

Guidance on the Evaluation
of Data Management Plans

TEMPLATE FROM THE SCIENCE EUROPE PRACTICAL GUIDE TO

THE INTERNATIONAL ALIGNMENT OF
RESEARCH DATA MANAGEMENT

# Introduction

## 1. What is the DMP evaluation guidance for?

When research (funding) organisations require researchers to develop DMPs they must be able to also follow up with evaluation and feedback for the researchers. This guidance is designed to support and guide the evaluation of DMPs, prompting the analysis of whether all required aspects have been covered. It is drafted in a generic way and deliberately written in plain language. Completely aligned with the guidance for researchers, the DMP evaluation rubric also aims to ensure the ‘FAIRness’ of data, even though this is not explicitly stated in all sections of the rubric.

## 2. Why this format?

This DMP evaluation guidance takes up the requirements and guiding questions from previous chapters of this guide. It provides criteria to help the reviewer assess whether the information provided in the DMP is sufficient to ensure that the research team will manage data as expected. It is presented in the form of a rubric and lists the different criteria and performance levels that indicate to what extent the criteria are met. This rubric contains two performance levels: ‘Sufficiently Addressed’ and ‘Insufficiently Addressed’. Insufficiently addressed refers either to a lack of information or to information deemed incorrect.

## 3. Can it be adapted?

Following the same structure as the previous chapters, the rubric provides the core criteria for analysis while leaving flexibility for organisations to adapt it to legislative or institutional circumstances and to disciplinary requirements. The core requirements and evaluation criteria in this guide can and should be adapted to take into account existing institutional or disciplinary policies and practices. Where these exist, they should be referred to, both in the guidance for researchers and for DMP reviewers. It is therefore important to stress that this rubric must be seen as guidance, not as a ready-made tool (such as a checklist). Organisations that want to develop checklists for the evaluation of DMPs can use the rubric as a framework and adapt it accordingly.

## 4. Who is it for?

The guidance will be helpful for anyone who is called to evaluate a DMP. This includes research officers, reviewers, or institutional data managers. Researchers will also find it useful as an additional source of information. Reviewers using the rubric are strongly encouraged not to use it as a tick-box exercise, but instead to use it to capture their comments and ratings. Sharing these comments with the researchers will provide additional support and clarification so that they can improve their DMPs.

## 5. When can it be used?

DMPs are reviewed at different stages of the research project lifecycle, depending on institutional policies. The rubric is designed to work alongside this process and can be used each time a DMP is reviewed. Reviewers must keep in mind that a DMP is a living document. The level of detail provided in a DMP might vary depending on which version is being assessed; for example, the first version included in a funding application, or a later version documenting the deposition in a repository.

# Evaluation Rubric for Data Management Plans

|  |  |  |
| --- | --- | --- |
| DMP Question | DMP Guidance | Performance Level |
| Guidance for Researchers | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| Administrative information | * Provide information such as name of applicant, project number, funding programme, version of DMP.
 | * contains the minimal information required to identify the applicant and the references of the project.
 | * provides no or limited information, which makes it hard to identify who is responsible for the project.
 |
| **1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **1a** How will new data be collected or produced and/or how will existing data be re-used? | * Explain which methodologies or software will be used if new data are collected or produced.
* State any constraints on re-use of existing data if there are any.
* Explain how data provenance will be documented.
* Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.
 | * gives clear details of where the existing data come from and how new data will be collected or produced. It clearly explains methods and software used.
* explains, if existing data are re-used, how these data will be accessed and any constraints on their re-use.
* explains clearly, if applicable, why new data must be collected, rather than re-using existing data.
 | * provides little or no details on where the data come from and what data will be collected or re-used.
* does not, if applicable, provide sufficient rationale for generating new data.
 |
| **1b** What data (for example the kind, formats, and volumes) will be collected or produced? | * Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.
* Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).
* Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.
* Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such ‘preferred formats’).
* Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).
 | * clearly describes or lists what data types will be generated (for example numeric, textual, audio, or video) and their associated data formats, including, if needed, data conversion strategies.
* explains why certain formats have been chosen and indicates if they are in open and standard format. If a proprietary format is used, it explains why.
* provides information about the estimated data volume.
* clearly states, if applicable, that no new data will be produced or generated by the project.

NB. Information derived from previously existing data sources (namely output, processed, and analysed data) are to be considered new data under this question. | * provides no or little details on what data types will be generated and does not provide a valid reason for this omission (for example a statement that no data will be produced or generated).
* only lists/describes the kinds of data without specifying their formats.
* only lists formats, without specifying the kinds of data.
* does not provide an estimate of data volume.
 |
| **2 DOCUMENTATION AND DATA QUALITY** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **2a** What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data? | * Indicate which metadata will be provided to help others identify and discover the data.
* Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
* Use community metadata standards where these are in place.
* Indicate how the data will be organised during the project mentioning, for example, conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
* Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.
* Consider how this information will be captured and where it will be recorded (for example in a database with links to each item, a ‘readme’ text file, file headers, code books, or lab notebooks).
 | * clearly outlines the metadata that will accompany the data, with reference to good practice in the community (for example uses metadata standards where they exist).
* clearly outlines the documentation needed to enable data re-use, stating where the information will be recorded (for example a database with links to each item, a ‘readme’ text file, file headers, code books, or lab notebooks).
* indicates how the data will be organised during the project (for example naming conventions, version control strategy and folder structures).
 | * provides little or no details on the metadata that will accompany the data.
* provides no information, or only a very vague mention of documentation, without providing any detail or explanation.
 |
| **2b** What data quality control measures will be used? | * Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.
 | * clearly describes the approach taken to ensure and document quality control in the collection of data during the lifetime of the project.
 | * provides no information or only a vague mention on how data quality is controlled and documented during the lifetime of the project.
 |
| **3 STORAGE AND BACKUP DURING THE RESEARCH PROCESS** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **3a** How will data and metadata be stored and backed up during the research? | * Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.
* Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.
 | * clearly (even if briefly) describes:
* the location where the data and backups will be stored during the research activities.
* how often backups will be performed.
* the use of robust, managed storage with automatic backup (for example storage provided by the home institution).

or* explains why institutional storage will not be used (and for what part of the data) and describes the (additional) locations, storage media, and procedures that will be used for storing and backing up data during the project.
 | * provides no information or very vague reference to how data will be stored and backed up during the project.
 |
| **3b** How will data security and protection of sensitive data be taken care of during the research? | * Explain how the data will be recovered in the event of an incident.
* Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.
* Consider data protection, particularly if your data is sensitive (for example containing personal data, politically sensitive information, or trade secrets). Describe the main risks and how these will be managed.
* Explain which institutional data protection policies are in place.
 | * clearly explains
* how the data will be recovered in the event of an incident.
* which institutional and/or national data protection policies are in place and provides a link to where they can be accessed.
* who will have access to the data during the research.
* clearly describes the additional security measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe, when sensitive data are involved (for example personal data, politically sensitive information, or trade secrets).
 | * provides little or no details on how the data will be recovered in the event of an incident, which institutional data protection policies are in place, and who will have access to the data during the research.
* provides little or no details about data protection and risk management, or the explanation is too vague, when sensitive data are involved (for example personal data, politically sensitive information, or trade secrets).
 |
| **4 LEGAL AND ETHICAL REQUIREMENTS, CODES OF CONDUCT** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **4a** If personal data are processed, how will compliance with legislation on personal data and security be ensured? | * Ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with:
* Gain informed consent for preservation and/or sharing of personal data.
* Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).
* Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible).
* Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party).
* Explain whether there is a managed access procedure in place for authorised users of personal data.
 | * clearly indicates if personal data will be collected/used as part of the project, and, if applicable, how compliance with applicable legislation will be ensured (for example by gaining informed consent, considering encryption, anonymisation, or pseudonymisation).
* describes the procedure to manage access to only authorised users.
 | * provides little or no details to demonstrate that personal data, if any, will be managed in compliance with applicable legislation.
 |
| **4b** How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable? | * Explain who will be the owner of the data, meaning who will have the rights to control access:
* Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses.
* Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement.
* Indicate whether intellectual property rights (for example Database Directive, sui generis rights) are affected. If so, explain which and how will they be dealt with.
* Indicate whether there are any restrictions on the re-use of third-party data.
 | * clearly explains, if applicable,
* who will have the rights to control access to which part of the data.
* what access conditions and re-use licenses will apply to the data.
* clearly explains, if applicable, how intellectual property rights will be managed.
* explains for multi-partner projects and multiple data owners how these matters are addressed in the consortium agreement.
* Alternatively, there is a clear statement that there are no such restrictions on the data.
* indicates, if applicable, whether there are any restrictions on the re-use of third-party data.
 | * does not address legal issues (or only for a subset of the data), and does not provide good reason or explanation for not doing so.
* does not address matters of rights to control access to the data in case of a multi-partner project and does not provide good reason or explanation for not doing so.
 |
| **4c** What ethical issues and codes of conduct are there, and how will they be taken into account? | * Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning.
* Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.
 | * provides details of what ethical issues have been considered that may affect data storage, transfer, use, sharing and/or preservation, and demonstrates that adequate measures are in place to manage ethical requirements.
* mentions, if applicable, whether ethical review is being pursued. If ethical approval has been obtained, refers to the relevant committee and documents.
* refers to relevant ethical guidelines and/or codes of conduct or alternatively provides a clear statement that explains why ethical issues have not been considered.
 | * provides little or no details to demonstrate that ethical implications and codes of conduct have been considered, and does not explain why they did not need to be considered.
 |
| **5 DATA SHARING AND LONG-TERM PRESERVATION** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **5a** How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons? | * Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism).
* Outline the plan for data preservation and give information on how long the data will be retained.
* Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents.
* Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions.
 | * clearly describes how the data and/or metadata will be made discoverable and shared.
* specifies when data will be shared and under which licence.
* includes the name of the repository, data catalogue, or registry where data will or could be shared.
* includes information on how long the data will be retained and gives precision on its timely release.
* clearly explains, if applicable, why data sharing is limited or not possible, and who can access the data under which conditions (for example, only members of certain communities or via a sharing agreement).
* explains, where possible, what actions will be taken to overcome or to minimise data sharing restrictions.
 | * provides little or no details on how and when data will be shared, or the explanation is not adequate or technically viable.
 |
| **5b** How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)? | * Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes.
* Indicate how it will be decided what data to keep. Describe the data to be preserved long-term.
* Explain the foreseeable research uses (and/ or users) for the data.
* Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.
 | * provides details of what data collected or created in the project will be preserved in the long term and clearly indicates for how long. This should be in alignment with funder, institutional, or national policies and/or legislation, or community standards.
* provides details of which (versions of) data and accompanying documentation will be retained or destroyed, and explains the rationale (for example contractual, legal requirements, or regulatory purposes).
* provides details of how the selection is made, and what possible interest there would be for re-use (or not).
* provides details on how the data, accompanying documentation, and any other required technology such as copies of software in specific versions will be archived in the long term.
* explains how data will be managed in a sustainable way beyond the lifetime of the grant.
* provides the name of the archive or trustworthy repository – or the way to curate and preserve data – that will be used to make data available for re-use.
 | * provides no further information or lacks adequate explanation on what provisions would be made for data preservation.
 |
| **5c** What methods or software tools are needed to access and use data? | * Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data.
* Indicate whether data will be shared via a repository requests handled directly, or whether another mechanism will be used?
 | * clearly indicates which specific tools or software (for example specific scripts, codes, or algorithms developed during the project, version of the software) potential users may need to access, interpret, and (re-)use the data.
* provides information, if relevant, on any protocol to access the data (for example if authentication is needed or if there is a data access request procedure).
 | * provides little or no details on which software developed during the project will be necessary to access and interpret the data, how it will be made available, or why that may not be possible.
 |
| **5d** How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured? | * Explain how the data might be re-used in other contexts. Persistent identifiers (PIDs) should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and re-use.
* Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.
 | * specifies how the data can be re-used in other contexts.
* clearly indicates if and which persistent identifiers (PIDs) are provided for all datasets, individual datasets, data collections, or subsets. If PIDs will not be used, it explains why.
* clearly presents the approach, and the choice of identifiers is justified and refers to international standards.
 | * makes no mention of PIDs nor provides a valid reason for not providing them.
* provides no clear information on what type of PID will be assigned to the data and whether individual datasets and/or collections or datasets will be issued with PIDs.
 |
| **6 DATA MANAGEMENT RESPONSIBILITIES AND RESOURCES** |
| **Guidance for Researchers** | Sufficiently AddressedThe DMP… | Insufficiently AddressedThe DMP… |
| **6a** Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)? | * Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible.
* For collaborative projects, explain the co-ordination of data management responsibilities across partners
* Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.
* Consider regular updates of the DMP.
 | * clearly outlines the roles and responsibilities for data management/stewardship (for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing), naming responsible individual(s) where possible.
* clearly indicates who is responsible for day-to-day implementation and adjustments to the DMP.
* explains, for collaborative projects, the co-ordination of data management responsibilities across partners.
 | * does not discuss responsibility for data management/stewardship activities and/or does not indicate who is responsible for day-to-day implementation and adjustments to the DMP.
* provides no description, in case of a collaborative project, on how data management responsibilities will be co-ordinated across partners.
 |
| **6b** What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)? | * Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in.
* 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.
* Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.
 | * provides clear estimates of the resources and costs (for example storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges) that will be dedicated to data management and ensuring that data will be FAIR and describes how these costs will be covered. Alternatively, there is a statement that no additional resources are needed.
 | * provides no answer or is vague about the resources required for data management and ensuring that data will be FAIR (for example resources are not listed or costed inappropriately), and/or does not describe how the costs will be covered.
 |