Making DMPs actionable and public

Kevin Ashley | Digital Curation Centre | kevin.ashley@ed.ac.uk
Sarah Jones | Digital Curation Centre | sarah.jones@glasgow.ac.uk
Daniel Mietchen | National Institutes of Health | daniel.mietchen@nih.gov
Stephanie Simms | California Digital Library | stephanie.simms@ucop.edu
Angus Whyte | Digital Curation Centre | a.whyte@ed.ac.uk

Defining machine-actionable DMPs

This term refers to information that is structured in a consistent way so that machines, or computers, can be programmed against the structure.

Data Documentation Initiative

Stakeholders and benefits

Researchers

Institutions / Organizations

Repositories / Infrastructure

Funders

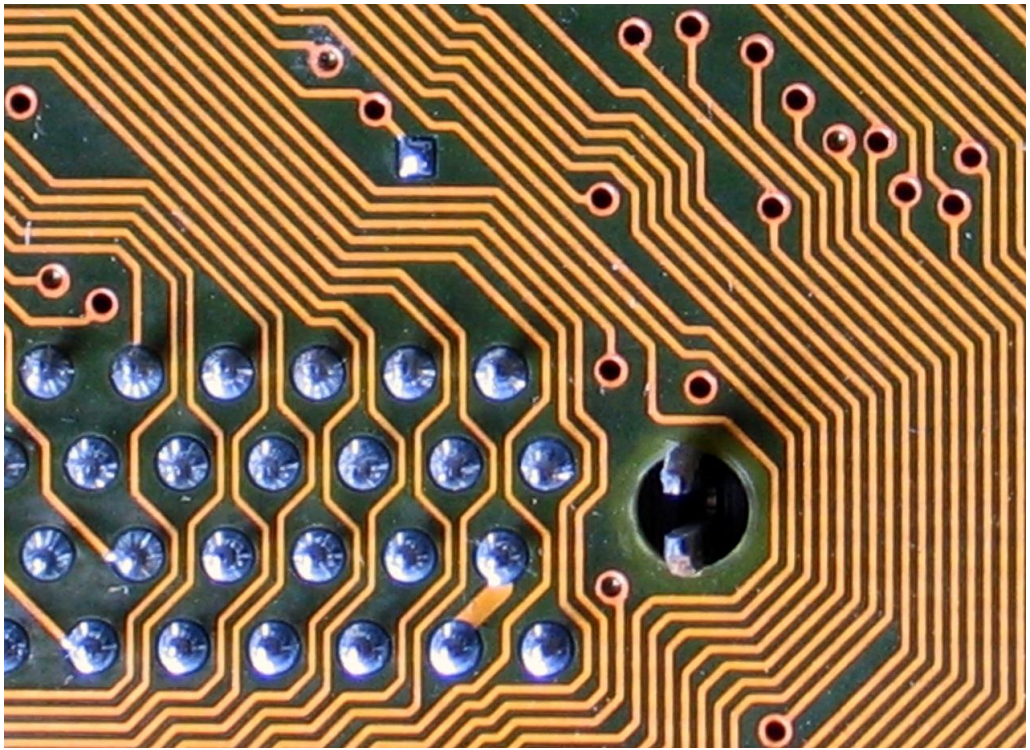
Publishers?

NIH/National Eye Institute. Public Domain

What do we want machine-actionable DMPs to do?

They should enable...

- Discovery
- Consistency
- Resource allocation
- Compliance monitoring



Review of current work (i): DMP vocabularies

- CASRAI DMP Ethics Review vocabulary
  - embargoes, intellectual property rights, security
- DMP extension of the DataID ontology
  - RDF, DMP lifecycle, includes re3data schema
- DCC themes for DMPs
  - Questions and guidance in DMPonline are tagged with themes to match/overlay them.
- Data Practices and Curation vocabulary (DPCVocab)
  - 3 categories—Research Data Practices, Data, Curation—with 187 terms

Review of current work (ii): DMP standards

- DART Project rubric for DMP evaluation
  - Analytic rubric to standardize review of NSF DMPs
  - UK group using it to create rubrics for UK funders
- Univ. of Colorado competition for best DMPs (2014 –)
  - $2k prize for best DMPs from 5 disciplines
  - Guidelines for DMPs on library website
- Ten Simple Rules for Creating a Good DMP (Michener 2015)
  - Establish relation to relevant policies
  - Define & explain data structure, provenance, preservation

Review of current work (iii): DMP platforms

- DMPTool-DMPonline merge
  - Opportunity to support standards at the point of use
- RIO Journal publishing DMPs
  - All in JATS XML, can be updated

- Your ideas here...

Themes in DMPonline

- Existing Data
- Data Description
- Data Format
- Data Type
- Data Volumes

- Data Capture Methods
- Documentation
- Metadata
- Data Quality

- Ethical Issues
- IPR Ownership and Licensing
- Data Security
- Storage and Backup

- ID
- Project Description
- Related Policies

- Expected Reuse
- Discovery by Users
- Method for Data Sharing
- Timeframe for Data Sharing
- Restrictions on Sharing
- Managed Access Procedures

- Data Selection
- Period of Preservation
- Preservation Plan
- Data Repository

- Responsibilities
- Resourcing

Refining the themes

- 28 themes reduced to 16
- Initially granular to represent each concern, e.g., how, when, under what restrictions or agreements data would be shared. Propose to merge similar concerns.
- Implement for all DMP templates worldwide.
- DMPonline/DMPTool and hosted instances have > 20k users. Opportunity to test the potential of basic tagging, e.g., text mining, before exploring more specific vocabulary.

Common DMP vocabulary

- Should map to and/or from
  - elements of data management workflows
  - policy requirements
  - suitable controlled vocabularies, e.g., discipline-specific (if available)
- Should keep in mind
  - we still need a human-readable document with a narrative
  - researchers resent form-filling exercises
  - needs to be updatable throughout lifecycle

Use cases for maDMPs

- Data discovery
- Capacity planning / Resource allocation
- Aggregation / Integration
- Policy compliance

Use case: maDMPs as a discovery tool

- Could facilitate discovery using any element of the core data model, across DMPs
- For example, it would be possible to watch out for new data
  - Of a particular kind (e.g. MRI scans of Alzheimer patients)
  - Acquired with a particular method/ instrument
  - Acquired by particular people/ labs/ institutions
  - With a particular license
- By inference, it would be possible to learn about
  - Different teams producing or curating the same or related data
    - Who is doing what around the Zika virus outbreak right now?
  - Ongoing replications of the same original studies
  - Field trips planned by different teams to the same location
- Making them public broadens the community that can make use of this tool

Funder use cases: Horizon 2020

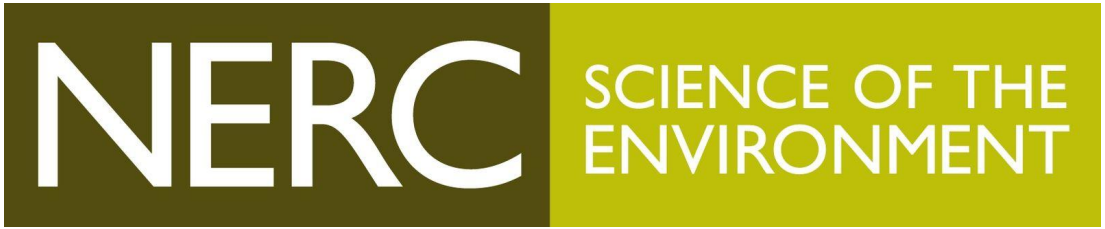
- Deposit of DMPs in repositories
  - Work planned under OpenAIRE, e.g., B2SHARE and Zenodo
- Compliance checking of data deposit in named repositories
  - DOI fed back into tool to update DMP

Horizon 2020
European Union funding
for Research & Innovation

Funder use cases: NERC

NERC: Natural Environment Research Council, UK

- Deposit of DMPs in NERC repositories
  - 7 disciplinary data centers
- Compliance checking of data deposit in named repositories
  - DOI fed back into tool to update DMP
- Support DMP lifecycle
  - trigger notification to begin next phase when project award made for funders with multi-stage requirements
  - push award details (grant IDs etc) back into DMPs



Repository use case (i)

Repository recommender service via re3data.org

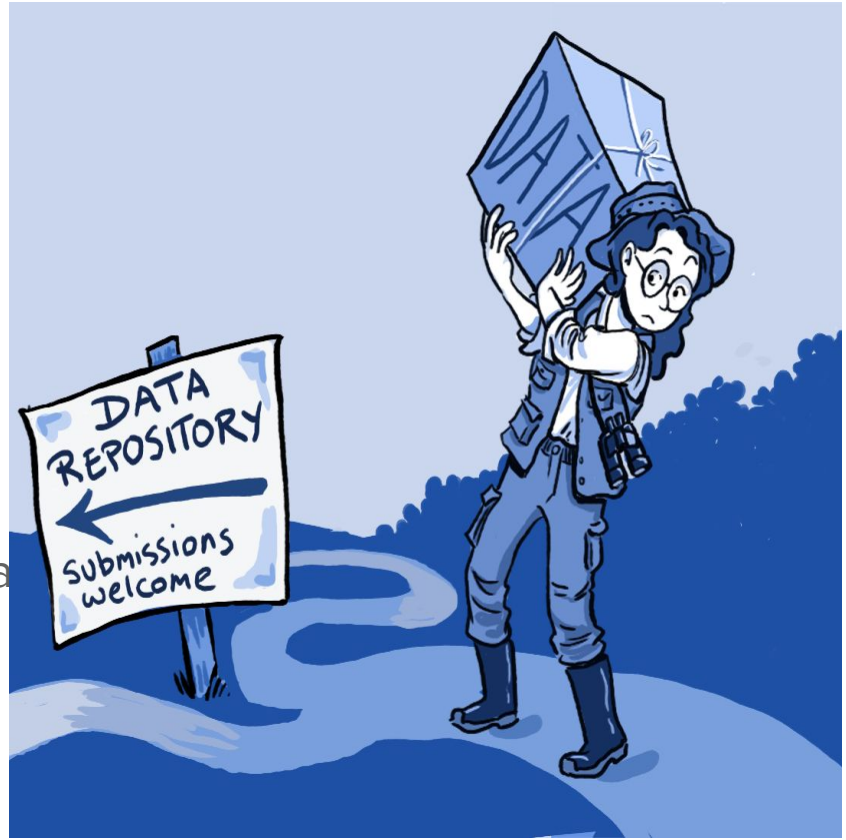
- Automated function for data tracking
- Provides info about metadata standards, etc. at beginning of project
- Can notify repository of data in pipeline for planning (repository use case ii)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Repository use case (ii)

Text/data mine to ping data repositories or other collections (e.g. in biobanks, museums) when mentioned in a DMP, e.g., using:

- Substance Editor
- SciCrunch
- DMPonline themes to help identify relevant text to mine



Persistent identifiers (PIDs)

- Assign a DOI to DMP of record (i.e., submitted with grant proposal). Use this to get award details back into a DMP .
- Leverage other PIDs to populate DMP over time:
  - Researcher IDs (ORCIDs)
  - Funder IDs (FundRef)
  - Grant IDs
  - Research Resource IDs (RRIDs)
    - antibodies, organisms, cell lines, tools
  - Etc.
- Also enables compliance monitoring

USGS, Public Domain

DMPonline API to create a plan

- Research offices can auto-generate plan with correct template, guidance, and basic plan details (PI, project name) when they know one is required
- Takes the burden off researchers
- Additional work planned to connect DMP tools with common research information management systems (e.g., PURE)
- Can integrate to assign DOI at point of grant submission

Layered DMP use case

- Higher-level plans to reduce burden of creating project-level plan where appropriate; interoperability across layers
- Can we map this to DMPonline/DMPTool? see how existing descriptions of large-scale infrastructure can be mapped to tools we have?
- ’Big Science’ projects (e.g, at CERN), infrastructure projects ELIXIR, individual

Prioritization & next steps

- Core vocabulary and standard way to extend
- DMPTool-DMPonline merged platform
  - begin deploying use cases
  - continue collecting new ones
- Education/outreach to make DMPs better
  - hypothetical use cases are good, but DMPs need to be better for them to work
- Continue integrating systems

Making DMPs public, open and discoverable

- On a website
- In a repository
- In a journal
- Elsewhere

- Formats
  - PDF, HTML, machine-actionable
- Versioning
- Licensing
- Archiving

In this section

What do I need to do?

Example data management plans

- Several research funders require a data management plan (DMP) to be submitted as part of a grant application.
- DMPs vary in length depending on the funder requirement and the complexity of the data that will be generated; most are 1-4 pages long
- The examples below are for illustration. Where possible, Leeds examples are used but there are also links to DMPs outside Leeds.

AHRC

The Professional Career and Output of Trevor Jones : this Technical Plan was highly rated by peer reviewers. This plan is for a large project which involves digitising an archive of film music. The Technical Plan format in this document comes from an old template. The AHRC provides a list of mandatory headings for your Technical Plan.

Virtual Holocaust Memory: data outputs from the project include several videos and a project web site. AHRC peer reviewers requested clarification about how the outputs would be looked after beyond the end of the project. The requirement was satisfied with a short letter from the RDL team with additional information about the Research Data Leeds repository service. The AHRC response underlines the importance of considering how and where data can be shared during and after your project.

ESRC

Realist Evaluation of Adapted Sex Offender Treatment Programs for Men with Intellectual Disability. This plan looks at managing quantitative and qualitative data, including audio interviews. The plan discusses ethical issues around a very sensitive dataset, including using consent and anonymisation to generate data which can be shared and reused.

Slawa Rokicki Dataverse (Harvard University)

Metrics 1 Download

Study on Mhealth and Reproductive Health in Teens

Rokicki, Slawa; Fink, Gunther, 2015, "Study on Mhealth and Reproductive Health in Teens", http://dx.doi.org/10.7910/DVN/23418, Harvard Dataverse, V1

Download Citation

If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description

The dataverse will include the necessary documents to replicate the Study on Mhealth and Reproductive Health in Teens. It includes the Data Management Plan, Statistical Analysis Plan, data, and code.

Files | Metadata | Terms | Versions

Search this dataset... Find

2 Files

Download

SMART Data Management Plan 11_18_13_2.docx
Unknown - 114.5 KB - Oct 22, 2015 - 1 Download
MD5: d12e40e85f0a131b11ece23462b5300b;

Download

SMART Statistical Analysis Plan V2_3.docx
Unknown - 167.1 KB - Oct 22, 2015 - 0 Downloads
MD5: 406fbb0a5c65b6df0185d6a8f2f51478;
Pre-specifies the statistical analysis for SMART.
SMART Statistical Analysis Plan

Download

Public DMPs: DMPTool

Stephanie

https://dmptool.org/public_dmps?public%3Apage=7

DMPTool

Home | DMP Requirements | Public DMPs | News | Help | Contact Us | About | Log In

PUBLIC DMPS

Use the search box or the A-Z links below to narrow down the list by plan title. Public DMPs are plans created using the DMPTool and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Search

A - F | G - L | M - S | T - Z | All

Plan Title	Funder Template	Institution	Owner	Download
Precipitation and Temperature Variation Monitoring Project	NSF-AGS: Atmospheric and Geospace Sciences	University of Tennessee	Chandler White	
Management Plan for the Annotation of Cryptosporidium baileyi	NIH-GEN: Generic	University of Georgia	Shelton Griffith	
A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy	NSF-SBE: Social, Behavioral, Economic Sciences	Non Partner Institution	Josh Fisher	
keenjhar lake	U.S. Geological Survey DMP Guidance	Non Partner Institution	sohail raza	
ASSESSMENT OF WATER QUALITY OF HOROOLO DRAIN	NSF-AGS: Atmospheric and Geospace Sciences	Non Partner Institution	Arif Asghar	
Multi-level Adaptive Agents for Search Space Control	NSF-ENG: Engineering	Missouri University of Science and Technology	Steven Corns	

Browse

Articles
Authors
Collections

In this collection

Papers published: **4**

Printed version: **Paperback**

OPEN ACCESS

Public Data Management Plans created with the DMPTool

This collection contains data management plans (DMPs) created with the DMPTool – a free online application that helps researchers build data management plans for specific U.S. funding agencies.

Data Management Plan for PhD Thesis "Climatic Limitation of Alien Weeds in New Zealand: Enhancing Species Distribution Models with Field Data"

Jennifer Pannell

Data Management Plan doi: 10.3897/rio.2.e8664

05-04-2016 Unique: 404 | Total: 657 Reprint: € 2,30

HTML XML PDF

A Political Ecology of Value: A Cohort-Based Ethnography of the Environmental Turn in Nicaraguan Urban Social Policy

Josh Fisher, Alex Nading

Data Management Plan doi: 10.3897/rio.2.e8720

05-04-2016 Unique: 226 | Total: 421 Reprint: € 2,30

HTML XML PDF

Coastal Data Information Program (CDIP)

Jennifer McWhorter, Darren Wright, Julie Thomas

Data Management Plan doi: 10.3897/rio.2.e8827

15-04-2016 Unique: 198 | Total: 321 Reprint: € 3,60

HTML XML PDF

Realizing the greater potential of DMPs

We all want to move DMPs beyond a culture of compliance to promote culture change

This involves such lofty goals as:

- Linking DMPs to their actual implementation
- Advancing open scholarship
- Using the DMP as a training platform to accomplish these things

Outlook

- Which workflows can we imagine around machine-actionable & public DMPs?
- What role can public DMPs play in education & training for data management?
- What if DMPs were accessible via public Jupyter notebooks by default?
- How can DMPs interact with each other, within & across layers?
- Which versions of a DMP should be archived for how long?
- Which resources should a DMP talk to/ be notified from?
- What actions could or should DMPs trigger?
- Who should know a DMP was updated?
- When should DMPs be updated?
- ...

Summary

Think of DMPs as key elements of a networked data management ecosystem:

- connected via a shared vocabulary
- actionable by humans and software
- versioned
- public