Science Europe Consultation on Research Data Management

Consultation available until 30 April 2018 at
http://scieur.org/rdm-consultation

Introduction

Science Europe and the Netherlands Organisation for Scientific Research (NWO) have launched an initiative for the voluntary international alignment of research data management (RDM) policies and practices.

The aim of the initiative is to develop a set of core requirements for research data management plans and a list of criteria to select trusted repositories where researchers can safely store their data for sharing. Science Europe and NWO wish to work towards the adoption of these requirements by as many stakeholders as possible and towards the alignment of their current rules.

About the Consultation

This consultation constitutes the next step towards the development of core RDM requirements. Interested organisations are invited to provide their feedback on draft proposed requirements for data management plans (DMPs) and criteria for trusted repositories (see invitation letter).

The consultation is structured as follows:

A. Information on the responding organisation and contact person
B. Core requirements for Data Management Plans
C. Criteria for Trusted Repositories

The consultation will run until 30 April 2018. In order to ease the preparation of the responses, the full list of questions is available in downloadable format here. However only responses submitted online by the deadline will be considered.

Should you have any questions, please contact the responsible person at the Science Europe Office:
marie.timmermann@scienceeurope.org

A. Information on responding organisation and contact person

Responding Organisation

1. The name of your organisation

2. Type of organisation
Choose one of the following answers

☐ Research Funding Organisation
☐ Research Performing Organisation
☐ University

☐ Learned Society or Scholarly/Researcher Association
☐ Research Infrastructure
☐ e-Infrastructure
☐ Scientific Publisher
☐ Data Repository/Curation Centre
☐ Other: _____

3. Does your organisation currently have an RDM policy in place?
Choose one of the following answers

☐ An RDM policy is in place
☐ An RDM policy is under development
☐ No RDM policy is in place
☐ A Data Management Plan (DMP) template is used, but no fully-fledged RDM policy is in place
☐ Other: _____

Contact Person

4. Your details

Your name _____
Your function within your organisation _____
Your email address _____
Your telephone number _____

5. Do you agree to be contacted by one of our experts in case we need more information regarding the feedback you are giving on behalf of your organisation?

☐ Yes ☐ No

B. Core requirements for Data Management Plans

The planning of data management from the beginning of the research process safeguards the full use of the data during the research and is essential for publishing, sharing and archiving data at the end of the research process. A data management plan (DMP) is a formal document that outlines how to handle data both during a research project and after the project is completed.

DMPs reduce the risk of losing data, help to anticipate complex ownership and user right issues in advance and improve future collaborations when data has been made openly accessible. There is a general acceptance that data management practices should follow the FAIR principles, meaning that data should be Findable, Accessible, Interoperable, and Re-usable.

Despite the general acceptance of these principles, there is globally a great variety in DMP templates and guidelines, which can be confusing for researchers. Thus, it is desirable to identify a core set of DMP items to facilitate increasingly multidisciplinary and globally collaborative research. There is an obvious need to have minimum standards for core DMP templates. The core set of DMP items need to be applicable to all fields of research. Also, guiding questions should be identified to ensure that key elements will be addressed in data management planning. Researchers do not have

to consider each topic in detail. They may refer to texts made available by research funding organisation, research performing organisations or to domain protocols broadly accepted by researchers in their field.

6. Does your organisation agree that there is a general need to align DMP requirements across national and European organisations?

☐ Yes ☐ No

6a. Why does your organisation not agree?

7. Would your organisation join others in supporting and implementing broadly accepted core principles?

☐ Yes ☐ No

7a. What timeframe do you expect to be needed to implement the requirements in your organisation's policy?

7b. What is the major reason or impediment for your organisation not to implement the requirements, and how could this be addressed?

8. Do you have general comments on the approach?

Science Europe Survey

The planning of data management from the beginning of the research process safeguards the full use of the data during the research and is essential for publishing, sharing and archiving data at the end of the research process. A data management plan (DMP) is a formal document that outlines how to handle data both during a research project and after the project is completed.

The Science Europe Working Group on Research Data conducted a survey of European Research Funding and Research Performing Organisations on data policies and data management between November 2017 and February 2018. Main topics covered by this survey were: research data policy coverage and stipulations; data management planning items; and evaluation of DMPs at RFOs or quality control at RPOs. Answers to the survey came from 18 countries and 22 Science Europe Member Organisations. The survey gave an overview of which DMP items were common in these organisations. In addition, a comparison was made of widely used templates and/or guidelines:

- Generic DMP Template of the Digital Curation Centre (DCC)
- Guidelines on FAIR Data Management in Horizon 2020
- Open Research Data and DMP requirements by the European Research Council
- The FAIR Guiding Principles for scientific data management
- Policy on data sharing by the Wellcome Trust
- The Finnish national DMP
- DMP items and guidelines by the Research Data Alliance (RDA) working groups and Force11

Taken together, it was possible to draft a core set of DMP items for a template consisting of five broad topics to be covered by a DMP together covering no more than 14 requirements/guiding questions. This is the minimum set of requirements that Science Europe is seeking consensus about. Indicative lists of related aspects to consider when developing a DMP have been added as well.

The proposed five core DMP items presented here are already widely used and provide the major aspects needed to build a useful DMP:

1. Data description and collection or reuse of existing data
2. Documentation and data quality
3. Storage and backup
4. Ethical and legal compliance, codes of conduct
5. Data sharing and long-term preservation

Science Europe welcomes your organisation's feedback on the core requirements as such, as well as on the 'Guiding questions'. The 'Related aspects to consider when developing a DMP (indicative list)' should be seen as a non-exhaustive list of aspects, supposed to serve as an example.

Core Requirement 1: Data description and collection or reuse of existing data

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
A. What is the type, format and volume of data?	• Explain what kind of data you are collecting or producing. • Explain what kind of existing data you will reuse. • What type, format and volume of data? Clearly note what format(s) your data will be in, eg. plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw). • Explain why you have chosen certain formats. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. • Do your chosen formats and software enable sharing and long-term access to the data? By using standardised, interchangeable or open formats ensures the long-term usability of data; these are recommended for sharing and archiving.

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
B. How will data be collected, created or reused?	• Are there any existing data that you can reuse? • Outline how the data will be collected and processed: eg. via surveys, interviews, laboratory experiments, or observations. This should cover relevant standards or methods, quality assurance and data organisation. • Briefly list the methodologies for data collection/generation or description of the existing data that will be used.

9. Does your organisation agree in general with the requirement 'Data description and collection or reuse of existing data' and the accompanying 'guiding questions' as described above?

☐ Yes ☐ No

9a. Why does your organisation not agree with the requirement as described above?

10. Are there any aspects missing?

☐ Yes ☐ No

10a. What aspects are missing?

11. Are there aspects that need more elaboration in the guiding questions?

☐ Yes ☐ No

11a. Which aspects need more elaboration in the guiding questions?

12. Are there aspects that are unnecessary?

☐ Yes ☐ No

12a. Which aspects are unnecessary and why?

Core requirement 2: Documentation and data quality

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
A. What metadata and documentation will accompany data?	• What metadata will be provided to help others identify and discover the data? • What standards or methodologies will you use? • Researchers are strongly encouraged to use community metadata standards where these are in place. • Your data and metadata vocabularies, standards or methodologies should facilitate interoperability (also inter-disciplinary interoperability) • Indicate how the data will be organised during the project, mentioning, eg. naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse. • Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data. • Consider how you will capture this information and where it will be recorded, eg. in a database with links to each item, in a 'readme' text file, in file headers, and so on.
B. Will you make sure unique and persistent identifier is in use (eg. DOI)?	• How might your data be reused in other contexts? Unique persistent identifiers should be applied so people can reliably and efficiently find your data. They also help you to track citations and reuse. • Will you pursue getting a unique persistent identifier for your data?
C. What data quality control measures do you use?	• What quality assurance processes will you adopt? • Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.

13. Does your organisation agree in general with the requirement 'Documentation and data quality' and the accompanying 'guiding questions' as described above?

☐ Yes ☐ No

13a. Why does your organisation not agree with the requirement as described above?

14. Are there any aspects missing?

☐ Yes ☐ No

14a. What aspects are missing?

15. Are there aspects that need more elaboration in the guiding questions?

☐ Yes ☐ No

15a. Which aspects need more elaboration in the guiding questions?

16. Are there aspects that are unnecessary?

☐ Yes ☐ No

16a. Which aspects are unnecessary and why?

Core requirement 3: Storage and backup

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
A. How will data be stored and backed up during the research?	• Describe where the data will be stored and backed up during the course of research activities. Also, how often the backup will be performed. • The use of robust, managed storage with automatic backup, for example, that provided by university IT teams, is preferable. Storing data on laptops, computer hard drives or external storage devices alone is very risky.
B. How will you take care of data security and personal data protection during the research?	• How will the data be recovered in the event of an incident? • Consider data security, particularly if your data is sensitive eg. detailed personal data, politically sensitive information, or trade secrets. Note the main risks and how these will be managed. • Take into consideration the EU General Data Protection Regulation (GDPR) on data privacy. • Are there any institutional data security policies in place?

17. Does your organisation agree in general with the requirement 'Storage and backup' and the accompanying 'guiding questions' as described above?

☐ Yes ☐ No

17a. Why does your organisation not agree with the requirement as described above?

18. Are there any aspects missing?

☐ Yes ☐ No

18a. What aspects are missing?

19. Are there aspects that need more elaboration in the guiding questions?

☐ Yes ☐ No

19a. Which aspects need more elaboration in the guiding questions?

20. Are there aspects that are unnecessary?

☐ Yes ☐ No

20a. Which aspects are unnecessary and why?

Core requirement 4: Ethical and legal compliance, codes of conduct

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
A. How will you manage ethical issues and codes of conduct?	• Have you gained consent for data preservation and sharing? • Follow the national and international codes of conducts, institutional ethical guidelines and check if ethical review is required for your research project. • How will sensitive data be handled to ensure it is stored and transferred securely? Investigators carrying out research involving human participants should request consent to preserve and share the data. • Consider how you will protect the identity of participants, eg. via anonymisation or using managed access procedures. • Ethical issues may affect how you store and transfer data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.
B. How will you manage IPR, copyright, ownership and other legal issues?	• Are there any restrictions on the reuse of third-party data? • State who will own the copyright and IPR of any existing data as well as new data that you will generate. For multi-partner projects, IPR ownership should be covered in the consortium agreement. Data sharing agreements may also be needed. • Explain how the data will be licensed for reuse. • Take in to consideration the EU General Data Protection Regulation (GDPR) on data privacy.

21. Does your organisation agree in general with the requirement 'Storage and backup' and the accompanying 'guiding questions' as described above?

☐ Yes ☐ No

21a. Why does your organisation not agree with the requirement as described above?

22. Are there any aspects missing?

☐ Yes ☐ No

22a. What aspects are missing?

23. Are there aspects that need more elaboration in the guiding questions?

☐ Yes ☐ No

23a. Which aspects need more elaboration in the guiding questions?

24. Are there aspects that are unnecessary?

☐ Yes ☐ No

24a. Which aspects are unnecessary and why?

Core requirement 5: Data sharing and long-term preservation

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
A. How and when will you share data (consider licences, data security/ protection, possible embargo reasons)?	• How will you share the data eg. deposit in a trustworthy data repository, use a secure data service, handle data requests directly or use another mechanism? The methods used will depend on number of factors such as the type, size, complexity and sensitivity of the data. • Outline the plans for data sharing and preservation – how long will the data be retained? • When will you make the data available? What is the expected timely release. For how long do you need exclusive use of the data and why? Will data sharing be postponed/restricted, eg. to publish, protect proprietary or seek patents? • How will potential users find out about your data? • Who will be able to use your data? If you need to restrict access to certain communities or apply data sharing agreements, explain why. What action will you take to overcome or minimise restrictions? • Will a data sharing agreement (or equivalent) be required?
B. How do you select data for preservation and where data will be preserved long-term (eg. data repository or archive)?	• What data must be retained/destroyed for contractual, legal, or regulatory purposes? • How will you decide what data to keep? • Describe the data to be preserved long-term. • What are the foreseeable research uses for the data? • Where will the data be deposited? If you do not propose to use an established repository, the data management plan should demonstrate that the data can be curated effectively beyond the lifetime of the grant. It helps to show that you have consulted with the repository to understand their policies and procedures, including any metadata standards, and costs involved.

Guiding questions	Related aspects to consider when developing a DMP (indicative list)
C. What methods or software tools are needed to access data?	• How will potential users can access your data? • Will you share data via a repository, handle requests directly or use another mechanism? • Carefully consider the sustainability of software needed for accessing the data.
D. Who will be responsible for data management (ie. data steward)?	• Outline the roles and responsibilities for all activities, eg. data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Individual(s) should be named where possible. • For collaborative projects, you should explain the co-ordination of data management responsibilities across partners. • Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?
E. What are the costs and time needed for data management and making data FAIR?	• Will additional resources be needed to prepare data for deposit or meet any charges from data repositories? • Have you costed in time and effort to prepare the data for sharing/preservation (data curation)? Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit and repository charges. • What costs if any will your selected data repository or archive charge?

25. Does your organisation agree in general with the requirement 'Storage and backup' and the accompanying 'guiding questions' as described above?

☐ Yes ☐ No

25a. Why does your organisation not agree with the requirement as described above?

26. Are there any aspects missing?

☐ Yes ☐ No

26a. What aspects are missing?

27. Are there aspects that need more elaboration in the guiding questions?

☐ Yes ☐ No

27a. Which aspects need more elaboration in the guiding questions?

28. Are there aspects that are unnecessary?

☐ Yes ☐ No

28a. Which aspects are unnecessary and why?

Regarding Core requirements 1–5

29. Are there aspects that would seriously refrain your organisation from adopting these requirements?

☐ Yes ☐ No

29a. Which aspects are these, and why would they refrain your organisation from adopting the requirements?

C. Criteria for trusted repositories

There is no generally accepted list of trusted repositories, whereas registries of repositories mention more than 2,000 of them. This is confusing not only for researchers but also for funders, institutions and journals that seek long term preservation of research data. Certification organisations and processes do exist to ensure that repositories meet minimum standards of quality and reliability. However, currently less than 10% of listed repositories worldwide comply with these certifications.

In due time repositories should account for the quality and trustworthiness of the services they offer; preferably by attaining some form of certification.

For the time being, minimal requirements for trusted repositories are proposed here to ensure that repositories meet minimum standards of quality and reliability. The proposed set of requirements are based on a comparison of criteria for suitable platforms for research data sharing and archiving for different disciplines, set by a selection of sources (funders, performers, certification bodies, as well as publishers).

Minimal requirements for trusted repositories

A trusted repository should have the following requirements:

1. **Provision of a persistent and unique identifier (PID)**
   - identify the dataset
   - ensure dataset persistence
   - enable searching and retrieval of datasets
   - maintain a repository-managed URI associated with each of those PIDs
   - keep permanent IDs as "tombstones" even if the data have been retracted

2. **Metadata**
   - ensure dataset persistence
   - enable finding of datasets
   - provide publicly available and maintained information even for retracted datasets

3. **Data access & Usage licenses**
   - enable access to the dataset under well-specified conditions
   - ensure dataset stability
   - enable retrieval of datasets
   - provide information about licensing and permissions

4. **Machine Accessibility**
   - enable searching (and preferably retrieval) of datasets by automated processes
   - ensure that at least intrinsic metadata is accessible in a structured and machine- readable form

5. **Long-term Preservation**
   - ensure persistence of metadata and datasets
   - explain the long-term preservation policies and plans
   - guarantee the sustainability of a repository

The above requirements are very well in line with criteria of existing certification systems.

30. Does your organisation agree in general with the criteria as described above?

☐ Yes ☐ No

30a. Why does your organisation not agree with the criteria as described above?

31. Are there any aspects missing?

☐ Yes ☐ No

31a. What aspects are missing?

32. Are there aspects that need more elaboration?

☐ Yes ☐ No

32a. Which aspects need more elaboration?

33. Are there aspects that are unnecessary?

☐ Yes ☐ No

33a. Which aspects are unnecessary and why?

34. Do you have any comments or suggestions on the proposed RDM requirements or criteria for trusted repositories in general?