Université de Montpellier: University of Montpellier - DMP template (english) 2

1. DESCRIPTION OF THE GENERATED AND COLLECTED DATA OR REUSE OF EXISTING DATA

1.1 Which data will be collected or generated during the project? Which data will be reused?

Exemple de réponse:
Example of reused data citation: Grabcad , 2021, online database, NIrma : Aadarsh S Chandran, read on the 7/04/2022

Recommandations:

- Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.
- Indicate the data format: .pdf, .xls, .doc, .txt, .rtc, etc. Give preference to open and standard formats as they facilitate sharing and long-term re-use of data.
- [For more information about 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): an estimation in the first version which will be specified in the following versions.

1.2 How will data be generated or collected?

Recommandations:

- Explain which methodologies or software will be used if new data is collected or produced.
- Specify the creation or collection date of the data.
- If material / existing data protected by specific rights or constraints is used during the project, specify it.

2. DATA DOCUMENTATION AND QUALITY CONTROL

2.1 Which metadata and documentation (e.g. data collection and methodology organisation) will accompany the data?

Recommandations:

- The documentation accompanying the data gives users the necessary information to use them and interpret them correctly. A readme file can at least be created to gather the information on the database (name of the source, file format, identifier, content description...).
- [Example of Readme file](#)
- Use community metadata standards where they exist.
- Tools to find metadata standards: https://www.dcc.ac.uk/guidance/standards/metadata and https://rdamsc.bath.ac.uk/
- Present the controlled vocabulary chosen for the indexation. If none has been chosen, explain why (lack of suitable solution for the processed data, etc.) and describe the suitable solution if applicable.
- Tools to find a controlled vocabulary: https://www.lexicool.com/dictionnaires-en-ligne-par-sujet.asp and https://skosmos.loterre.fr/fr/
- Indicate how the data will be organised during the project, mentioning for example naming conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- [For more information about naming conventions](#)

2.2 Which data quality control measures will be implemented?

Recommandations:

- 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.
- Indicate the potential quality control certifications.

3. ETHICAL AND LEGAL REQUIREMENTS

3.1 Which institution will own the intellectual property rights for the data created during the project?

Exemple de réponse:
See article 5 "Intellectual Property" in the consortium agreement of the XXX project (LIRMM).

Recommandations:

- Research data is part of the general context of opening public data by default (Open Data). However, their dissemination can be subjected to conditions, particularly when it comes to personal and sensitive data, copyright or a contract with third parties.
- Explain who will be the owner of the data and which rights they have.
https://opendatacanvas.org/pgd-etape-1-acquisition-1-5

3.2 How will compliance with the provisions of the legislation on personal data be ensured if personal data are processed?

Recommandations:

- 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). https://datascience.etalab.studio/pseudo/
- For any question about the processing of personal data, please contact the [Data Protection Officer (DPO)](#)

Exemple de réponse:

Data protection laws are complied with. Firstly, the participant signs an informed consent form in which it is explained that every personal data will stay strictly confidential (only accessible for researchers that are part of this experiment). Secondly, every data which can be subjected to computer processing will be pseudonymised.

Personal data are stored on paper and archived during 10 years in a secured room inside building XXX. Data are pseudonymised with a unique username for each participant. The link between the pseudonymised data and personal information is only possible accessing paper archives which are never digitalised. Published data do not allow to identify participants, even with cross checking or a new analysis of the data through future methods.

Automatic procedure of pseudonymisation:

Nominative information is only stored on paper, in a locked room. Digitalised data are linked to a unique username for each participant. The unique username is as follows: 000NnPp o 3 digits = unique number for the research participant o Nn = 2 first letter of the surname o Pp = 2 first letters of the name.

3.3 How will ethical issues be considered, and codes of conduct respected?

Recommandations:

- Follow the national and international codes of conduct and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.
- Some research projects require a preliminary opinion from an ethics committee, for example:
    - Committee to protect people for the Jardé Law – 5th of March 2012.
    - Ethics Committee for Animal Experiments for the use of animals in experimentations.
- Some research protocols require preliminary declarations, authorizations or approvals, for example:
    - The Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity.
    - Importation or exportation of human biological samples destined to research.
- Contact for the [UM ethics committee](#)

4. DATA STORAGE AND BACKUP

4.1. How will data and metadata be stored and secured throughout the research project? How will data sharing be ensured between partners?

Exemple de réponse:

The collected data is stored with replication on the Meso@LR infrastructure, whose environment is hosted at CINES, in a Restrictive Regime Zone (access control). The data stored within the Meso@LR infrastructure will be accessible via the NextCloud tool of the xxx laboratory to any user present in the structure's directory. All people associated with the project will possess an account.

Recommandations:

- Describe where the data will be stored and backed up during the research process and how often the backup will be done. It is recommended to store the data in at least two separate locations.
- Focus on safe storage systems, with automatic save, like those given by project institute's informatic services. Data storage on laptops, external hard drives and storage devices as USB keys or commercial clouds (Google Drive, Dropbox, etc...) is not recommended.
- Institutional depository and High-Performance Computing solution: meso-lr@umontpellier.fr
- Explain how data will be recovered in the event of an incident. Explain who will have access to the data during the research process and how access to the data will be controlled, particularly in collaborative research.
- Explain whether ethical issues will affect how data is stored and transferred, who will be able to see or use it, and what conservation periods are applied to it.
- Demonstrate that these aspects are well considered and planned. Explain whether a specific access procedure has been put in place for users authorized to access personal data.

5. DISSEMINATION, SHARING AND REUSE OF DATA

5.1 Which data will be shared? According to which selection requirements?

Recommandations:

- Which requirements are used to select data (and metadata) to be preserved and disseminated?
- During the project, and then between 5 and 10 years after its end, the data to be retained include those allowing the validation of the results and those with evidentiary value or potential for reuse.
- Indicate which data should not be disclosed or should be destroyed for contractual, legal, or regulatory reasons.

5.2 How will data be shared?

Recommandations:

- Explain where and under which conditions the data will be shared. It is recommended to deposit the selected data in a certified or trusted disciplinary repository. Find a depository.
- In the lack of a disciplinary solution, it is possible to deposit the data in the institutional space of the UM within the national multidisciplinary repository Recherche Data Gouv.
- Grant a clear and explicit license to the data, even if it is in restricted access. A list of the licenses that can be used by public administrations in France has been established by decree. It is also possible to use a CC-BY or CC-BY-SA license among Creative Commons.
- Using a perennial identifier is strongly recommended to improve the visibility and accessibility of your dataset. It makes it easier to cite the dataset and link it to a publication or other research output. It makes it easy to cite the dataset and link it to a publication or other research product. Deposit in a data warehouse often results in the automatic assignment of a perennial identifier, including a DOI, but this should better be ensured.
- If a period of exclusivity of access to the data is necessary (for example for the scientific exploitation of the data), indicate its duration (embargo concept: the producer of the data deposits his data in a public repository but these will only be accessible after a defined period).
- Define, at beginning of project, levels of access to your data (public, restricted, embargoed, closed, or a combination).
- Indicate whether potential users will need specific tools for accessing and (re)using data.

Exemple de réponse:

At the end of the project, the data and associated metadata will be deposited in the UM institutional space of the Recherche Data Gouv platform. They will be assigned a DOI, as well as a Share license (CC-0).

6. DATA ARCHIVING AND SUSTAINABILITY

6.1 If some of the data is concerned, how will it be selected and archived?

Exemple de réponse:

Sample of software archived in *Software Heritage* and reported in *HAL*

Recommandations:

- The following questions can be asked in order to select the data:
    - Are they conclusive or indicative of scientific contribution?
    - Is the data reproducible?
    - How much for reproduction?
    - Will they be eligible for further treatment?
    - Can they be used later for other uses?
    - What would be the consequences of data loss?
- In the context of personal data, in the case of additional data processing or in the case of processing operations for purposes different from those for which the data were collected, the persons concerned must be informed.
- To preserve data and ensure its readability in the very long term, permanent archiving can be considered in addition to the deposit in a repository, but it has a cost. The main French archiving service is CINES.
- *CINES* has developed a format validation service to verify that the formats used are sustainable and can be archived.
- For software and source code, the recommended format validation service for data retention is Software Heritage.
- Learn more about archiving