



# Creating and Implementing Data Management Plans (DMPs)

---

**S. Venkataraman**

Research Data Specialist  
Digital Curation Centre

Open Science Forum, University of Luxembourg Workshop, 15<sup>th</sup>  
November, 2018, Luxembourg

# What will we cover?

1. How are funders currently evaluating DMPs
2. What makes a 'good' DMP
3. Lessons from DART and UK implementation
4. Exercise

# Implementing DMP requirements

Lots of funders have DMP requirements but the picture in terms of review differs and monitoring is patchier still



Arts & Humanities  
Research Council



wellcome



CANCER  
RESEARCH  
UK



Science & Technology  
Facilities Council



Engineering and Physical Sciences  
Research Council

# AHRC

- Established a technical peer review college
- If a proposal has a significant technical component, 1 of the 4 reviewers will be a technical reviewer
- Reviewers see technical plan AND full application. They submit a written technical review, but don't sit on academic review panel.
- A poor DMP can be a reason for rejecting a proposal, but conditional awards are more likely

[www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan](http://www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan)



# BBSRC

An application's Data Management Plan **will be assessed by reviewers** and BBSRC responsive mode Research Committees or assessment Panels. The plan will be **considered separately from the scientific excellence** of the proposed research; however, an application's credibility will suffer if peer review agrees the statement is inappropriate. In the case where a highly-rated proposal has an inappropriate Data Management Plan, Research Committees and Panels may choose to **offer conditional awards** and/or provide specific feedback to the applicants. Appropriate plans are expected to be those where the proposed data sharing activities are in-line with current best practice in the field and both the scientific and cost benefits are considered.

[www.bbsrc.ac.uk/funding/apply/application-guidance/data-management](http://www.bbsrc.ac.uk/funding/apply/application-guidance/data-management)

# ESRC and NERC

- DMPs reviewed by proposal reviewers
- ESRC guidelines to assist reviewers to check DMPs  
[www.esrc.ac.uk/files/funding/guidance-for-peer-reviewers/data-management-plan-guidance-for-peer-reviewers](http://www.esrc.ac.uk/files/funding/guidance-for-peer-reviewers/data-management-plan-guidance-for-peer-reviewers)
- NERC data centres co-write full DMP post-award
- ESRC may withhold final grant payment if data are not offered for deposit with UKDA
- Those who do not meet data requirements risk having award payments withheld or becoming ineligible for future funding from NERC.

# Horizon 2020

- DMPs are a deliverable, checked primarily by project officers and in some cases external reviewers too
- Guidelines are being developed to give reviewers pointers on what to check. These are based on the template.
- The reviewer has access to the full project documentation
- Process is only just evolving and this is a pilot so feedback may be variable initially

# What makes a good DMP?

Has the researcher taken time  
to reflect on what to do?

- There are no absolute right answers.
- You want to be reassured that due consideration has been given and the approach seems reasonable.
- Look for proper engagement with the issues.



# And does it fulfil...

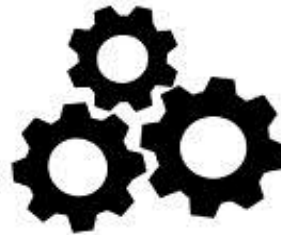
F  
indable



A  
ccessible



I  
nteroperable



R  
eusable



Image CC-BY-SA by [SangyaPundir](#)

# Key things to check

- Is the plan appropriate?
  - adopting relevant standards
  - practices in line with norms for that field
  - use of support services e.g. university storage, subject repositories...
- Does it seem feasible to implement?
- Has sufficient detailed information been provided?
- Has advice been sought where needed?
- Are restrictions and costs properly justified?

# 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!
- Does the team know which are suitable?
- Does the chosen repository have preferences?

# 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

# Make it easy for reviewers to evaluate

*“Online resource development will cost £21,000”*

versus

*“Online resource development, 60 days at £350”*

- Don't make reviewers dig around for information
- Be consistent in what you say in DMP and proposal

# Advice from reviewers

- First impressions count
  - Stick to page limits, follow the template if mandated, provide information in the relevant section...
- Beware blanket copy/paste
  - A limited amount of information can be provided as boilerplate text. Always read and adjust to your project
- Avoid hyperbole and buzzwords
  - Stick to clear statements and the strength of your technical approach will evidence itself

# Common themes to cover

- Data Description
- Standards and Metadata
- Data Sharing
- Archiving and preservation

# Data Description

- Is there a **full** description of the data to be produced? Statistics about the size, quantity and duration help reviewers to get a proper sense of scale.
- If third-party data will be reused, or the project will work with human subjects, has sharing been considered in the consent and licence agreements?



# Standards and Metadata

- Are metadata standards being used?
- Will sufficient metadata and documentation be provided to allow others to find, understand and reuse the data?
- Is the choice of file format explained, so it is clear that appropriate decisions have been made?

# Data Sharing

- Is it clear which data will be shared and with whom?
  - Are opportunities to share data openly maximised? e.g. by seeking consent to share, anonymising data...
  - If data can't be shared, are the reasons why explained?
- How will the data be shared? e.g. deposit in repository
- If an embargo period is planned, is that in line with norms for that discipline?

# Archiving and Preservation

- Will the research data be deposited in a suitable community database, repository or archive?
- Is it clear which data should be preserved and for how long?
- Are there any costs associated with preservation, and if so, how will these be covered?

# DART - DMPs as A Research Tool

- A 3 year IMLS funded project run across 5 US unis
- Developed an analytic rubric to standardize the review of data management plans
- Analysed DMPs to inform expansion or development of research data services at academic libraries

<https://osf.io/kh2y6>



# Lessons from DART

- Lots of useful info about what is going well and not
  - Basics often missing, 20-50% of plans didn't mention file formats
  - Data sharing most formalised in Geo and Bio (high use of repos)
  - Many others proposed to share via website or on request
  - Metadata wasn't addressed in majority of plans
- Analysing DMPs can help you to target support services
- When evaluating, assess what the DMP guidelines stipulate, not what you think the DMP should include!

# UK rubrics project

Collaboration established by Mary Donaldson to develop rubrics for major research funders.

- All RCUK
- CRUK
- Wellcome
- NIHR
- H2020
- A rubric for the 'ideal' non-funded DMP

<https://research-data-network.readme.io/docs/compliance-tools>

# Basic framework

Performance criteria (based on funder reqs )	Detailed	Performance Levels Incomplete	Not addressed
1. Summary of digital outputs and technologies	Plan provides a clear description ...	There is insufficient information to assess whether the overall plans are sound.	It is unclear what will be created or how.
2a. Standards and formats			
2b. Hardware and software			
2c....			

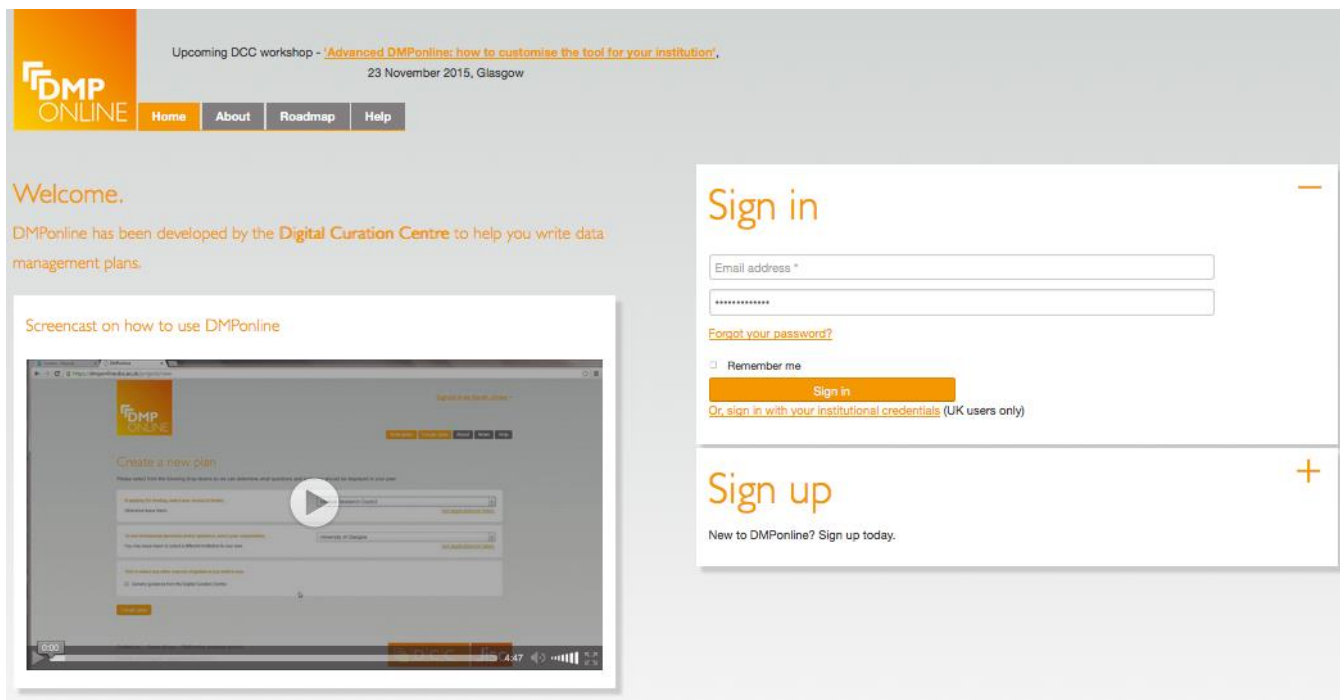
# DMPonline



# DMPonline

A web-based tool to help researchers write DMPs

Includes templates for funders, organisations, disciplines



The screenshot shows the DMPonline website interface. At the top left is the DMP ONLINE logo. To its right, a navigation menu contains links for Home, About, Roadmap, and Help. A header message reads: "Upcoming DCC workshop - '[Advanced DMPonline: how to customise the tool for your institution](#)', 23 November 2015, Glasgow". Below the navigation menu, a "Welcome." section states: "DMPonline has been developed by the [Digital Curation Centre](#) to help you write data management plans." A video player titled "Screencast on how to use DMPonline" is embedded, showing a preview of the tool's "Create a new plan" form. On the right side of the page, there are two white panels. The top panel is titled "Sign in" and contains a form with fields for "Email address \*" and a password field (masked with asterisks). Below the password field are links for "Forgot your password?", a "Remember me" checkbox, and a "Sign in" button. A note below the button says "Or, sign in with your institutional credentials (UK users only)". The bottom panel is titled "Sign up" and includes a "+" icon and the text "New to DMPonline? Sign up today."

<https://dmponline.dcc.ac.uk>

# How the tool works

## Create a new plan

Please select from the following drop-downs so we can determine what questions and guidance should be displayed in your plan.

If you aren't responding to specific requirements from a funder or an institution, [select here to write a generic DMP](#) based on the most common themes.

If applying for funding, select your research funder.

Otherwise leave blank.

European Commission (Horizon 2020)

[Not applicable/not listed.](#)

To see institutional questions and/or guidance, select your organisation.

You may leave blank or select a different organisation to your own.

University of Glasgow

[Not applicable/not listed.](#)

Tick to select any other sources of guidance you wish to see.

- DCC guidance
- EUDAT
- School of Humanities
- Computing

Create plan

Click to write a generic DMP

Or choose your funder to get their specific template

Pick your uni to add local guidance and to get their template if no funder applies

Choose any additional optional guidance

# Plan details: summary

This plan is based on:

Funder | Economic and Social Research Council

[Answer questions](#)

[Export](#)

The ESRC requires that all applicants seeking ESRC funding include a statement on data sharing in the relevant section of the Je-S application form. If data sharing is not possible, the applicant must present a strong argument to justify their case.

Sections	Questions
Existing data	<ul style="list-style-type: none"><li>- An explanation of the existing data sources that will be used by the research project (with references)</li><li>- An analysis of the gaps identified between the currently available and required data for the research</li></ul>
Information on the data that will be produced	<ul style="list-style-type: none"><li>- Methodologies for data collection</li><li>- Data volume and data type, e.g. qualitative or quantitative data</li><li>- Data quality, formats, standards documentation and metadata</li></ul>
Planned quality assurance and back-up procedures (security/storage)	<ul style="list-style-type: none"><li>- Quality Assurance</li><li>- Back-Up</li></ul>
Management and archiving of collected data	<ul style="list-style-type: none"><li>- Plans for management and archiving of collected data</li></ul>
Overcoming data sharing difficulties	<ul style="list-style-type: none"><li>- Expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.</li></ul>
Consent, confidentiality, anonymisation and other ethical considerations	<ul style="list-style-type: none"><li>- Explicit mention of consent, confidentiality, anonymisation and other ethical considerations</li></ul>
Copyright and intellectual property ownership of the data	<ul style="list-style-type: none"><li>- Copyright and intellectual property ownership of the data</li></ul>
Responsibilities for data management and curation	<ul style="list-style-type: none"><li>- Responsibilities for data management and curation</li></ul>

Summary of the sections and questions in your DMP

# Overview of sections in a DMP

The screenshot displays the 'DMP ONLINE' interface. At the top left is the logo. At the top right, it shows the user is signed in as 's.jones@arts.gla.ac.uk'. Below the logo, there are buttons for 'View plans', 'Create plan', 'About', 'News', and 'Help'. The main heading is 'My project (NERC Template)'. A callout box points to the 'Outline DMP' and 'Full DMP' tabs, stating 'Enables multiple phases'. Below the tabs is a progress bar indicating 'No questions have been answered'. The main content area lists several sections, each with a dropdown arrow on the right:

- Plan details
- Outline DMP
- Full DMP
- Organisation (3 questions, 0 answered)
- Roles and Responsibilities (1 question, 0 answered)
- Data Generation Activities (1 question, 0 answered)
- In-Project Data Management Approach (1 question, 0 answered)
- Metadata and Documentation (1 question, 0 answered)
- Data Quality (1 question, 0 answered)
- Exceptions or Additional Services (1 question, 0 answered)
- New Datasets (3 questions, 0 answered)
- Third Party/Existing Datasets (1 question, 0 answered)

A callout box at the bottom right states: 'Summary page with dropdown buttons to expand and answer each section'.

# Answering questions

My project (DCC Template) 1/13

Plan details **Generic DMP** Share Export

**Data Collection** (2 questions, 0 answered) +

**Documentation and Metadata** (1 question, 1 answered) -

What documentation and metadata will accompany the data?

**B** *I* Paragraph ▾ ▾

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.

**Notes who has answered the question and when**

**Progress bar updates how many questions remain**

Save

Answered less than a minute ago by Sarah Jones

**DCC Guidance** -

**Questions to consider:**

- What information is needed for the data to be read and interpreted in the future?
- How will you capture / create this documentation and metadata?
- What metadata standards will you use and why?

**Guidance:**

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.

Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.

# Site and funder-specific guidance

Plan details **Initial DMP (within first 6 months)** Mid-term Review DMP Final review DMP Share Export

**For each data set specify the following:** (5 questions, 0 answered)

The DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

Data set reference and name

**B** *I* [List] [List] [Link] [Grid] ▾

Save

Not answered yet

Data set description

**B** *I* [List] [List] [Link] [Grid] ▾

Save

**Guidance** Share note

**EC Guidance** +

**EUDAT guidance** -

EUDAT offers a number of [services](#) to host and manage research data. All data hosted within the EUDAT Collaborative Data Infrastructure (CDI) will be assigned a persistent identifier through the Handle system. This unique dataset reference can be used in data citations and to enhance discoverability.

**Guidance** Share note

**EC Guidance** +

**EUDAT guidance** -

The EUDAT Collaborative Data Infrastructure (CDI) provides research data hosting and management services for others, regardless of data type or format. The use of open, non-proprietary data formats is strongly encouraged though to enable interoperability and reuse. The [B2FIND catalogue](#) can be used to search for available datasets to reuse.

# Sharing plans

## Withdrawal of services for young people

Plan details   ESRC Data Management Questions   **Share**   Export

You can share your plan to allow others to read or edit it. Please insert the email address of the person you wish to share it with. You can leave a note to explain why you are sharing the plan, or what you wish them to look at.

### Collaborators

User name	Permissions	
Sarah Jones	Owner	
Laura Molloy	<input type="text" value="Read only"/>	<a href="#">Remove user access</a>

### Add collaborator

Permissions:

Allow colleagues to read-only, read-write, or become co-owners

# Co-writing DMPs

Signed in as [Laura Molloy](#) ▾

**DMP ONLINE**

[View plans](#) [Create plan](#) [About](#) [News](#) [Help](#)

## My project (DCC Template)

1/13

Plan details **Generic DMP** Export

**Data Collection** (2 questions, 0 answered) +

**Documentation and Metadata** (1 question, 1 answered) -

This section is locked for editing by Sarah Jones.

What documentation and metadata will accompany the data?

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.

Answered 7 minutes ago by Sarah Jones

**Ethics and Legal Compliance** (2 questions, 0 answered) +

**Storage and Backup** (2 questions, 0 answered) +

**Selection and Preservation** (2 questions, 0 answered) +

**Data Sharing** (2 questions, 0 answered) +

**Responsibilities and Resources** (2 questions, 0 answered) +

[Export](#)

Sections are locked for editing when they're being worked on by colleagues



# Exporting DMPs

## Withdrawal of services for young people

### ESRC Data Management Questions

Can export as plain text, PDF, html...

#### Existing data

Questions	Answers
- An explanation of the existing data sources that will be used by the research project (with references)	The ESDS archive has been systematically searched using a series of search terms related to and derivative of 'public service withdrawal', 'impacts', and/or 'children' and 'young people'. Our overall assessment is that there are no datasets that will adequately address the aims of this project. The following datasets are tangentially related to this project.
- An analysis of the gaps identified between the currently available and required data for the research	Given the contemporary nature of the proposed project, we know of no datasets that cover users' (especially young people's) views and experiences of austerity measures and service withdrawal. The proposed project will therefore capture new and unprecedented data, for which there is an evident demand among national and regional stakeholders, decisionmakers and service-providers (see 'Pathways to Impact' attachment). Moreover, whilst several datasets incorporate longitudinal data, none includes data gleaned from oral history and multigenerational family interview methods central to the proposed project. The proposed project therefore represents an extension to the methods and data quality of the tangentially-related projects listed in section 1. Finally, the novel use of 'impact' activities to generate data (e.g. workshops, mapping software) exceeds the scope of all extant and even tangentially-related datasets.

#### Information on the data that will be produced

Questions	Answers
- Data volume and data type, e.g. qualitative or quantitative data	The project shall generate new quantitative data (Bristol Online Survey outputs, SPSS data and outputs), qualitative data (digital audio files, audio transcripts, digital photographic and video data, workshop outputs, NVivo files), and mapping data (TIFF files). Metadata, in the form of pdfs and Excel spreadsheets, shall be used to facilitate the management and archiving of these data. Data shall be stored in password-protected folders on the host institution's secure servers. Data transfer between the PI and Co-I shall take place via face-to-face meetings.
	Quantitative data shall be generated from an anonymous online survey (target 10,000 responses). The survey will be administered via Bristol Online Surveys (BOS) software: a secure, quality-assured, widely-used online survey tool. Data will be exported to SPSS for analysis.

# Plans created within DMPonline tool can be updated ...

## Withdrawal of services for young adults

Plan details | **Initial DMP (within first 6 months)** | Mid-term Review DMP | Final review DMP | Share | Export

This page gives you an overview of your plan. It tells what your plan is based on and gives an overview of the questions that you will be asked.

Project name	Withdrawal of services for young adults
ID	-
Grant number	-
Principal Investigator/Researcher	Sarah Jones
Project data contact	-
Description	-

This plan is based on:

Funder	European Commission (Horizon 2020)
Institution	University of Glasgow

# Plans created within DMPonline tool can be updated ...

## Withdrawal of services for young adults

Plan details

Initial DMP (within first 6 months)

Mid-term Review DMP

Final review DMP

Share

Export

This page gives you an overview of your plan. It tells what your plan is based on and gives an overview of the questions that you will be asked.

Project name	Withdrawal of services for young adults
ID	-
Grant number	-
Principal Investigator/Researcher	Sarah Jones
Project data contact	-
Description	-

This plan is based on:

Funder	European Commission (Horizon 2020)
Institution	University of Glasgow

# DMP exercise



# Thank you!

---

In collaboration with:

