Data Management Plans

Material from: Sarah Jones

Digital Curation Centre, Glasgow

Twitter: @sjDCC

Presented by: Marcela Alfaro Córdoba

CODATA Research Data Science School, 6 December 2018, Sao Paulo #dataSaoPaulo18

What is a data management plan?

A brief plan written at the start of a project to define:
- how the data will be created?
- how it will be documented?
- who will access it?
- where it will be stored?
- who will back it up?
- whether (and how) it will be shared & preserved?

DMPs are often submitted as part of grant applications, but are useful whenever researchers are creating data.

Why write a DMP?

NON PECUNIAE INVESTIGATIONIS CURATORE

SED VITAE FACIMUS PROGRAMMAS DATORUM PROCURATIONIS

(Not for the research funder, but for life we make data management plans)

- Make informed decisions to anticipate and avoid problems

- Avoid duplication, data loss and security breaches

- Develop procedures early on for consistency

- Ensure data are accurate, complete, reliable and secure

- Save time and effort to make your life easier!

Don't undervalue research data

PUBLICATIONS AND DATA

DCC Checklist for a DMP

The DCC assessed existing funder requirements, DMP templates and other best practice to see what should be included in plans. This was synthesised down into common themes and questions.

- 13 questions on what's asked across the board

- Prompts / pointers to help researchers get started

- Guidance on how to answer

www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

Common themes in DMPs

1. Description of data to be collected / created
    (i.e. content, type, format, volume...)

2. Standards / methodologies for data collection & management

3. Ethics and Intellectual Property
    (highlight any restrictions on data sharing e.g. embargoes, confidentiality)

4. Plans for data sharing and access
    (i.e. how, when, to whom)

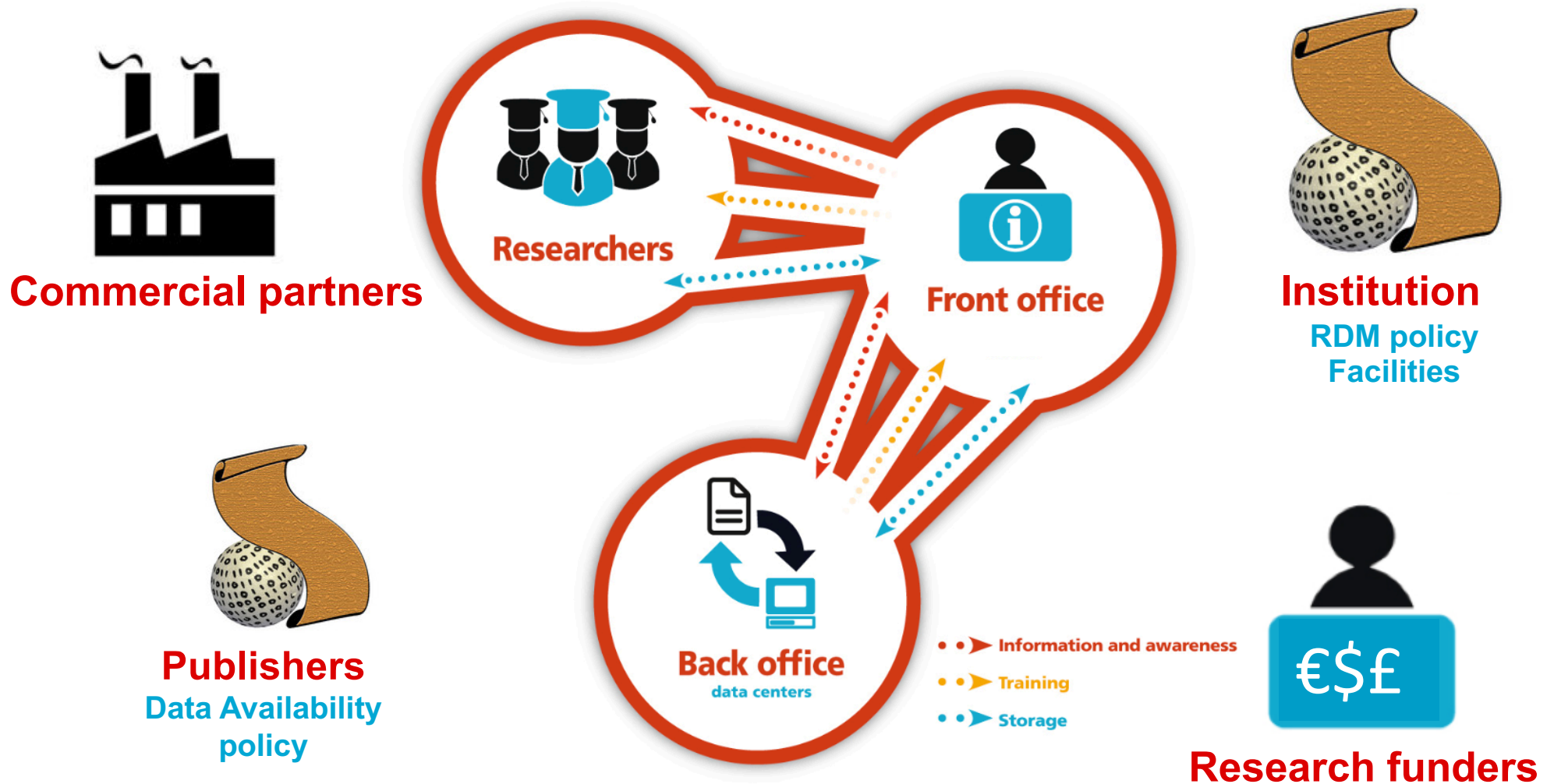5. Strategy for long-term preservation

Planning trick 1: think backwards

What data organisation would a re-user like?
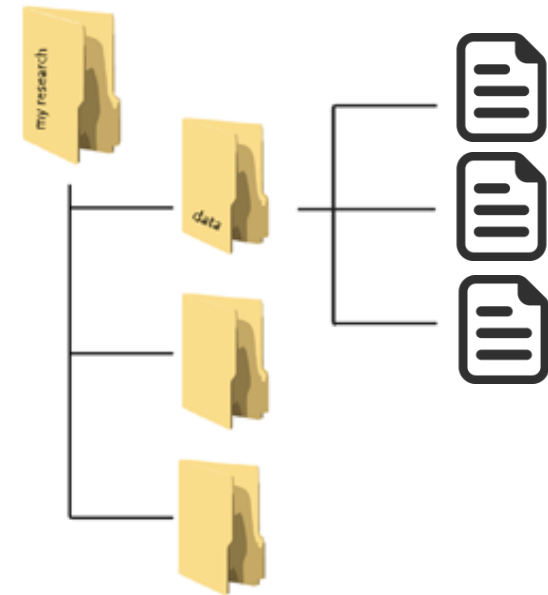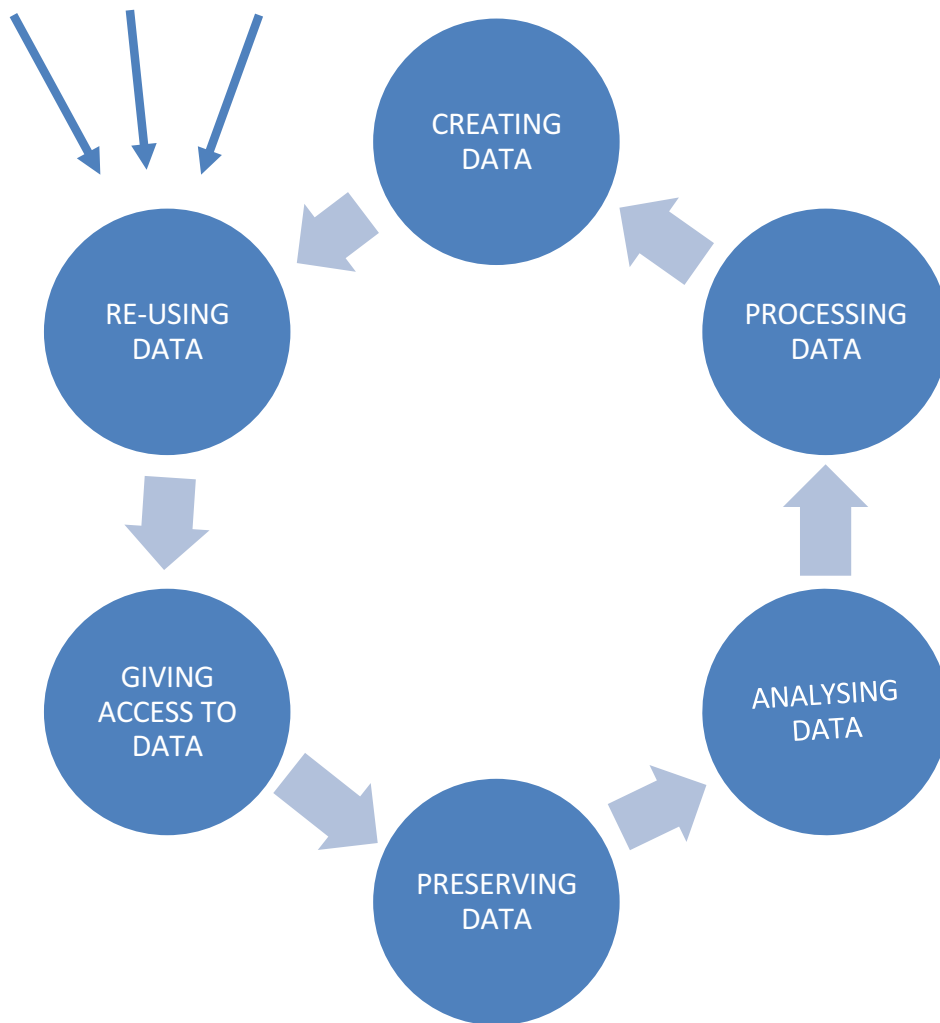
Design how you will organise data in the project (folder structure, file naming convention, …)

Planning trick 2: include RDM stakeholders

Commercial partners

Publishers
Data Availability
policy

Researchers

Front office

Back office
data centers

● ●▶ **Information and awareness**
● ●▶ **Training**
● ●▶ **Storage**

Institution
RDM policy
Facilities

€$£

Research funders

www.openaire.eu/briefpaper-rdm-infonoads

Planning trick 3: ground your plan in reality

Base plans on available skills, support and good practice for the field – show it's feasible to implement

What makes a good DMP?

- Clear, detailed information that is relevant to the science

  - adopting recognised standards

  - practices in line with norms for that field

  - use of support services e.g. university storage, subject repositories…

- Realistic approach that is feasible to implement

- Evidence of consultation and seeking advice

- Proper justification of restrictions and costs

Have you taken time to reflect on what to do?

Is the information specific enough?

"we will use suitable formats to ensure that our data can be preserved and sustained over the long term"

- Which standards? Name them!

- Show that you know which are suitable

- Does your chosen repository have preferences?

Are decisions justified?

"data will be made available upon request to bona fide medieval historians"

- Why is it restricted?

- Could other communities not reuse the data?

- Will the research team be around to handle access requests in the future?

A better response...

"We will provide MP3 audio files for online dissemination. While this is not an open format, it is well-established and the most widely supported. High-resolution WAV files will be used for the archival master recordings."

- Be clear, specific and detailed

- Justify decisions

Example plans

Plans from several funders and disciplines via DCC
www.dcc.ac.uk/resources/data-management-plans/guidance-examples

108 DMPs from the National Endowment for the Humanities
www.neh.gov/divisions/odh/grant-news/data-management-plans-successful-grant-applications-2011-2014-now-available

Scientific DMPs submitted to the NSF (USA) provided by DataOne
https://www.dataone.org/data-management-planning

DMPs published in RIO journal
http://riojournal.com/browse_user_collection_documents.php?collection_id=3&journal_id=17

Share yours! - www.dcc.ac.uk/share-DMPs

Data description examples

The final dataset will include self-reported demographic and behavioural data from interviews with the subjects and laboratory data from urine specimens provided.

From NIH data sharing statements

Every two days, we will subsample E. affinis populations growing under our treatment conditions. We will use a microscope to identify the life stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook and then copy the data into an Excel spreadsheet. The Excel spreadsheet will be saved as a comma separated value (.csv) file.

From DataOne – E. affinis DMP example

Metadata examples

Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.

From ICPSR Framework for Creating a DMP

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and describe all columns, units, abbreviations, and missing value identifiers. These notes will be transcribed into a .txt document that will be stored with the data file. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. EML is one of the accepted formats used in ecology, and works well for the types of data we will be producing. We will create these metadata using Morpho software, available through KNB. The metadata will fully describe the data files and the context of the measurements.

From DataOne – E. affinis DMP example

Data sharing examples

The videos will be made available via the bristol.ac.uk website (both as streaming media and downloads) HD and SD versions will be provided to accommodate those with lower bandwidth. Videos will also be made available via Vimeo, a platform that is already well used by research students at Bristol. Appropriate metadata will also be provided to the existing Vimeo standard.

All video will also be available for download and re-editing by third parties. To facilitate this Creative Commons licenses will be assigned to each item. In order to ensure this usage is possible, the required permissions will be gathered from participants (using a suitable release form) before recording commences.

From [University of Bristol Kitchen Cosmology DMP](#)

We will make the data and associated documentation available to users under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed.

From [NIH data sharing statements](#)

Examples restrictions

Because the STDs being studied are reportable diseases, we will be collecting identifying information. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement.

From [NIH data sharing statements](#)

1. Share data privately within 1 year.
   Data will be held in Private Repository, but metadata will be public

2. Release data to public within 2 years.
   Encouraged after one year to release data for public access.

3. Request, in writing, data privacy up to 4 years.
   Extensions beyond 3 years will only be granted for compelling cases.

4. Consult with creators of private CZO datasets prior to use.
   Pis required to seek consent before using private data they can access

From [Boulder Creek Critical Zone Observatory DMP](#)

Archiving examples

The investigators will work with staff at the UKDA to determine what to archive and how long the deposited data should be retained. Future long-term use of the data will be ensured by placing a copy of the data into the repository.
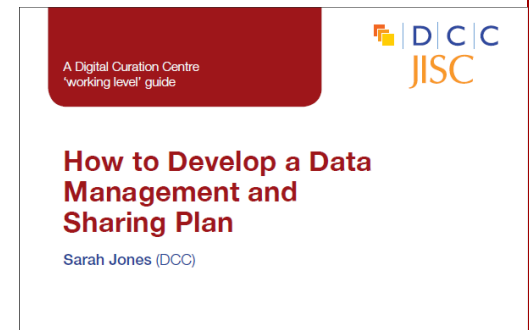
From ICPSR Framework for Creating a DMP

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service. For example, SPSS Portal format and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Appropriate documentation necessary to understand the data will also be provided. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule. Biological samples (output 3) will be deposited with the UK BioBank for future use.

From Writing a Wellcome Trust Data Management and Sharing Plan

DCC support on DMPs

- Webinars and training materials

- How-to guides and other advisory documents

- Checklist on what to cover in DMPs

- Example DMPs

- DMPonline

www.dcc.ac.uk/resources/data-management-plans

A Digital Curation Centre
'working level' guide

D|C|C
JISC

How to Develop a Data Management and Sharing Plan

Sarah Jones (DCC)

DCC Checklist for a Data Management Plan

D|C|C

DMP ONLINE

Guidance from elsewhere

Framework for Creating a Data Management Plan

This framework can be used as an outline in assembling data management plans to accompany grant applications. Note that some funders have page limits for data management plans—NSF limits plans to two pages.

Elements of a Data Management Plan

This list of elements is informed by a gap analysis that ICPSR conducted of existing recommendations for data management plans and other forms of guidance made available for researchers generating data. The result of the gap analysis was a comparison of existing forms of guidance. Elements that are highly recommended for inclusion in effective data management plans are noted.

See our bibliography for additional readings germane to the elements of a data management plan.

Data Description (Recommended)

Provide a brief description of the information to be gathered -- the nature, scope, and scale of the data that will be generated or collected

Why this is important
A good description of the data to be collected will help reviewers understand the characteristics of the data, their relationship to existing

Example 1:
This project will produce public-use nationally representative survey data for the United States covering Americans' social backgrounds, enduring political predispositions, social and political values, perceptions and evaluations of groups and candidates, opinions on questions of public policy, and participation in political life.

Example 2:
This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and corre nationally representative sample of over 8000 respondents. The sensitive nature of these data will require that the data be released thr

> Think about why the questions are being asked – why is it useful to consider that topic?

> Look at examples to help you understand what to write

www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html

What is DMPonline?

A web-based tool to help researchers write DMPs

https://dmponline.dcc.ac.uk

Main features in DMPonline

- Templates for different requirements (funder or institution)

- Tailored guidance (funder, institutional, discipline-specific etc)

- Ability to provide examples and suggested answers

- Supports multiple phases (e.g. pre- / during / post-project)

- Granular read / write / share permissions

- Customised exports to a variety of formats

- Shibboleth authentication

A single platform for all things DMP

Agreed to converge on a single codebase, based on DMPonline with additional features from DMPTool

Bring together features and strengths of each tool

Co-manage, co-develop and issue joint roadmap

DMPRoadmap: https://github.com/DMPRoadmap

Key messages

- Data management is part of good practice whether you plan to make the data open or not

  - **it benefits you!**

- The process of planning is as important as the DMP. Think about the desired end result and plan for this.

- Approach DMPs in whatever way best fits your project. Don't just let funder requirements drive things.

What we learn from history

- "No battle plan survives contact with the enemy"

    - Helmuth von Moltke the elder

- "... I have always found that plans are useless, but planning is essential"

    - Dwight D. Eisenhower

- The data management plan is an evolving object. It should change, along with the research.

Thanks for listening

DCC resources on DMPs:

www.dcc.ac.uk/resources/data-management-plans

Follow us on twitter:

@DMPonline and #ukdcc