Implementing Effective Data Practices

May 6, 2020
CNI Virtual Membership Meeting

Maria Gould * Natalie Meyers * Maria Praetzellis
Jenny Muilenburg * Judy Ruttenberg
Introduction

"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

- DOI, ORCID, ROR, etc.
- Workflows
- Systems
- Standards
- Infrastructure
- Metadata indexing
- System connections
- Search & 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
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**.
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.
“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.”
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...
Margaret Levenstein, ICPSR

If you said in the data management plan, “Oh, yes give me money to do this research and I’ll share the data” then when you go to the journal you say “Oh I can’t share my data it’s confidential.” There’s actually some transparency in this whole thing. On the other hand, you can give the journal persistent identifiers and these things all reinforce one another. There’s some accountability measures built into a data management plan if it’s machine actionable 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
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
- 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 PID/DMPs/data management needs greater visibility
- Lack of centralized funding
- Lack of standardization on use of PID/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