Data Management Plans:
what, why, who, when & how?

Maggie Hellström

ICOS Carbon Portal & Lund university

DATA FAIRNESS
LECCE 1-5 JUL 2019 INTERNATIONAL SUMMER SCHOOL FOR ENVIRONMENTAL & EARTH SCIENCE INFRASTRUCTURES

Background to DMPs

- Initial reasons to ask for data management plans were mainly related to the need to estimate usage of technical resources, such as computations and storage
- eventually, it was realized that DMPs could also help researchers plan the data activities throughout a project, including allocating responsibilities
- increasingly, funders and policy makers saw possibilities to use DMPs as a way to foster best practices (including FAIR)
- the responsibility of checking plan content has, however, not always been clear, leading to copy & paste behavior
- DMPs are increasingly being seen as dynamic, "living" documents
- machine readability & interpretability is now seen as very important, enabling cross-linking with unis and funders

DMP flavours: what to cover

- Data types, volumes, formats, organization
- Need for e-infrastructure resources
- Allocation of responsibility for important tasks
- Plans for long-term sustainable storage
- Legal & ethical issues (if applicable)

- Discussion on how to ensure highest realistically possible degree of FAIRness

Tools & templates

There are a number of DMP templates, and tools to help fill these out:

- Data Steward Wizard tool (https://ds-wizard.org/)
- Digital Curation Centre's DMP-Online tool & template (https://dmponline.dcc.ac.uk/)
- Science Europe's "core requirements" template (https://www.scienceeurope.org/)
- European Commission's DMP for Horizon2020 template Guidelines on FAIR Data Management in Horizon 2020 (available for DMPtool and DSW)
- …

Science Europe: concise

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. Data description and collection or re-use of existing data

a. How will new data be collected or produced and/or how will existing data be re-used?

b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. Documentation and data quality

a. What metadata and documentation (for example the methodology of data collection and way of organizing data) will accompany data?

b. What data quality control measures will be used?

Science Europe: concise

3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

c. How will possible ethical issues be taken into account, and codes of conduct followed?

Science Europe: concise

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR?

European Commission: detailed

- The EC DMP approach is multi-step & dynamic:
  - start simple when applying,
  - add more detail during first year of approved project,
  - revisit regularly,
  - make a final updated version at the end
- Ensuring FAIRness is given extra weight

European Commission: detailed

In the following slides, the questions in the simplified (initial stage) DMP template are listed:

A-1. Provide a summary of the data addressing the following issues:

- State the purpose of the data collection/generation
- Explain the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- State the expected size of the data (if known)
- Outline the data utility: to whom will it be useful

European Commission: detailed

B-1. Making data findable, including provisions for metadata:

- Outline the discoverability of data (metadata provision)
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used
- Outline the approach towards search keyword
- Outline the approach for clear versioning
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

European Commission: detailed

B-2. Making data openly accessible:

- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions

European Commission: detailed

B-3. Making data interoperable:

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

B-4. Increase data re-use (through clarifying licenses):

- Specify how the data will be licenced to permit the widest reuse possible
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
- Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
- Describe data quality assurance processes
- Specify the length of time for which the data will remain re-usable

European Commission: detailed

C-1. Explain the allocation of resources, addressing the following issues:

- Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
- Clearly identify responsibilities for data management in your project
- Describe costs and potential value of long term preservation

D-1. Address data recovery as well as secure storage and transfer of sensitive data

E-1. To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

F-1. Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Try it yourself!

Split up into groups of 3-4 people and fill out a Science Europe DMP for one of your ENVRIs (45 mins)!

- Link to form: https://forms.gle/Ws7J4PLWtGYbPLzz9
- Concentrate on sections 2, 3 and 5
- Put your names
- Remember to go all the way to the end and submit when you are done!

At the end, the groups will share their inputs & experiences (30 mins).