Health Research Board (HRB) Ireland: Health Research Board DMP Template

Data description and collection or re-use of existing data

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

Guidance:

Explain which methodologies or software will be used if new data are collected or produced and specify which community standards (if any) will be used.

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.

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

Guidance:

Give details on the types of data – quantitative, qualitative; generated from surveys, interviews, medical records, clinical measurements, tissue samples, genotypic data, etc.

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).

Consider and detail which data will have value to other research users and could be shared. We recognize that there are many reasons why data cannot be shared or made openly available. If you do not intend to make the data you’ve generated available to others, please provide justification for your decision.

Documentation and data quality

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

Guidance:

Guidance

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 and potential community standards available.

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.

Consider what other documentation is needed to enable re-use - 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.

When describing data, please remember that file and folder names as well as variables and metadata may contain personal or sensitive data. Even if your research data contains personal data, related metadata can be published if it does not contain identifiers which could be used to identify a study subject.

What data quality control measures will be used?

Guidance:

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.
Consider how data minimisation, pseudonymisation or anonymisation will affect data quality.

Storage and backup during the research process

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

Guidance:

Describe how and where the data will be stored, backed-up and managed 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 should be avoided. If external servers are used please ensure that they are compliant with GDPR and any other legislation related to the data collected.

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

Guidance:

Detail the key risks to the confidentiality and security related to human participants or other sensitive data and how this information will be managed.

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.

Explain which institutional data protection policies are in place.

Legal and ethical requirements, codes of conduct

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

Guidance:

Ensure that when dealing with personal data protection laws (for example GDPR and the Health Research Regulations) are complied with:

- Ensure that the preservation and/or sharing of personal data is fully consistent with the terms of the informed consent under which the data were provided by participants.
- 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 or governed access procedure in place for authorised users of personal data.

If you are unclear on how best to comply with legislation on personal data and security, please speak to your institutional Data Protection Officer.

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

Guidance:

Explain who will be the owner of the data and 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 (eg. CC-BY, CC-BY-NC, etc.)
- 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.

What ethical issues and codes of conduct are there, and how will they be taken into account?

Guidance:

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.

Data sharing and long-term preservation

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

Guidance:

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 (For data related to clinical trials - specify how long the data will be made available for). 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.

We recognize that there are many reasons why data cannot be shared or made openly available. If you do not intend to make the data you've generated available to others, please provide justification for your decision. Restrictions should be minimized as much as possible.

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

Guidance:

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 and consider how this data will be curated and preserved beyond the lifetime of the grant. Indicate where the data will be deposited, preferably in a trusted repository.

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.

What methods or software tools are needed to access and use data?

Guidance:

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?

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

Guidance:

Explain how the data might be re-used in other contexts. Persistent identifiers 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.

Data management responsibilities and resources

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

Guidance:

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. Alongside the PI, specify who is responsible for ensuring of the completion of these tasks.

Specify who is responsible for the management of sensitive and confidential data as well as monitoring its implementation throughout the lifecycle of the data.

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.

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)?

Guidance:

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