

UPSTEM Data Management Plan

This document has been prepared using the Data Management Plan template of the EPFL.

Goal of a DMP

The main objective of a Data Management Plan (DMP) is to anticipate the project needs and requirements in terms of:

- resources (e.g., servers, hard drives, data curation and preservation, software tools, etc.)
- good practices (e.g., standardized documentation, metadata sharing, naming convention, data security, open formats, regular backup, etc.)
- data protection and ethics authorizations (e.g., submission to the cantonal ethics commission)

A DMP is a live document supposed to be updated to reflect the actual data management. At the end of a research project, the associated DMP becomes a precious document that enables research reproducibility, allowing other researchers to understand your processes.

PART 1. GENERAL INFORMATION

Personal Information

P1a. Fellow/PI's full name(s) and email(s)	<i>Julia Venturini</i> julia.venturini@swissyoungacademy.ch
P1b. Fellow/PI's ORCID(s)	0000-0001-9527-2903
P1c. DMP contact person's full name and email	<i>Marco Cavallaro</i> marco.cavallaro@swissyoungacademy.ch

Project details

P1d. Project title	Unveiling the leaky Pipeline in STEM (UPSTEM)
P1e. Project description	The UPSTEM project aims to understand the reasons behind the leaky pipeline in STEM fields in Switzerland. This project surveys former researchers who left academia in the past 15 years.
P1f. Name of fellowship/funding programme	Swiss Young Academy Joint Projects
P1g. Institution(s) involved	Swiss Young Academy
P1h. Starting date of the project	January 2024

Data global context

P1j. Could any of the data/code that you create, produce, collect or process contain personal data? ¹	Yes
P1k. Could any of the data that you collect or process contain sensitive data ² or data that you are legally obliged to protect?	Yes
P1l. Could any of the data that you collect or process contain confidential data or intellectual property of a third party that you are legally obliged to protect?	No
P1m. Will you publish any of your data/code (Open Research Data)? ³	No
P1n. Are ethics authorizations necessary for research involving humans and/or personal data from the competent ethics committee?	Yes

¹ Personal data: information that relates to an identified or identifiable individual. Not only data that can directly identify a person (e.g., name) is considered personal data, but also data that can make a person identifiable through the combination of data (e.g., combining age, e-mail address and information related to usage of social networks may allow identifying a person). You should consider that information together with all the means reasonably likely to be used by either you or any other person to identify that individual. See also [Art. 3.c](#) of the FADP Swiss law.

² Sensitive data: information that includes but is not limited to religious, ideological, political or trade union-related views or activities; health, genetic, biometric, or concerning the intimate sphere or the racial origin; ethnic data or social security measures; administrative or criminal proceedings and sanctions. See also [Art. 3.c](#) of the FADP Swiss law.

³ 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.

Part 2. DATA/Code collection⁴

P2a. Dataset identifier #	P2b. Collection ⁴ method	P2c. Description of dataset, type of data/code	P2d. Origin / source / equipment	P2e. Naming convention, format	P2f. Responsible person	P2g. Contains personal data? ⁵
1	Websurvey	<i>One full dataset comprising participants' answers to the questionnaire, along with email addresses of the participants interested in updates on the study, invitation to final workshop, and/or potential qualitative follow-up study (interviews and/or focus groups).</i>	<i>Questionnaire completed in the QuestionPro software</i>	<i>UPSTEM study_Dataset_YYYY-MM-DD</i>	<i>Marco Cavallaro</i>	Yes
2	Websurvey	<i>Participants' answers to the questionnaire.</i>	<i>Questionnaire completed in the QuestionPro software</i>	<i>UPSTEM study_Responses_Y YYY-MM-DD</i>	<i>Marco Cavallaro</i>	No
3	<i>Collected from the websurvey, where participants can voluntarily enter their e-mail address in case they wish to be informed of the developments of the project</i>	<i>Email addresses of the participants interested in updates on the study, invitation to final workshop, and/or potential qualitative follow-up study (interviews and/or focus groups).</i>	<i>Questionnaire completed in the QuestionPro software</i>	<i>UPSTEM study_Emails_YYYY-MM-DD</i>	<i>Marco Cavallaro</i>	Yes

⁴ Acquired, produced, processed, or coded.

⁵ If any "Yes" in this column, please fill in the *Annex 2: Data Protection Analysis*

Understandability of acquired, produced, processed, or coded data

P2h. For whom are you collecting ⁶ this data/code?	<i>For a project, UPSTEM, supported by the Swiss Young Academy, that aims to gain understanding of the reasons behind the leaky pipeline in STEM fields in Switzerland. Data will only be accessible to the UPSTEM project members.</i>
P2i. What documentation do you provide with your data/code for your audience to make data/code understandable for future reuse?	<i>Our data will not be published in a public data repository due to the exploratory nature of this pilot study. The results and conclusions of the study will be made openly available by the Swiss Young Academy in the form of a report.</i>

Part 3. Data/Code storage or Management infrastructure

P3a. Storage or management system	P3b. Role of system	P3c. System provider(s)	P3d. Access rights management	P3e. Estimated size of dataset
Servers of the Swiss Academy of Sciences (SCNAT) / Swiss Academies of Arts and Sciences at the House of the Academies	<i>Data processing, Versioning, Storing, Collaboration</i>	<i>Self-hosted</i>	<i>Only UPSTEM project members can access via authentication</i>	<i>Up to 200 KB</i>

⁶ Acquired, produced, processed, or coded.

Part 4. Data/CODE long-term preservation⁷

It is expected that the preservation for all datasets of PART 2 is discussed here. For column P4a., use all and the same identifiers # as in P2a.

P4a. # (same as P2a.)	P4b. What will be preserved?	P4c. Preservation purpose	P4d. Retention Period	P4e. Deletion process of the preserved data	P4f. Preservation infrastructure	P4g. When/How often will preservation be done?	P4h. Responsible person
1	<i>The questionnaire and list of emails will be extracted from this file, which will then be deleted.</i>	<i>not applicable</i>	<i>not applicable</i>	<i>The dataset #1 will be erased upon extraction of the datasets #1 and #2 by an UPSTEM project member.</i>	<i>not applicable</i>	<i>not applicable</i>	<i>Marco Cavallaro</i>
2	<i>The complete dataset will be preserved.</i>	<i>Use for the preparation of a report</i>	<i>3 years, starting from the survey dissemination date until May 31, 2028.</i>	<i>The dataset #2 will be erased on May 31, 2028, by an UPSTEM project member.</i>	<i>Servers of the SCNAT / Swiss Academies of Arts and Sciences at the House of the Academies</i>	<i>Only once, at the end of the project.</i>	<i>Marco Cavallaro</i>
3	<i>The complete dataset will be preserved. The deletion of personal data can happen anytime respondents who left their emails ask for it.</i>	<i>Use for communicating updates about the final workshop and a potential qualitative (interviews or focus groups) follow-up study</i>	<i>3 years, starting from the survey dissemination date until May 31, 2028.</i>	<i>The dataset #3 will be erased on May 31, 2028, by an UPSTEM project member.</i>	<i>Servers of the SCNAT / Swiss Academies of Arts and Sciences at the House of the Academies</i>	<i>Only once, at the end of the project</i>	<i>Marco Cavallaro</i>

⁷ 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](#)).

Annex 1. Data Protection Analysis

Data and datasets collected and/or produced

A1a. How many participants will be included?	The survey will be filled out on a voluntary basis. We cannot estimate at this stage a number of participants.	
A1b. Participants' geographical location: do you consider only Swiss residents? UE residents? Others?	We foresee no limitation with regards to the geographical location, as long as the participants correspond to our target group (STEM scholars with a completed PhD that worked for a Swiss university/higher education institute and left academia during the last 15 years (2010-2025))	
A1c. Consent: have you planned to get or already obtained the subjects' consent? ⁸	Yes	Before starting the survey, participants are informed about the scope of the study, its duration, potential impact, risks related to potentially sensitive questions, and how data is handled. They are also made aware of the possibility to withdraw from the questionnaire at any point.

Security measures

A1d. Is the data processed in a manner that ensures appropriate security ⁹ of the personal data using appropriate technical or organizational measures?	Yes	
A1e. Which technical and organizational measures do you plan to ensure IT security?	The data is stored locally on the servers of the Swiss Academy of Sciences (SCNAT) / Swiss Academies of Arts and Sciences at the House of the Academies and is backed up off-site at a colocation in Bern. The data is released via Server Message Block (SMB) and Hypertext Transfer Protocol Secure (HTTPS). This means that the data is always stored on the well-secured SCNAT systems. Access rights to the share for those who receive full access are	

⁸ Research involving human participants or their personal data cannot be permitted, unless the participant has given his/her consent: it is necessary to collect participants' consent before proceeding to data collection and processing.

⁹ This might include protection against unauthorized or unlawful processing and against accidental loss, destruction, or damage.

	managed by SCNAT IT. Full access to the share is only granted to one employee of the Administrative Office of the Swiss Young Academy. The folder containing the UPSTEM project data on this share is accessible only to members of the project group and one employee of the Administrative Office of the Swiss Young Academy.	
A1f. How do you ensure that technical and organizational measures are appropriate and correctly implemented regarding the risks?	The proposed solution for data management and backup was discussed and implemented together with the SCNAT IT department.	
A1g. Do you plan to work with a data processor ¹⁰ ?	No	
A1h. Do you have a process in place for granting and revoking appropriate user access (incl. privileged access rights and roles)?	Yes	The members of the project group receive an account from the SCNAT IT and thus access to the data as 'editors', i.e. they can edit, upload and download data, comment on it and view it in the preview - but not delete it or release it for others. The access rights will not be changed unless the speaker of the projects asks the staff member of the Administrative Office of the Swiss Young Academy (as the person responsible for the share) to adjust them in the event that members of the project group leave the project or new members join the project group. After the project has ended, all access to the corresponding cloud folder will be deleted by SCNAT IT at the instruction of the staff member of the Swiss Young Academy Administrative Office.

¹⁰ E.g., private company or other entity that carries out the data processing on the behalf of your research group or laboratory.