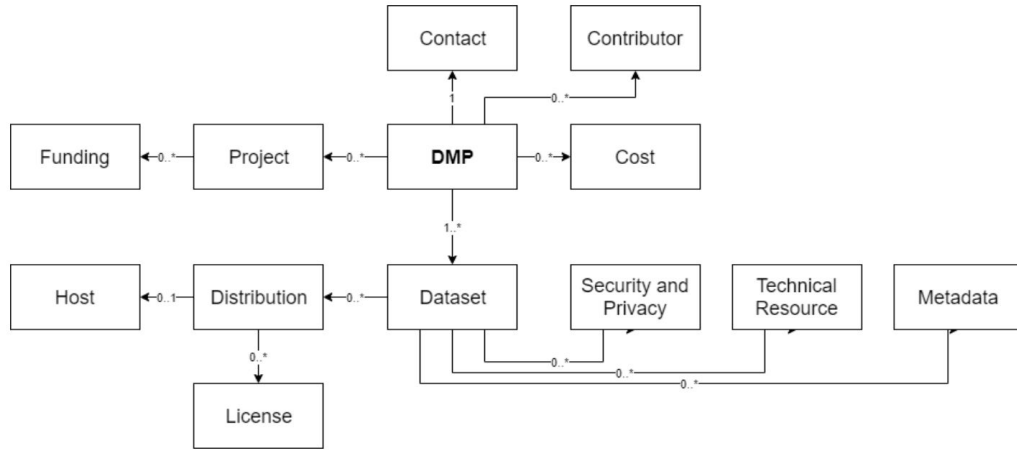
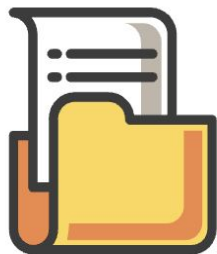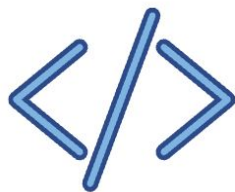Repository operator   Infrastructure provider   Research support staff   Institutional administrator

University of California
CDL
California Digital Library

UC3

Developing the maDMSP

Tomasz Miksa, Paul Walk, Peter Neish. RDA DMP Common Standard for Machine-actionable Data Management Plans. http://doi.org/10.15497/rda00039

What's in a maDMSP?

Data type

Tools, software, code

Standards

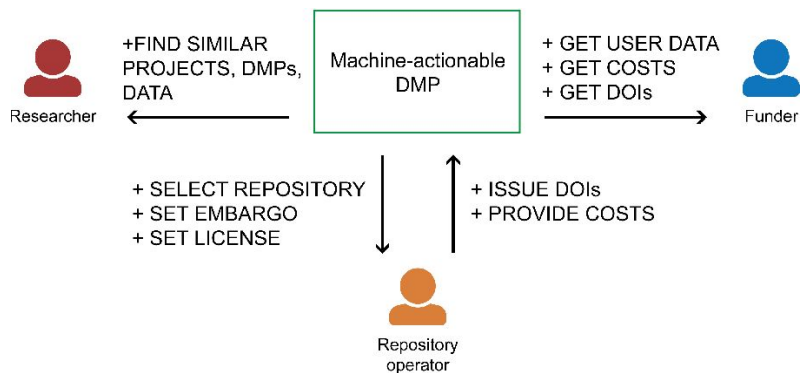Preservation and Access

Oversight

University of California
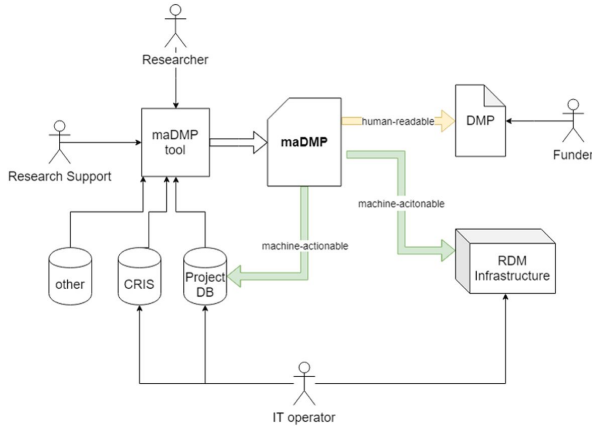CDL
California Digital Library

☀UC3

What can you do with a maDMSP?

- **Facilitate guidance** – maDMSPs help librarians and administrators provide guidance at scale and with fewer resources
- **Ensure compliance** – DMSPs are a requirement of the NIH Data Management & Sharing Policy, and maDMSPs are specified in the CHIPS and Science Act
- **Promote research integrity** – maDMSPs promote best practices and support research reproducibility and data security
- **Track impact** – maDMSPs help grant administrators track the impact of an organization's research programs

What integrations could you do with a maDMSP?

Miksa, T, et al.. "Application Profile for Machine-Actionable Data Management Plans". Data Science Journal, https://doi.org/10.5334/dsj-2021-032

- Automate notifications to stakeholders about key data lifecycle events using maDMSPs. Example use cases: sensitive data alerts, managing big data, enabling data transfers, linking planned datasets to outputs
- Build prototype integrations connecting maDMSPs with existing research systems like RIMS or researcher profiles to exchange info
- Engage academic departments to test maDMSPs for research workflows and data management needs, providing user feedback to optimize the systems
- Demonstrate improved communication workflows between campus RDM stakeholders (library, research office, IT/security, IRB, research/HPC) using maDMPSPs as a connecting platform

University of California
CDL
California Digital Library

UC3

Example of a maDMSP landing page

DMPTool   **DMP ID:** 10.48321/D17598   **Version:** 10 May 2023

This page describes a data management plan written for the National Science Foundation (nsf.gov) using the DMPTool.

EAGER: The FAIR Island Project for Open Science

Read the data management plan

Contributors to this project

Brian Riley: Data-curation, University of California, Office of the President (UCOP),
https://orcid.org/0000-0001-9870-5882

Catherine Nancarrow: Project-administration, University of California, Office of the President (UCOP),
https://orcid.org/0000-0001-8659-3115

Erin Robinson: Investigation, Metadata Game Changers (United States) (metadatagamechangers.com),
https://orcid.org/0000-0001-9998-0114

John Chodacki: Investigation, University of California, Office of the President (UCOP),
https://orcid.org/0000-0002-7378-2408

Maria Praetzellis: Data-curation, National Institutes of Health (NIH), https://orcid.org/0000-0001-5047-3090

Neil Davies: Investigation, University of California, Berkeley (UCB), https://orcid.org/0000-0001-8085-5014

Project details

Project Start: July 31, 2021

Project End: September 01, 2023

Created: May 13, 2021

Modified: May 10, 2023

Ethical issues related to data that this DMP describes? no

University of California
CDL
California Digital Library

UC3

API details

Our project looks to experiment with integrations and extensions from the DMPTool system - driven by our robust API.

- integrate organizational maDMSP data into local systems
- supports maDMSP retrieval, creation and modification

The DMPTool will monitor external systems and funder APIs for research output and award information regardless of whether the maDMSP was created within the DMPTool UI or directly through the API.

Adding additional API functionality throughout project.

See the API documentation for further details

Project development

Piloting additional maDMP workflows

Piloting the integration or creation of prototypes and possible workflows for maDMPs.

Connecting with additional open systems for tracking outputs

We'll be developing additional integrations with funder, APIs as well as larger indexes and aggregators such as Open Alex.

Developing a new data model for the DMPTool

This new data model will generate maDMPs more effectively and facilitate additional structured metadata.

Experimenting with the use of AI/ML

Investigating entity extraction to structure narrative DMPs, as well as AI to facilitate the generation of plans reflecting best practices.

INSTITUTE *of* **Museum** and **Library** SERVICES

NSF National Science Foundation

University of California
CDL
California Digital Library