**Data Management Plan PhD Template**

**About your research**

|  |  |
| --- | --- |
| Name and email address |  |
| Thesis director/s |  |
| Working title for the thesis |  |
| Describe your research | *Approximately 50 words* |
| Duration of your research | Start date: *DD-MM-YYYY*  End date: *DD-MM-YYYY* |
| Linked project | *Is this a thesis related to a project? Which one?* |
| Funding | *Have you received funding to complete your dissertation? Whose?* |

**About this data management plan**

|  |  |
| --- | --- |
| Creation date | *DD-MM-YYYY* |
| Last update | *DD-MM-YYYY* |
| Version and date | *Make a new version every time there are significant changes (new datasets, significant changes in your research, or other factors)* |

**1. Data collection**

*Describe the data you will create/collect*

**1.1 Specify the origin/provenance of the data and describe if it is generated or reused**

☐ Your own data or data from the research group in which you participate

☐ Academic collaborators

☐ Commercial collaborators

☐ Publicly available databases/files

☐ Commercial data providers

☐ Others, specify:

|  |
| --- |
|  |

Questions to Consider:

* Will you generate the data or will you collect it from another source?
* Specify which sources
* Explain the standard, methodology or software use~~d~~ to collect or generate new data
* Detail whether the data you have generated or collected can be reused in the future
* If reusing existing data, explain how issues such as copyright and IPR have been addressed [*[More information]](https://bibliotecnica.upc.edu/en/investigadors/dades-recerca-fair" \l "drets-autoria)*

## Data description, type, formats and volume

Give a short description of the data, type and formats and estimated data volume.

* ***Data type:*** for example, if you are going to work with measurements, simulations, observations, text (text, MS Word), images, audio-visuals or samples, statistics (spreadsheets), with computational models, with data from a qualitative survey (questionnaires), recordings (audio, video), software (code), etc [See [*Controlled Vocabularies for Repositories - COAR ]*](https://vocabularies.coar-repositories.org/resource_types/)
* ***Format***: preferably use open standards so that the data can be read by multiple programs,facilitating preservation, and sharing with other users. If you use them, explain why and strategies to convert them. [See recommendation about [*formats]*](https://confluence.csuc.cat/display/RDM/Recomanacions+de+formats)
* **Reuse of existing data / Data Origin:**source from which you captured the data
* ***Data size:*** specify approx. (if you know it) how much space your data will need (bytes, number files, objects, columns, etc.). For example, you can use: (<10GB, 10-30GB, 30-50 GB, 2TB, ...)
* **Data utility:**You can detail it later but it is important to know if the original source allows you to reuse it.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| You can use a table similar to the one below (which contains examples) to describe the type of data in your research   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Data type | Format | Reuse of existing data | Data origin | Estimated size | Data utility (how will be reused?) | | Observational | (.csv); (.xlsx); (.txt); (.doc); (.odt) | No | Interview | <10 GB aprox. |  | | Experimental - Spreadsheets, scripts, image | (.csv); (xlsx); (.mat); (.tiff) | No | Laboratory | <2 GB | Public but license not defined | | Notes Books | (.ipynb) | No | Research |  |  | | Observational | (.doc); (.pdf) | Yes. Public Web | Public Web [name xxx] [URL] | 2 GB |  |     *The data use proprietary formats because [explanation] but when published the dataset it format will be migrates to open standards to f*acilitate *preservation and reuse.* |

## Specify sensitive/personal data

* If you work with [*personal*](https://formaciooberta.eapc.gencat.cat/contingutsdelscursos/abpd/020/inici.html) *or* [*sensitive data*](https://webs.uab.cat/dretsautor/2015/06/22/que-son-les-dades-sensibles/), you have a legal obligation to process it according to applicable regulations ([*GDPR*](https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32016R0679), [*LOPDGDD*](https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673), ...)
* You must also contact with the Ethics Committee of the UPC. You can send an email to proteccio.dades@upc.edu

☐ I’m not working with personal or sensitive data

☐ I’m working with personal or sensitive data and I have already contacted with the Ethics Committee of the UPC.

☐ I’m working with personal or sensitive data but I haven't contacted with the Ethics Committee of the UPC yet.

**2. Documentation and data quality**

*Consider what information is needed to facilitate read, comprehension and reuse of the data*

## 2.1 Naming convention and folder structure

Indicates how the data will be organized during the research: conventions, versions, name and structure of folders [see: [*Guidelines for organizing, naming, and versioning files]*](https://confluence.csuc.cat/display/RDM/Organize%2C+name+and+version+files)

|  |
| --- |
|  |

## 2.2 Version control

Describe how you will control the versions

☐ No version control (e.g., original files are overwritten)

☐ Software with version control, indicate it:

☐ Software with change tracking option

☐ Version number and date in the file or folder name

☐ Making a copy of the script in which the data is processed

☐ Other, specify:

|  |
| --- |
|  |

## 2.3 Documentation associated to the data

* Specify the documentation that will accompany the data (e.g., [Readme files, data dictionaries, codebooks](https://confluence.csuc.cat/display/RDM/Description+of+the+data), ...) necessary to understand and reuse data. This documentation may include information about how data was collected, analytical and procedural information, definitions of variables, units of measurement, ...
* Also use standards and specify formats (preferably open and non-proprietary) and metadata to enable your data to be interoperable and reusable within your discipline

☐ Readme files

☐ Data dictionaries

☐ Codebooks

☐ Other, specify:

|  |
| --- |
|  |

For more information, see “[*Disciplinary metadata standards*](http://www.dcc.ac.uk/resources/metadata-standards)” of DCC

## 2.4 Metadata standards used for data discovery

If in the future you plan to publish your research data in an open repository such as [CORA. Repositori de Dades de Recerca](https://dataverse.csuc.cat/about.xhtml) you can indicate which metadata will be provided to help others identify and discover the data. This information is usually available in the information section of the repository's interface.

* Metadata repository: if you deposit your data in a repository, this have standard metadata like Dublin Core, DDI and use the OAI PMH protocol for interoperability. Also has a controlled vocabulary for data description and this date indexed in other aggregators and directories.

|  |
| --- |
|  |

For more information, see “[*Metadata standards*](http://en.wikipedia.org/wiki/Metadata_standards)” at Wikipedia

**3. Data storage and security during your research**

*Describe where the data is stored during your research and the technical measures that will be implemented to ensure data integrity (data backup), recoverability (prevention of data loss), security (to prevent unauthorized access)*

## 3.1 Measures to be taken to reduce the risk of data loss

☐ Access restrictions

☐ Encryption

☐ Data processing

☐ Regular backups

☐ Other, specify:

|  |
| --- |
|  |

## 3.2 Specify any restrictions (commercial, ethical or confidentiality ) that may affect your data and explain how will you manage access restrictions and data security during the research? (including back up)

|  |
| --- |
|  |

## 3.3 Major data security risks

*Identify the main risks, such as: accidental deletion of data, loss, or theft of data. Describe the consequences of potential data loss*

|  |
| --- |
|  |

## 3.4 Where will you store your data?

Describes where the data will be stored during research activities and how often it will be backed up. It is recommended to store data in at least two separate locations.

* The use of robust, managed storage provided by university Informatic teams is preferable: the following [*table*](https://dadesrecerca-upc-edu.translate.goog/ca/emmagatzema?_x_tr_sl=ca&_x_tr_tl=en&_x_tr_hl=es&_x_tr_pto=wapp) presents the platforms available at the UPC with their main characteristics that are recommended for preferential use over third-party platforms (Google, DropBox, Microsoft...) that may present administrative or ethical incompatibilities.
* Storing data on laptops, standalone hard drives or external storage devices such as USB sticks is not recommended.

☐ In the network of your department or research group

☐ In the university network

☐ Physical storage (e.g., USB, external hard drive)

☐ Cloud service (e.g., Dropbox)

☐ Other, specify:

|  |
| --- |
|  |

***Ex. 1*** During the course of the research, all data will be stored on local servers maintained and automatically backed up by [institution name]. Every night the data will be automatically backed up. The data will be replicated over multiple sites/data centres.

# 4. Legal and ethical requiriments

## 4.1 Are you working on personal data?

Remember that If you work with personal data it is necessary contact with the Ethics Committee of the UPC [see point 1.3]

☐ Yes, specify:

☐ No

|  |
| --- |
|  |

## 4.2 Are you working on sensitive data? such as health, biometric or genetic data, data of racial or ethnic origin, philosophical or religious beliefs, political ideas or other data with a high risk for the affected individuals? If so, indicate with which data.

☐ Yes, specify:

☐ No

|  |
| --- |
|  |

## 4.3 If you work with sensitive or personal data, how will you comply with the corresponding regulations regarding data management and security? Specify the techniques or tools used

*Make sure that when handling personal data, you comply with data protection laws*

* *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*

☐ Gain informed consent for preservation and/or sharing of personal data

☐ Anonymisation of personal data for preservation and/or sharing

☐ Pseudonymisation of personal data

☐ Encryption of personal data

☐ Managed access procedure in place for authorised users of personal data

|  |
| --- |
|  |

## 4.4 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.*
* *Indicate whether there are any restrictions on the re-use of third-party data.*

|  |
| --- |
|  |

## 4.5 What ethical issues and codes of conduct are there, and how will they be considered?

*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*

|  |
| --- |
|  |

# 5. Access, share and reuse the data

## 5.1 Do you have any restrictions on data sharing as regards the existing regulation ([General Data Protection Regulations](https://www.boe.es/doue/2016/119/L00001-00088.pdf)) or others (ethics, commercial, security, intellectual property, or copyright)?

*Specify which ones.*

|  |
| --- |
|  |

## 5.2 Who are the potential users of your data and how are they going to find them?

*Briefly describe who might be interested in your research and how you will distribute it (e.g. data repositories, website, conference publications, etc.)*

|  |
| --- |
|  |

## 5.3 Specify the licenses that you will apply to the data to enable maximum reuse

*The use of* [*Creative Commons*](https://bibliotecnica.upc.edu/en/propietat-intellectual/llicencies-acces-lliure-creative-commons-gnu) *licenses is recommended (CC - BY o CC Zero) or* [*GNU*](https://bibliotecnica.upc.edu/en/propietat-intellectual/llicencies-acces-lliure-creative-commons-gnu#quines-llicencies-gnu-es-poden-atorgar-al-programari-lliure)

|  |
| --- |
|  |

# 6. Deposit and conservation of the data

*Keep in mind that all favourably evaluated theses must be published in the respective open access institutional repositories*

## 6.1 What criteria will you use when selecting the data for long-term preservation?

☐ Type of data (raw, processed) and ease of generation

☐ Relevance of content to others

☐ Ease of reuse of the format by others

☐ Data linked to a publication

☐ Investigation verification

☐ Time available

☐ Available financial resources

☐ Others, specify:

## 6.2 How long do you intend to preserve the data?

*Various international standards recommend a minimum of 10 years*

|  |
| --- |
|  |

## 6.3 In which repository will you store your data?

*Consider* [*specific requirements*](https://confluence.csuc.cat/pages/viewpage.action?pageId=180453488) *in terms of format, metadata, size, cost, etc., that the repository may have to deposit data*

☐ UPC data instutional repository (e.g., CORA. Repositori de Dades de Recerca)

☐ Thematic repository (international), specify:

☐ Multidisciplinary repository (e.g., Zenodo, Figshare, Dryad)

☐ Others, specify:

November, 2024