

# Implementing Effective Data Practices

May 6, 2020

CNI Virtual Membership Meeting

Maria Gould \* Natalie Meyers \* Maria Praetzellis  
Jenny Mulenburg \* Judy Ruttenberg

ASSOCIATION  
OF RESEARCH  
LIBRARIES

ASSOCIATION  
OF RESEARCH  
LIBRARIES



University of California

**CDL**

California Digital Library

# Introduction

NSF 19-069

## Dear Colleague Letter: Effective Practices for Data

---

May 20, 2019

“The purpose of this Dear Colleague Letter (DCL) is to describe — and encourage — effective practices for managing *research data*, including the use of persistent identifiers (IDs) for data and machine-readable data management plans (DMPs).”

# Partners



- Advance open science by design
- Accelerating Public Access to Research Data
- Tool-builders and experts

NSF-funded conference in Washington, DC.  
December 11-12, 2019



**Grant No.  
1945938**

# Organizing Committee

**John Chodacki**, California Digital Library

**Cynthia Hudson-Vitale**, Penn State University Libraries

**Natalie Meyers**, University of Notre Dame Libraries

**Jenny Muilenburg**, University of Washington Libraries

**Maria Praetzellis**, California Digital Library

**Kacy Redd**, Association of Public and Land-grant Universities

**Judy Ruttenberg**, Association of Research Libraries

**Katie Steen**, Association of American Universities

Facilitator: **Joel Cutcher-Gershenfeld**

# **Participants & expert interviews**

# Value in data practices

## Persistent IDs

Discovery

Disambiguation

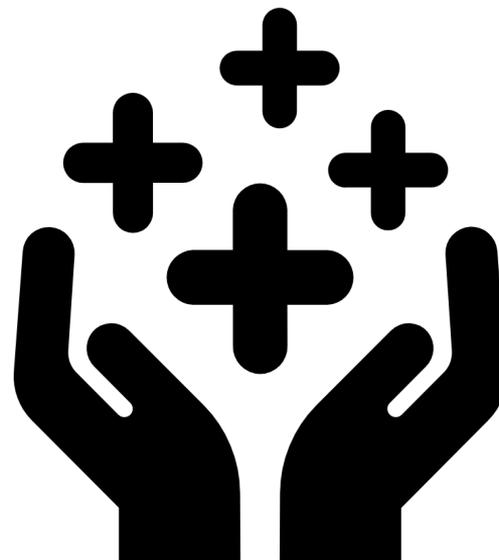
Credit

Tracking, linking, connecting

Automating compliance

Reproducibility

Metascience



# Value in data practices

## Machine-readable DMPs

Communication & progress reporting

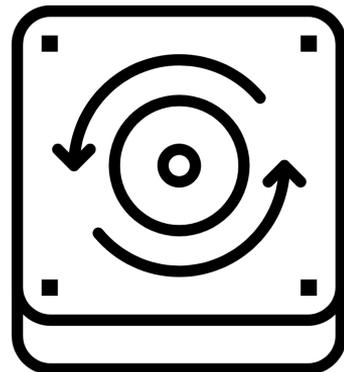
Repository planning

Campus planning

Syncing with pub process

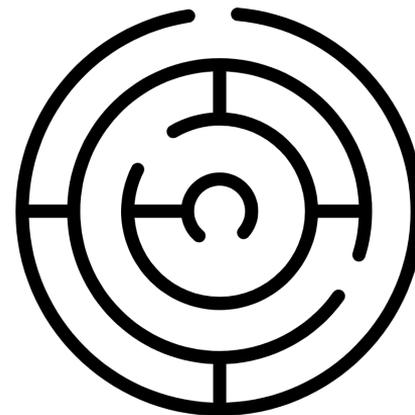
Risk identification

Transparency & accountability



# Persistent IDs: challenges

- Lack of awareness among investigators
- Confusion around granularity
- No communication between data repositories and publishers
- Concern over sustainability of PID organizations and infrastructure

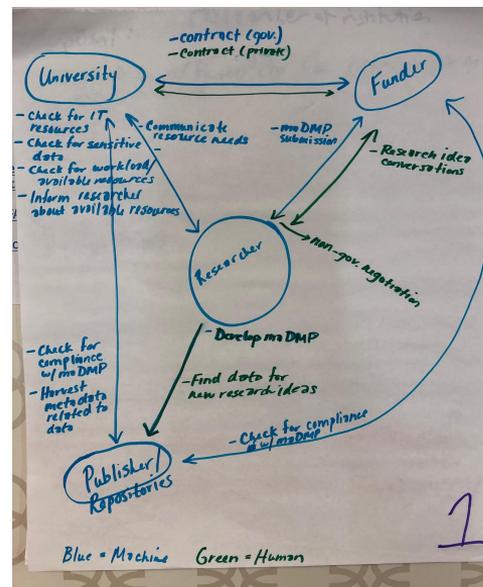


# DMPs: challenges

- Lack of awareness of tools
- Investigator burden
- DMPs are part of grant application  
(PDF), software (old), workflow (local and idiosyncratic)
- DMPs are neither public, nor shared

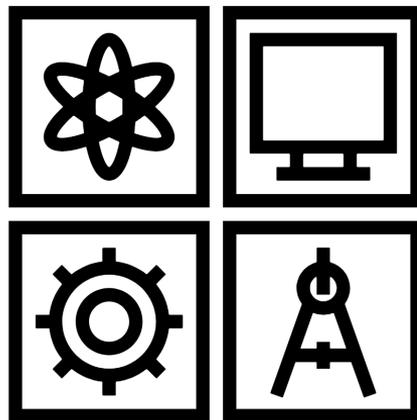
# Five takeaways

1. Center the researcher
2. Integrate library & scientific communities
3. Sustain PID infrastructure, including community and organizations
4. Unbundle the DMP
5. PIDs will unlock discovery



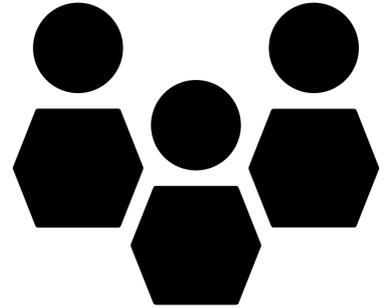
# Five takeaways

1. Center the researcher
2. **Integrate library & scientific communities**
3. Sustain PID infrastructure, including community and organizations
4. Unbundle the DMP
5. PIDs will unlock discovery



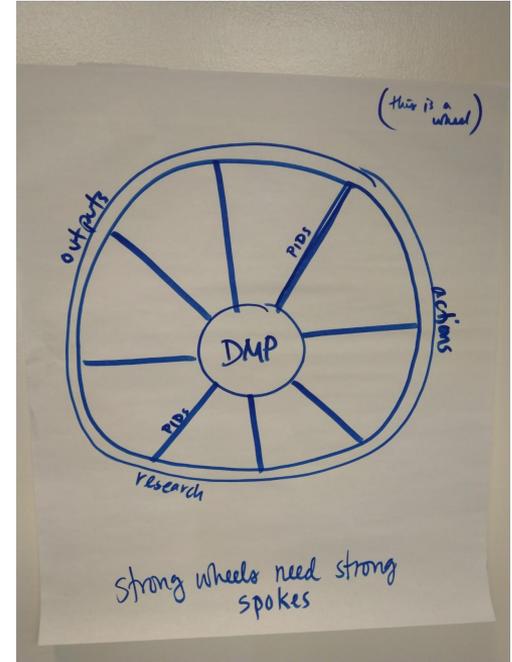
# Five takeaways

1. Center the researcher
2. Integrate library & scientific communities
3. **Sustain PID infrastructure, including community and organizations**
4. Unbundle the DMP
5. PIDs will unlock discovery



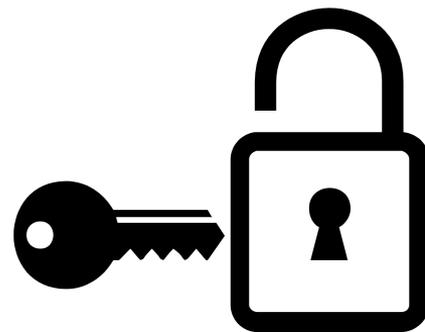
# Five takeaways

1. Center the researcher
2. Integrate library & scientific communities
3. Sustain PID infrastructure, including community and organizations
4. **Unbundle the DMP**
5. PIDs will unlock discovery



# Five takeaways

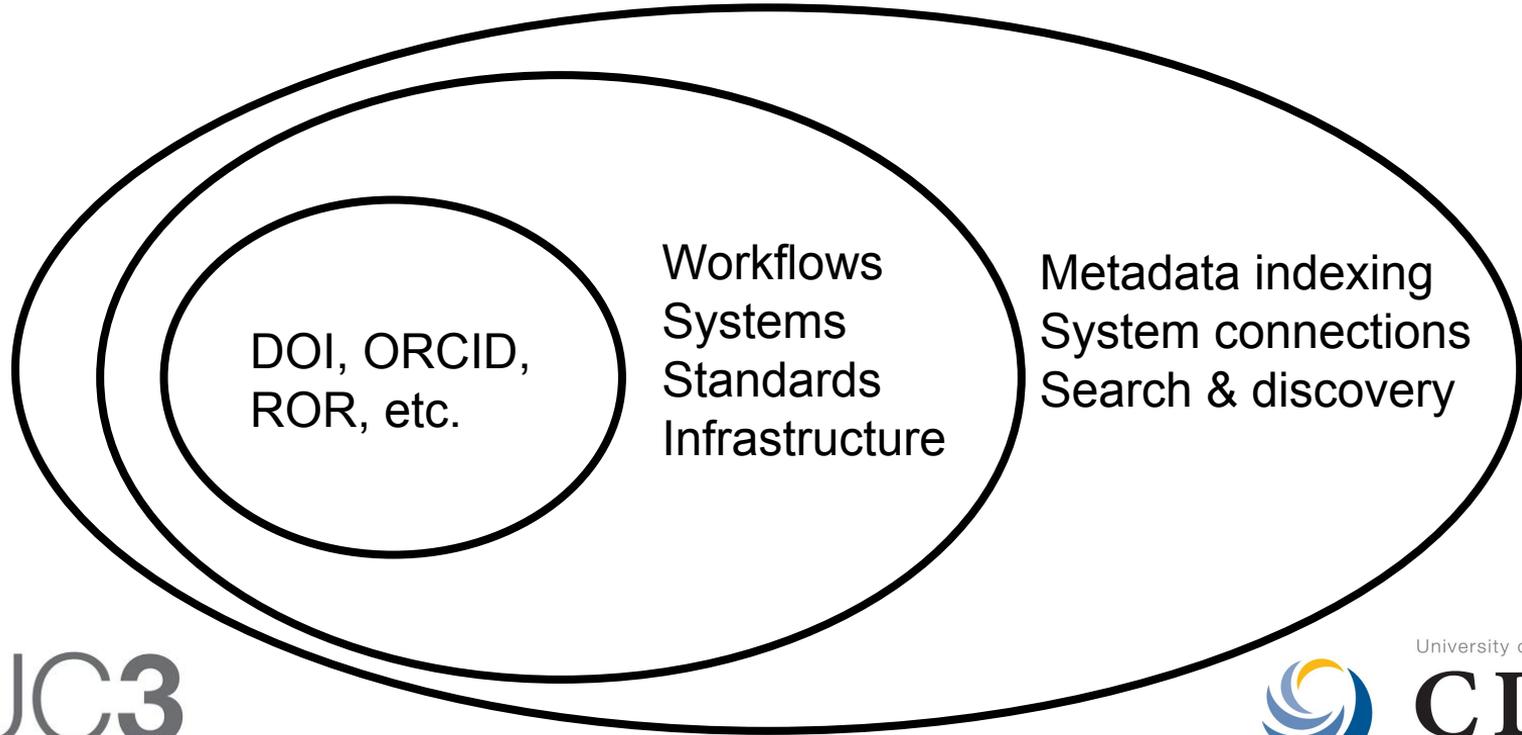
1. Center the researcher
2. Integrate library & scientific communities
3. Sustain PID infrastructure, including community and organizations
4. Unbundle the DMP
5. **PIDs will unlock discovery**



# Persistent identifier landscape(s)



# PIDs: Networking the research landscape to unlock discovery



# DMPTool

Platform for DMP creation and guidance with the goal of creating a bridge for researchers to discuss research data with librarians

46k+ users with 43k+ data management plans at 268 participating institutions

Create next-generation, machine-actionable DMPs



# PIDs for DMPs

Partnering with DataCite to mint PIDs for DMPs and utilizing this PID as an anchor for collecting changes over time. Evaluating opportunities for further incorporation into the PID graph, ie grant ids,

Using the PID infrastructure and DataCite's Event Data service to keep all players up to date on changes over time.



# The FAIR Island Project

Develop the **optimal data policies** and **technical infrastructure necessary** to create an environment where all data and knowledge collected on field stations are curated and made **openly available as quickly as possible**.



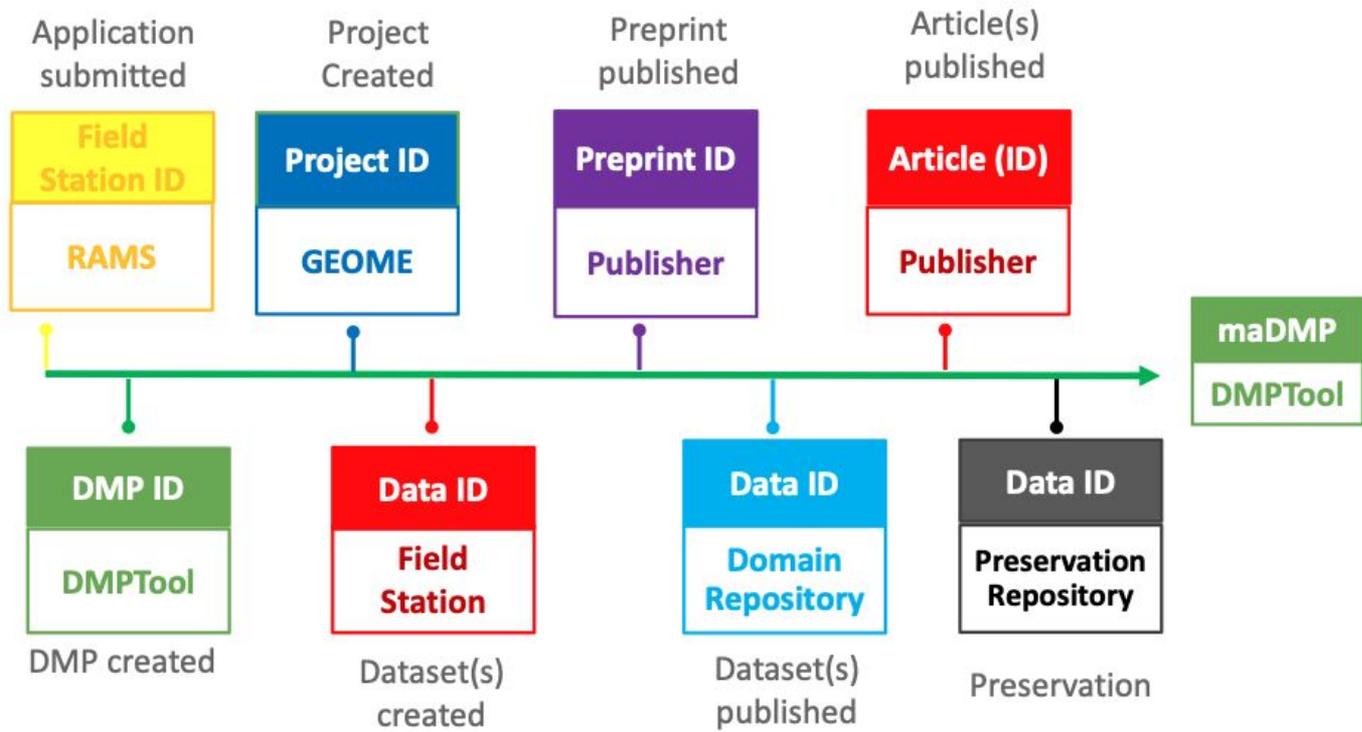
CDL  
California Digital Library

# Project Goals

Testing the effectiveness of optimized RDM/Open Science Policies.

Demonstrate the capabilities of machine-actionable data management plans (maDMPs) in a bounded, incubated, monitored working research environment and analyse the downstream effects of these policies in the resulting release of data.





# Ben Pierson, Bill and Melinda Gates Foundation

***“This is what we can do right now*** in our guidance, in our policies, to make it better for program officers and grantees getting research grants to make sure that their outputs are in fact persistent and machine readable.”

**PRECONFERENCE INTERVIEW NOV 6 2019**  
ASHLEY FARLEY & BEN PIERSON, GATES  
FOUNDATION

# Dina Paltoo, NIH

As far as machine readable data management plans, I think this is really important because it also fosters that connection.

It allows, it creates, some consistency. It also allows for the ability for investigators to be able to update their data management plan . . .

. . . So everything can be linked and connected. The data can be attached to whatever the award is. It can be attached to the data management plan, and can be followed appropriately. And, then others who would be able to find and use that data would be able to know where it came from.

# Margaret Levenstein, ICPSR

If you said in the data management plan, “Oh, yes give the journal persistent identifiers and things all the data are when you go to the journal you say “Oh accountability measures built into a data management plan if it’s confidential there’s actually some machine actionable things that are really much harder to leverage if it’s a pdf stored someplace nobody can see.



# Stakeholders for Machine actionable DMP content

*Everyone has a stake in this*, the researchers themselves, the office of grants and sponsored projects . . . the library . . . the IT folks, particularly if the DMP is revised to call for larger amounts of storage . . .

Clifford Lynch



**Cliff Lynch**



# Exposing Data Management Plans

RFC open April 9th. Twelve recommendations covering five main points:

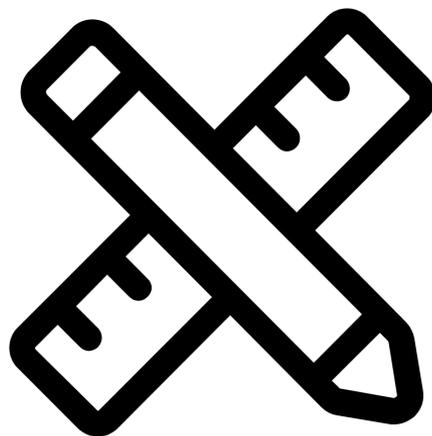
1. FAIR DMPs for FAIR data production
2. Ethical exposure of DMP content
3. Standardised metadata for DMPs
4. Controlled vocabularies in DMPs
5. Persistent identifiers in DMPs, and for DMPs

Open For Comment and Review at  
<https://tinyurl.com/RDA-ExposingDMP>

# Next steps: Draft recommendations

## Stakeholder groups

- **Libraries**
- Funder
- Publisher
- Researcher
- Vice Provost of Research
- Tool Builder



# Next steps: Draft recommendations

- Will be presented via webinar for comment
- Contributing to AAU/APLU Guide for Accelerating Public Access to Research Data

<https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/workshop-on-public-access-report-aplu-aaau-2019.pdf>

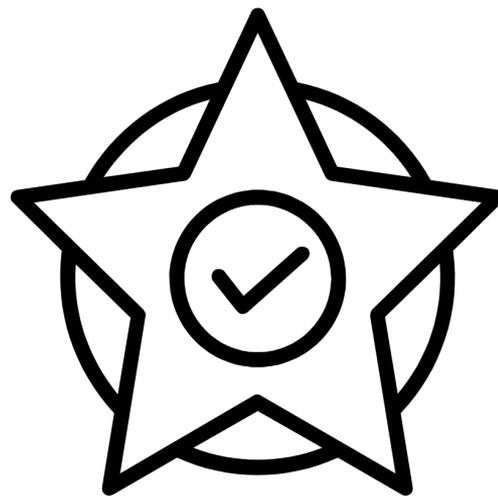
- December meeting agenda

<https://www.arl.org/implementing-effective-data-practices/>

# Example: Library recommendations

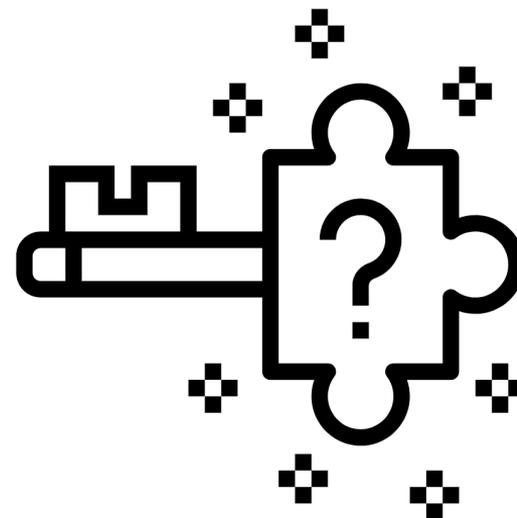
For each group, we identified:

- Problem statement
- Missing pieces
- Actionable agenda



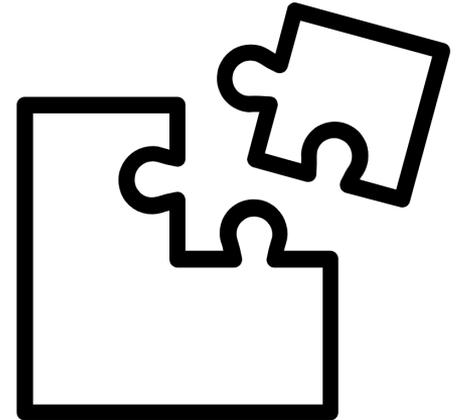
# Example: Problem Statement

- Library expertise in PIDs/DMPs/data management needs greater visibility
- Lack of centralized funding
- Lack of standardization on use of PIDs/DMPs/data management



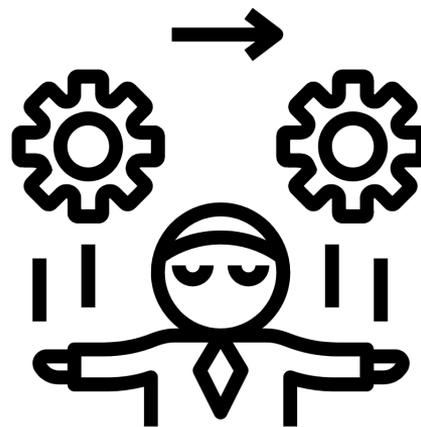
# Example: Missing Pieces

Collaboration and coordination among campus partners to elevate use of these services, and infrastructure to support that use



# Example: Actionable agenda

- Library + stakeholder partnerships to develop institutional policies on data management and sharing
  - includes institutional expectations for DMPs/PIDs
- Library is point of contact for DMP/PID expertise, infrastructure, education, training



# Thank You

Questions?

contact: [judy@arl.org](mailto:judy@arl.org)