

# Guide to develop a data management plan for doctoral students

(Doc.CO22/25) (4 RDM\Plans Gesto Dades\PlandeGestiodeDatos\_Doctorandos-EN.docx, 23.11.22)

This document intends to assist doctoral students in the development of their Data Management Plans (DMP).

This document was prepared by the CSUC Working Group on Research Support, which is composed of representatives from the following universities: University of Barcelona, Autonomous University of Barcelona, Polytechnic University of Catalonia, Pompeu Fabra University, University of Girona, University of Lleida, Rovira i Virgili University, Open University of Catalonia, University of Vic, Central University of Catalonia, Ramon Llull University and University of the Balearic Islands.

This document is licensed under the Attribution of Creative Commons (<http://creativecommons.org/licenses/by/4.0/>).

Digital version: <http://hdl.handle.net/2072/370509>

NOTE: Point 6. Sensitive/personal data has been prepared by the University of Lleida and is necessary for the processing of data from UdL doctoral theses.

## About your research

---

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

## About this data management plan

---

Creation date	<i>DD-MM-YYYY</i>
Last update	<i>DD-MM-YYYY</i>
Version and date	<i>Make a new version every time there are significant changes (new datasets, significant changes in your research, or other factors)</i>
Sensitive/personal data	<p><i>If you work with personal or sensitive data, you have a legal obligation to process it according to applicable regulations. Personal data is any information that allows a person to be identified (name, address, location, etc.).</i></p> <p><i>If you work with <b>personal data</b>, see the points:</i></p> <p><b>2. Data storage and security</b></p> <p><b>6. Sensitive/personal data</b></p> <p><input type="checkbox"/> <i>I'm not working with personal data</i></p> <p><input type="checkbox"/> <i>I will work with personal data [see point 2 and 6]</i></p>

## 1. Data collection

---

*Describe the data you will create/collect*

### 1.1 Will you use existing data during your search? If not, indicate the origin of the data you are going to use

- No
- 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 (indicate which ones):

### 1.2 Data Description

*Describe the data that you are going to create or the third-party data that you are going to reuse and specify:*

- *if you are going to use protocols or standards used in your research area*
- *the tools, instruments, equipment, hardware, or software you are going to use*

*If you reuse data from third parties, make sure to have the appropriate permissions and to be aware of the terms and conditions of the data.*

### 1.3 Data type and formats

*Keep in mind:*

- *the type of data: 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.*
- *the longevity of the file formats: preferably use open standards so that the data can be read by multiple programs, facilitating preservation, and sharing with other users.*

#### 1.4 Specify the data volume

- < 10 GB
- 10-30 GB
- 30-50 GB
- 50 GB-250 GB
- 250 GB-500 GB
- 500 GB-2 TB
- 2 TB

## 2. Data storage and security

---

*Make sure that all your research data is securely stored and backed up regularly.*

### 2.1 Specify any restrictions (commercial, ethical or confidentiality) that may affect your data

- Contractual obligations
- Legal obligations: protection of personal data (LOPDGDD, RGPD...) [see 4.1]
- Legal obligations: copyright, intellectual property [see 4.1]
- Ethical restrictions [see 4.1]
- Commercial aspects (p. ex. patentability)
- Formal security standards
- No obligations
- Other, specify:

*Briefly explain the restrictions*

*For more information: [LOPDGDD](#), [RGPD](#)*

### 2.2 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.*

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

- Access restrictions
- Encryption
- Data processing
- Pseudonymization
- Anonymization
- Regular backups
- Other, specify:

*Also specify the procedures you will use to guarantee the privacy of personal data.*

## **2.4 Where will you store your data?**

*In case of processing and storing personal data, answer the point: 6.6 Where will you store the data in case of processing personal data.*

- 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:

*Briefly explain the storage and copying conditions.*

### 3. Data documentation

---

*Document data to facilitate comprehension and reuse*

#### 3.1 Name and structure of the files and the folders

*Describe how your data files and folders will be organized and named.*

*When creating or collecting data and working with many files, these files can easily become disorganized. To save time and avoid mistakes, it is important to make decisions about how files will be organized and named, and what version control will be applied whether working individually or in groups.*

*Having conventions for the structure and organization of research files and data will improve their management and make research data more consistent. If these are also documented, other researchers will be able to understand them and reuse them in the short, medium and long term. You can learn more in the [Best Practices Guide in research data management: Organizing, Naming, and Versioning Files](#).*

#### 3.2 Version control

*Describe how you will control the versions. Also, specify what you will do if you delete data.*

- 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:

#### 3.3 Which metadata standards do you intend to use?

- I will not use any standard (specify the metadata needed to understand the data)
- Generic metadata schema (e.g., Dublin Core)
- Windows automatic metadata schema (e.g., from Word, Excel)
- Specialized thematic metadata schema, indicate:
- Another metadata schema, indicate:



*Specify how they will be created (in a "readme" file, in a spreadsheet, embedded in the data) and what documentation you will produce to make the data understandable to others.*

*For more information, see “Disciplinary metadata standards” of DCC or “Metadata standards” at Wikipedia.*



## 4. Access, share and reuse the data

---

### 4.1 Do you have any restrictions on data sharing as regards the existing regulation (General Data Protection Regulations) or others (ethics, commercial, security, intellectual property, or copyright)?

*Specify which ones.*

*For more information about existing regulation, see [General Data Protection Regulation](#)*

### 4.2 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.).*

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

*The use of Creative Commons licenses is recommended (CC - BY o CC Zero) or GNU.*

*Licenses are legal texts through which an author or the holder of the rights of a work authorizes third parties to reuse it under certain conditions. This is a non-exclusive grant of rights that may include reproduction, distribution, public communication and transformation. Without any indication, it must be understood that the work is offered with "All rights reserved" and therefore permission must be requested for its reuse, except in the cases provided for by the applicable law.*

*You can learn more in the [Best Practices Guide in research data management: rights and licenses for research data](#).*

## 5. Deposit and conservation of the data

---

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

### 5.1 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 formed by others
- Data linked to a publication
- Investigation verification
- Time available
- Available financial resources
- Others, specify:

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

*Various international standards recommend a minimum of 10 years.*

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

*Consider [specific requirements](#) in terms of format, metadata, size, cost, etc., that the repository may have to deposit data.*

*The repository for depositing data at the UdL is [CORA-Repositori de dades de Recerca](#). It is a federated and multidisciplinary data repository that allows Catalan universities, CERCA research centers and other research entities to publish research data sets in FAIR mode and following the EOSC guidelines.*

- Institutional repository
- Thematic repository (international), specify:
- Multidisciplinary repository (e.g., Zenodo, Figshare, Dryad)
- Others, specify:

## 6. Sensitive/personal data

---

### 6.1 Will you work with personal data? *In case of working with personal data specify which data*

*Personal data:* Any information about an identified or identifiable natural person. An identifiable natural person is one whose identity can be determined, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that person.

*Example:* Patient's medical record number, first or last name, voice, image, an email address, an IP address, date of birth, a person's lab results, NIF (tax identification number), age, nationality, etc.

- No
- Yes
- I'm not sure

### 6.2 In my research, is it essential to identify specific people?

*Anonymization:* The process that converts personal data into data that cannot be used to identify any individual, i.e., an irreversible process that prevents the collected personal data from identifying a specific natural person. Anonymization includes both the application of technical anonymization measures and security measures to prevent re-identification.

*Pseudonymization:* The processing of personal data in such a way that it can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data is not attributed to an identified or identifiable natural person.

- No, because personal data are anonymous
- Yes, because it is unavoidable to identify specific people

### 6.3 Why are personal data anonymous?

*The concept of anonymization should not be confused with pseudonymization. The latter allows individuals to be identified, either directly or indirectly, as long as additional information is available..*

*If you have any doubts about whether the data you are handling is anonymous or not, we provide this link:*

*[Anonimización y seudonimización | AEPD](#)*

- Because no identifiers of the affected people are captured or saved (mail addresses, IP numbers of devices connected to the internet, mobile phone numbers, names and surnames or ID, etc.)
- Because participation in the project only generates a code for each participant, which makes it impossible for the UdL research team to identify their identity
- Because under no circumstances are the voices or images of the affected people recorded
- Because it is not possible to deduce, in a reasonable way and without disproportionate efforts, from the results of the execution of the Projects (content of the surveys and interviews, content of the medical analytics, etc.), the identity of the people
- Because, although identifiers of the affected people are captured or saved, the necessary technical measures have been implemented to avoid re-identification ("anonymization" through data encryption or encryption)

### 6.4 Why do personal data identify specific people?

- Because identifiers of the affected people are captured or saved and no technical measures have been implemented to prevent re-identification
- Because recordings are made of the voice or image of the people affected
- Because in a reasonable way and without disproportionate effort it is possible to deduce, from the results of the execution of the Project, the identity of the people
- Because personal data pseudonymization systems are used

### 6.5 What is the age of the people whose personal data you will process?

*The Organic Law 2/2018, of December 5, on the protection of personal data and the guarantee of digital rights, in Article 7, establishes the following:*

*"1. El tratamiento de los datos personales de un menor de edad únicamente podrá fundarse en su consentimiento cuando sea mayor de catorce años. Se exceptúan los supuestos en que la ley exija la*

*asistencia de los titulares de la patria potestad o tutela para la celebración del acto o negocio jurídico en cuyo contexto se recaba el consentimiento para el tratamiento.*

*2. El tratamiento de los datos de los menores de catorce años, fundado en el consentimiento, solo será lícito si consta el del titular de la patria potestad o tutela, con el alcance que determinen los titulares de la patria potestad o tutela."*

- People under 14 years old
- People 14 years old or older

**6.6 If you are working with special category personal data, indicate wich ones:**

***Genetic data:** personal data relating to the inherited or acquired genetic characteristics of a natural person, which provide unique information about their physiology or health, obtained from the analysis of a biological sample from that person.*

***Biometric data:** personal data obtained from a specific technical processing, relating to the physical, physiological, or behavioral characteristics of a natural person, which allow or confirm the unique identification of that person, such as facial images or fingerprint data.*

***Health data:** personal data relating to the physical or mental health of a natural person that reveal information about their health status, including the provision of healthcare services.*

- Health
- Ethnic or racial origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Genetic data
- Biometric data
- Data concerning sexual life or sexual orientation
- Criminal records or similar
- Data of vulnerable groups or at risk of social exclusion
- Other high-risk data

**6.7 Confirm which tools you will use to manage personal data**

- I confirm my obligation to use only UdL tools, or those for which UdL has acquired the license
- I confirm that it is essential for my research to use additional tools, for which I assume full responsibility for their use and consequences

**6.8 What external tools to UdL will you use for personal data management? Why are they essential?**

*For example, external platforms to UdL, such as Google Forms, Mailchimp, Dropbox, WeTransfer, Whatsapp, Gmail, etc., should not be used.*

**6.9 I confirm that even though I use external tools to UdL for managing personal data, this data will also be stored in a UdL environment**

- I confirm that, in any case, the data will remain stored in a UdL environment (preferably electronic)

**6.10 Data of the participants in the project. In the design and development of the project's methodology and objectives, are researchers external to UdL or other third parties (e.g., a natural or legal person, companies, public authorities, services, or any other organizations) involved?**

- No, because researchers from other universities or research centers are not involved
- Yes, therefore, at the start of the activity, the corresponding co-responsibility agreement between all participating entities will be formalized

**6.11 In case of an affirmative response in the previous section, it is mandatory to identify the third parties (name and email address) involved in the project and explain their participation**

**6.12 How will I collect information in my research? If data is collected through databases from other public or private organizations, indicate which organizations or entities. Also specify any other sources of information, if applicable**

*Questionnaires, surveys, informed consents, and other relevant documents must be shared via the UNIDISC or OneDrive cloud.*

- *First: share the folder or document from the cloud with the address [dpd@udl.cat](mailto:dpd@udl.cat)*
- *Second: copy the link of the folder or document to the “comments” in this question 12, paste and add the link using the “insert link” icon.*

- Through surveys
- Through questionnaires
- Through UdL databases
- Through databases of other public or private organizations
- Through other sources of information

**6.13 I confirm that I will comply with the principles established in Article 5 of the GDPR, which states that personal data must be:**

- **Processed lawfully, fairly, and transparently**
- **Processed for specified, explicit, and legitimate purposes**
- **Adequate, relevant, and limited to what is necessary**
- **Accurate and kept up to date**
- **Retained only for as long as necessary**
- **Processed securely and confidentially**

I confirm

**6.14 I request an opinion from the Research Ethics and Transfer Committee (CERT) of the University of Lleida.**

*The request for an opinion from the CERT is voluntary, except when data protection regulations establish the obligation to obtain such an opinion.*

*Remember to share this form with the Data Protection Officer with “co-owner” permissions at the email address: [dpd@udl.cat](mailto:dpd@udl.cat)*

- Yes
- No