|  |
| --- |
| EPFL DATA MANAGEMENT PLAN (DMP) TEMPLATE |

**Scope**

This DMP template has been designed to be applicable **to any EPFL research project** that produces, collects or processes research data that has not already a DMP from a funder or publisher.

You should develop a single DMP for your project to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell them out. Depending on the project and research field, however, you may not need to address each of these questions in your DMP.

Throughout the DMP, it’s indicated if the required information is Mandatory (M), Mandatory if applicable (MA), or Recommended (R).

**Goal**

The main goal of a DMP is to **anticipate your needs** and **requirements** in terms of

* + **resources** (e.g. servers, hard drives, data curation and preservation, software tools, etc.)
	+ **data protection** **and ethics authorizations** (e.g. submission to the cantonal ethics commission, or to the HREC)
	+ **good practices** (e.g. standardized documentation, metadata collection, naming convention, data security, open science, regular back-ups, etc.)

Standardized procedures enable any collaborator to understand your processes.

# PERSONAL INFORMATION (M)

|  |  |
| --- | --- |
| **1a. Fellow’s/PI’s surname(s)**  |  |
| **1b. Fellow’s/PI’s first name(s)**  |  |
| **1c. Fellow’s/PI’s Orcid (if applicable)** (https://orcid.org/) |  |

# PROJECT AND FELLOWSHIP ID (M)

|  |  |
| --- | --- |
| **2a. Project title**  |  |
| **2b. Type of Fellowship/Funding Programme** |  |
| **2c. Institution**  |  |
| **2d. Laboratory** |  |
| **2e. Supervisor’s full name (optional)**  |  |
| **2f. Starting date of the project**  |  |
| **2g. What financial resources do you plan for data management?** |  |

All personal information contained in the DMP will be treated as confidential.

You can save/rename the DMP as follows: *SURNAME\_Firstname\_DMP*

Review and feedback for your DMP: *researchdata@epfl.ch*

Please send the DMP to: *research@epfl.ch*

Project Manager Contact: *[Add Project Manager Information]*

# INTRODUCTION (M)

3a. Could **any** of the data/code that you create, produce, collect or process contain personal data? [[1]](#footnote-1),

[ ]  Yes [x]  No (If you indicated “Yes”, your DMP must include the *Data Protection Annex*, see below.)

3b. Could **any** of the data that you collect or process contain sensitive data like ethnic data, genetic, biometric or health data[[2]](#footnote-2) or data that you are contractually obliged to protect

[x]  Yes [x]  No (If you indicated “Yes”, your DMP must include the *Data Protection Annex*, see below.)

3c. Could **any** of the data that you collect or process contain confidential data, as intellectual property of a third party that you are contractually obliged to protect?

[x]  Yes [x]  No

3d. Will you publish **any** of your data/code (Open Research Data), either because of an obligation (e.g. funder or publisher requirement), because it is the norm in your community, or because you are committed to doing so yourself?

[ ]  Yes [x]  No (If you indicated “Yes”, your DMP must include the *Open Data Annex*, see below.)

# Part 1: Data /Code that will be collected, produced and processed

Please, use the guidelines to fill in the table.

(M = Mandatory/ MA = Mandatory if applicable / R=Recommended)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P1a. # (M) | P1b. Contains personal data? (M) | P1c. Collected, Produced or Processed, coded (M)  | P1d. Description of dataset (M) | P1e. Formats (M) | P1f. Description of Metadata (R) | P1g. Origin (M)  | P1h. Naming, organization (R) | P1i. Responsible person (R) |
| 1 | [x]  Yes[[3]](#footnote-3)[x]  No | *i.e. collected data from field studies* | *i.e. Optical micrographs* | *i.e..tiff (raw)**.png (processed)* | *i.e. XML file containing scale settings, resolution, acquisition setup, possible annotations.* | *i.e. Self-produced with Leica DM2700M with Suite v3.8* | *i.e. TypeOfPicture\_nameOfPicture\_no.tiff* | *Surname Name* |
| 2 | [x]  Yes[x]  No |  |  |  |  |  |  |  |
| 3 | [x]  Yes[x]  No |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |  |

# Part 2: Understandability of collected, produced or processed data

This part is about how you make your datasets understandable for you and others.

Please use the guidelines to answer each question. (M=Mandatory/ MA=Mandatory if applicable/ R=Recommended)

**P2a. For whom are you collecting this data/code? (R)**

*e.g. for yourself, for a consortium, for other colleagues in the same field*

|  |
| --- |
|  |

**P2b. What documentation do you provide with your data/code for your audience and to make data/code understandable for future reuse? (MA)**

*e.g. README.txt, data codebooks, log files, protocols, …*

|  |
| --- |
|  |

**P2c. How you will ensure the quality of your data/code? (R)**

*e.g. cleaning processes, peer-review of dataset, community standards such as codebooks or libraries etc.*

|  |
| --- |
|  |

# Part 3: Data/Code Management infrastructure

Please, use the guidelines to fill in the table. (M = Mandatory / MA = Mandatory if applicable / R=Recommended)

Please, reuse the same ID # as in the Part 1, if you are talking about the same data/code.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| P3a. Data/Code Management System (M) | P3b. Role of data management system (M) | P3c. Backup strategy (M) | P3d. System Provider(s) (M) | **P3e. Access Management (M)** | P3f. EstimatedSize of Dataset (M) | P3g. # (MA) |
| *i.e. External SSD drive**.* | *i.e. Data processing, Versioning,* *Collaboration,* *Storing, etc.* | *i.e. 1 / 2 weeks(manual)* | *i.e. Self-hosted* | *i.e. PI, lab members, guests* | *i.e. Several TB* | *i.e. 1, 3* |
| *i.e. NAS Drive* | … | … | … | … | … | *i.e. 2* |
|  |  |  |  |  |  | *…* |
|  |  |  |  |  |  | … |

# Part 4: Data/CODE long-term preservation[[4]](#footnote-4)

Please, use the guidelines to fill in each field. (M = Mandatory / MA = Mandatory if applicable / R = Recommended)

Please, reuse the same ID # as in the Part 1, if you are talking about the same data/code.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| P4a. # (MA) | P4b. What concretely will be preserved? (M) | P4c. Preservation Purpose (M) | P4d. Deletion Process (M) | P4e. Retention Period (M) | P4f. Preservation infrastructure (MA) | P4g. Regularity (R) | P4h. Responsible Person (M) |
| 1 | *Part of the dataset, the graphs used for my article* | *Contractual obligation* | *After 3 years, there will be a secure deletion process* | *20 years, starting from YYYYMMDD* | *EPFL cold storage* | *At the end of the project* | *[Your Name]* |
| 2 |  |  |  |  |  |  |  |
| … |  |  |  |  |  |  |  |

 **P4i. Describe how this preservation will be obtained. What data curation process(es) need to be applied to your data/code on the long-run? (R)**

*e.g. metadata improvement, format migration, integrity check, measures to ensure accessibility, etc.*

|  |
| --- |
|  |

# Annex 1: “As Open As possible”-Research Data/Code (ORD)

This part is intended for use when datasets are published to an ORD platform/repository. We are aware that not all data can be published, but you can at least strive for publication of description of this kind of data. For clear definitions, please check guidelines.

(M = Mandatory / MA = Mandatory if applicable / R = Recommended)

**A1a. What is the intended purpose of opening up your data/code? (M)**

Describe here the intended purpose of opening up your data in relation to

* the objectives of the project,
* in reference to scientific practices in your community,
* for transdisciplinary reuse, or
* considering its value for society.

*i.e. For dataset number #1 mentioned in the tables above, I commit to publish at least the data underlying my article(s), at the moment of submission of my research article.*

|  |
| --- |
|  |

**List of all datasets/code that will be opened, where and how**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A1b. # (MA) | A1c. Title (M) | A1d. Author(s) (M) | A1e. Type/Formats (R) | A1f. Repository/Platform (M) | A1g. License (M) | A1h. Anonymization Process, if personal data (MA) |
| 1 | *Optical micrographs – a collection for the future* | *[Your Name]* | *.tiff* | *Zenodo* | *CC0* |  |
| 2 |  |  |  |  |  |  |
| … |  |  |  |  |  |  |

**A1i. What actions (if any) are planned to promote or facilitate the reuse of published open data? (R)**

*Write here, how you are going to engage the research community in reusing your data: will you write a blog article, a data paper, etc.*

|  |
| --- |
|  |

# Annex 2: Form for a data Protection Analysis

**Need help? Please contact the EPFL DPO:** **dpo@epfl.ch** **or** **chiara.tanteri@epfl.ch**

**Part 1: Datasets collected/produced**

|  |  |  |
| --- | --- | --- |
| **Question** | **Your answer** | **Further information/links/contacts** |
| **A2a. How many participants** will be included?  |  |  |
| **A2b. Data Subjects**: do you consider only Swiss residents? UE residents? Others?  |  |  |
| **A2c. Have you got the consent** (explicit consent in case of sensitive data) of the data subject (before proceeding to the data processing)? | [x]  Yes[x]  No |  |

**Part 2: The 9 criteria to evaluate if we face to “likely to result in a high risk” processing operations**

The current FADP does not require a Data protection impact assessment (DPIA), but if several of the following criteria are met, I will recommend that you do a DPIA. If GDPR applies to your project and if 2 or more of the following criteria are met you must conduct a DPIA.

Note that the new FADP (which is planned to entry in force at the end of 2022), the data controller (i.e. the person responsible to set the purpose and the means of the data processing) must carry out a DPIA in case of high risks for data subjects.

|  |  |  |
| --- | --- | --- |
| **Question** | **Your answer** | **Further information/links/contacts** |
| **A2d. Evaluation or scoring**, including profiling and predicting, especially from “aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements” | [x]  Yes[x]  No |  |
| **A2e. Automated-decision making** with legal or similar significant effect: processing that aims at taking decisions on data subjects producing “legal effects concerning the natural person” or which “similarly significantly affects the natural person” | [x]  Yes[x]  No |  |
| **A2f. Systematic monitoring**: processing used to observe, monitor or control data subjects, including data collected through networks or “a systematic monitoring of a publicly accessible area” | [x]  Yes[x]  No |  |
| **A2g. Sensitive data or data of a highly personal nature** | [x]  Yes[x]  No |  |
| **A2h. Data processed on a large scale** | [x]  Yes[x]  No |  |
| **A2i. Matching or combining datasets** | [x]  Yes[x]  No |  |
| **A2j. Data concerning vulnerable data subjects** (e.g. children) | [x]  Yes[x]  No |  |
| **A2k. Innovative use or applying new technological or organizational solutions** | [x]  Yes[x]  No | *Please, add more description if applicable* |
| **A2l. When the processing in itself “prevents data subjects from exercising a right or using a service or a contract”** | [x]  Yes[x]  No |  |

**Part 3: Security measures**

|  |  |  |
| --- | --- | --- |
| **Question** | **Your answer** | **Further information/links/contacts** |
| **A2m. Is the data processed in a manner that ensures appropriate security** of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organizational measures? | [x]  Yes[x]  No |  |
| **A2n. Do you share personal data** with another EPFL entity? | [x]  Yes[x]  No |  |
| **A2o. Do you share personal research data** with anyone outside of EPFL? (please specify EU/non EU/ USA)  | [x]  Yes[x]  No | *If yes, please describe which organizational and technical measures you have taken* |
| **A2p. Do you plan to work with a data processor** (e.g. private company that carries out the data processing on the behalf of the EPFL)? | [x]  Yes[x]  No |  |
| **A2q. Do you have a process in place** for granting and revoking appropriate user access (incl. privileged access rights and roles)? | [x]  Yes[x]  No |  |
| **A2r. Do you have standards** for isolating sensitive data and procedures and technologies in place to protect it from unauthorized access and tampering? | [x]  Yes[x]  No |  |
| **A2s. Which technical and organizational measures** do you take to ensure IT security?  | *Please, specify here the measures you plan to take (e.g. for technical measures: encryption, encryption key management process, …)*  |  |
| **A2t. How do you ensure** that they are appropriate and correctly implemented regarding to the risks?  | *e.g. tests, audits, …* |  |
| **A2u. Do you implement Privacy-by-Design** or by-Default? | [x]  Yes[x]  No | *Please, add more information* |
| **A2v. How do you ensure the personal data traceability?** Note that within the new FADP, you must keep the event logs related to personal data processing for a period of 2 years | *e.g. event logs, data value history* |  |

1. Personal data is information that relates to an identified or identifiable individual. If you cannot directly identify an individual from that information, then you need to consider whether the individual is still identifiable. You should take into account that information together with all the means reasonably likely to be used by either you or any other person to identify that individual. See FADP Swiss law ([**Art. 3**](https://www.fedlex.admin.ch/eli/cc/1993/1945_1945_1945/en#art_3).c.) [↑](#footnote-ref-1)
2. E.g. 1. religious, ideological, political or trade union-related views or activities, 2. health, the intimate sphere or the racial origin, 3. social security measures, 4. administrative or criminal proceedings and sanctions (Art 3.c.) [↑](#footnote-ref-2)
3. Please fill in *Data Protection Annex* [↑](#footnote-ref-3)
4. Digital preservation can be defined as a "series of managed activities necessary to ensure continued access to digital materials for as long as necessary" [(Digital Preservation Handbook)](http://www.dpconline.org/docman/digital-preservation-handbook/299-digital-preservation-handbook/file). [↑](#footnote-ref-4)