Why is this a good DMP?

Marjan Grootveld and Ellen Leenarts, DANS

TUD Seminar “Towards cultural change in data management - data stewardship in practice”, May 24th, 2018
Where do we go today?

- DANS and RDM
- DMP developments
- Horror and hooray
- How do you like these DMP sample texts?
Mission: promote and provide permanent access to digital research resources

Institute of Dutch Academy and Research Funding Organisation (KNAW & NWO) since 2005

First predecessor dates back to 1964 (Steinmetz Foundation), Historical Data Archive 1989
DANS and RDM

- RDM and DMP training and coaching
- Within European projects: RDM support, data stewardship skills, DMPlanning
- LCRDM working groups, a.o. on DMP support
- H2020 DMP reviewer for Research Executive Agency (REA/EC)

[Links to websites for more information]

https://doi.org/10.5281/zenodo.1120245
http://datasupport.researchdata.nl/
https://www.edugroepen.nl/sites/RDM_platform/SitePages/Home.aspx
https://www.cessda.eu/DMguide
Reviewing DMPs

When you have reviewed (draft) DMPs, how much time did you need on average?

1. Less than 30 minutes
2. 30-60 minutes
3. 60-90 minutes
4. More than 90 minutes
Harmonising funder templates

“Science Europe presents a framework for the creation of domain-specific protocols that can be used as standardised templates, reducing the administrative burden on both researchers, research organisations, and funders.”

DART - DMPs as A Research Tool

- A three-year project run across five universities in the USA
- 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
Collaboration established by Mary Donaldson to develop rubrics for major (UK) research funders.

Very useful summary of the process for writing DMP assessment rubrics

- Define the purpose: internal / to provide feedback?
- Develop DMP performance criteria that you would like to use
- Decide how many performance levels you need and what you would like them to be
- Develop descriptions of what constitutes the levels of performance

https://docs.google.com/document/d/1spKO9Fx0SFjcbYFLhGmuhA6veX06M5DjWGF_X7toePc/edit
## Basic framework

<table>
<thead>
<tr>
<th>Performance criteria (based on funder reqs)</th>
<th>Detailed</th>
<th>Performance Levels</th>
<th>Not addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Incomplete</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not addressed</td>
<td></td>
</tr>
</tbody>
</table>

2.1 Findability

2.1.1 Outline the discoverability of data (metadata provision)

<table>
<thead>
<tr>
<th>Performance criteria</th>
<th>Detailed</th>
<th>Performance Levels</th>
<th>Not addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan mentions at least one specific metadata schema with its URL</td>
<td>There is insufficient information to assess which metadata will be provided</td>
<td>It is unclear how people outside the consortium can discover the data</td>
<td></td>
</tr>
</tbody>
</table>

2.1.2. Outline the identifiability of data and refer to standard identification mechanism.

<table>
<thead>
<tr>
<th>Performance criteria</th>
<th>Detailed</th>
<th>Performance Levels</th>
<th>Not addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan mentions that long-term repository will provide DOI</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...
Lessons from DART and the UK group

- Each template requires its own rubric.
- When evaluating, assess what the DMP guidelines stipulate, not what you think the DMP should include!
- Only assess one aspect with each criterion.
- (Because of this,) Making a rubric is teamwork.

Hooray stories

Who wants to share a good reviewing experience?
What was so good about that DMP?
Horror stories

And who wants to share a disappointing DMP reviewing story? What was so bad about that DMP?
Let’s look at some sample texts
Common themes to address in DMPs

• Data description
• Standards and metadata
• Data sharing, accessibility, licensing
• Archiving and preservation
• FAIR – but may be interwoven with the above

Based on Sarah Jones: Developing and reviewing Data Management Plans.
DCC-DANS training, Amsterdam, 2017
The project will apply a mixed methods approach collecting both qualitative and quantitative data. Primary data will be mainly collected in the case study countries/regions (through surveys, interviews and focus groups), while secondary data will be collected from publicly available EU and national sources (such as Eurostat/Eurobarometer, academic and policy literature, party political programmes and policy documents and online media).

A brief description of each dataset is provided in table 2, including the data source, file formats and estimated volume to plan for storage and sharing.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
<th>Source</th>
<th>File format</th>
<th>Volume</th>
</tr>
</thead>
</table>
Sample text 2: DMP narrative

1 EXECUTIVE SUMMARY

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the applicants with regard to all the datasets that will be generated by the project.

According to Digital Curation Centre (DCC), a DMP contributes to save time and effort, makes research process easier, helps to validate if the necessary support is considered, and enables to make sound decisions.

The DMP supports project partners to:
- understand the data and use it when needed;
- ensure continuity if project staff leave or new researchers join;
- avoid unnecessary duplication e.g., re-collecting or re-working data.

https://doi.org/10.5281/zenodo.936394
Sample text 3: data collection

“We will begin fieldwork only after having obtained clearance from the <relevant> national data protection authority on the full anonymisation of the interviews.”
Sample text 4: data types and formats

<table>
<thead>
<tr>
<th>2.1 Describe the data that will be collected/generated and which you find relevant for reuse.</th>
<th>The data that will be collected and generated during this research project include: CAD drawings of wind tunnel models, Raw PIV images, Velocity and Uncertainty fields, Tecplot layouts for data visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2 Which type and format of data are these?</td>
<td>CAD drawings: .stp, Raw PIV images: .im7, .tif, Velocity and Uncertainty fields: .vc7, .dat, Tecplot layouts: .lay, .lpk</td>
</tr>
</tbody>
</table>

https://doi.org/10.5281/zenodo.1243736
## Summary of Data Format

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>Formats Used During Data Processing</th>
<th>Formats for Sharing, Reuse, and Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical or Textual Tabular Data</td>
<td>Microsoft Excel (.xls/.xlsx)</td>
<td>Comma-separated values (.csv)</td>
</tr>
<tr>
<td>Qualitative Textual Data</td>
<td>Microsoft Word (.doc/.docx)</td>
<td>Rich Text Format (.rtf) or text (.txt)</td>
</tr>
<tr>
<td>Audio Data</td>
<td>mp3 format (.mp3)</td>
<td>Audio recordings will be deleted after their transcription and only the processed transcripts will be shared and preserved.</td>
</tr>
<tr>
<td>Topic Modeling Data</td>
<td>Mallet format (.mallet)</td>
<td>Comma-separated values (.csv)</td>
</tr>
<tr>
<td>Simulation Model Data</td>
<td>Text model format (.mdl)</td>
<td>The mathematical model will be saved using standard differential equations symbols in .rtf, .csv and .txt files because the simulation model is developed using a proprietary software. Simulated values will be saved as numerical data, as specified above.</td>
</tr>
<tr>
<td>Statistical Data</td>
<td>STATA format (.dta)</td>
<td>Comma-separated values (.csv), Stata format (.dta)</td>
</tr>
</tbody>
</table>
The first relevant set of applicable standards will accrue to data collection methodology, expected to follow the guidelines spelt out by the ESOMAR and WAPOR survey research associations. In particular, these guidelines have clear rules for data collection that involves minors, requiring that informed consent of parents/legal guardians is obtained prior to any data collection, and that protection of any privacy-critical information is guaranteed. The lead of work package 5 will monitor contract compliance of the fieldwork agency through careful inspection of data and documentation files delivered after fieldwork.

Confidential DMP

ESOMAR: formerly known as The European Society for Opinion and Market Research
WAPOR: World association for public opinion research
“Both the collected and the generated data, anonymised or fictional, are not envisioned to be made openly accessible.”
2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.

Data will to the extent possible, legally and practically, be made available as our results are published and/or at the end of project period. A data embargo will typically be necessary to ensure that we can publish our results before other researchers have access to our own data.
“The source code will be released under an open source licensing scheme, whenever IPR of the partners is not infringed.”
The data will be treated as requiring the highest level of protection. A dissociation procedure is established to separate the actual data from the identity of the consultants.

The dissociation file will contain the data to identify each person that has accepted to participate in the Project and the link to a randomly assigned and unique identifier code (id), which will be used to identify him/her in the videos and in the rest of information about him/her.

This file will be segregated from the rest of the project information. There will be a dissociation file for each partner, and it will not be shared between the partners. Each partner will be responsible for the safekeeping of his file following the security measures listed below:
2.4.5 Specify the length of time for which the data will remain re-usable
This may vary for the type of data, and difficult to specify at this stage of the project.
Specify the length of time for which the data will remain re-usable

Data will be made available for ten years, from 12 months after the project end in line with University Policy and EC Horizon 2020 policies. The metadata will be kept indefinitely.
The Data Management Plan will be assessed bi-annually and further developed as the project progresses, under the leadership of the Project Management Team and to be decided on by the Work Package (WP) leaders together. All updates to the DMP will at latest be included in time for the periodic evaluation of the project, in project month 12, 30 and 48.
What makes a good DMP?

- Does the plan show data and Open RDM awareness?
  - Is the focus on the data or on publications?
    - A publication that describes the data $\neq$ depositing the data.
  - Is Openness on the author’s mind?
  - Does the plan distinguish between storing and archiving?

- Does the plan show that data management has many stakeholders? IT department, ethics committee, long-term repository...

- Overall approach to DMPlanning
  - Is the plan specific, or mainly intentional?
    - “should”, “possibly”, “where suitable/ appropriate/ relevant”...
  - Is the whole project team involved?
  - (When) will the DMP be evaluated and updated?
Common themes to address

• Data description

• Standards and metadata

• Data sharing, accessibility, licensing

• Archiving and preservation

• FAIR – but may be interwoven with the above

Based on Sarah Jones: Developing and reviewing Data Management Plans. DCC-DANS training, Amsterdam, 2017
Data description

• Is there a full description of the data that will 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?

• Is there a clear view on who may use the data?

Based on Sarah Jones: [Developing and reviewing Data Management Plans](https://example.com). DCC-DANS training, Amsterdam, 2017
Standards and metadata

• Are metadata standards being used?
  • If not, what alternative does the plan describe?

• Will sufficient metadata and documentation be provided to allow others to find, understand and reuse the data?
  • Persistent identifier specified?

• Is the choice of file format explained? Is it clear that appropriate decisions have been made?

Based on Sarah Jones: Developing and reviewing Data Management Plans. DCC-DANS training, Amsterdam, 2017
Data sharing, accessibility, licensing

- 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?
  - Is it clear that “sharing” means “outside the consortium”?

- How will the data be shared? E.g. through deposit in repository?
  - Will usage licences be attached?
  - Will there be an explicit process to apply for access to restricted data (if any)?

- If an embargo period is planned, is that in line with norms for that discipline?

Based on Sarah Jones: Developing and reviewing Data Management Plans. DCC-DANS training, Amsterdam, 2017
Archiving and preservation

- Will the research data be deposited in a suitable community database, repository or archive?
  - Certified as Trustworthy Data Repository?

- Is it clear which data should be preserved and for how long?
  - Criteria available?

- Are there any costs associated with preservation, and if so, how will these be covered?

Based on Sarah Jones: Developing and reviewing Data Management Plans. DCC-DANS training, Amsterdam, 2017
Feedback and further suggestions?

More sample texts in [this webinar](#)
Public DMPs (for different funders) at [DCC](#) or in [RIO](#)

Acknowledgements: thanks to all DMP authors and the DCC

[marjan.grootveld@dans.knaw.nl](mailto:marjan.grootveld@dans.knaw.nl) and [ellen.leenarts@dans.knaw.nl](mailto:ellen.leenarts@dans.knaw.nl)
Will you be the winner of the Dutch Data Prize 2018?

Have you made research data accessible for re-use? Or do you know someone who has?

Nominate yourself, another researcher or a research group for the Dutch Data Prize 2018!

Go to researchdata.nl to nominate. Deadline July 1st.

The prize:
• a sculpture
• € 5,000 to make the dataset (even more) accessible.