DMP – "INHALE" Project

1. Data collection and documentation

1.1 What data will you collect, observe, generate or re-use?

The data produced by this research project falls into two categories:

- Primary data: including recorded and textual data generated through pupils' interviews.
- Secondary data: including textual data from observation collected by teachers.

The specific data types, storage formats and volumes are listed in Table 1 below

Origin	Type	Equipment/ Software	Format	Size
Primary data	Interview	Recorder	.mp3	<100 GB
	Transcription	Whisper, Nvivo	.docx	<100 GB
	Questionnaire answers	RedCap	.csv	<100 GB
Secondary data	Observation notes	Word	.docx	<100 GB
	Academic transcript	PDF	.pdf	<100 GB

1.2 How will the data be collected, observed or generated?

- **Structured Interviews:**
  - **Method**: Conducting in-depth structured interviews with three classes of 9th grade Harmos, each with 22 students.
  - **Instruments**: Standardized questionnaires such as the Psychological Distress K10, the School Burnout Inventory, and the Stress Resilience Scale will guide the interviews.
  - **Procedure**: Interviews will be conducted in a controlled environment to ensure consistency. Each interview will be recorded and transcribed using Whisper software. In addition, the answers to the questionnaire and the demographic information will be recorded in RedCap during the interview.
- **Observational Notes:**
  - **Method**: Teachers will observe and document students' behavior before and after the interview sessions.
  - **Instruments**: Observation sheets or digital note-taking tools.
  - **Procedure**: Teachers will use a standardized format to ensure consistency in the observations.

- **Quality measurement**:
  - **Sound check**: carried out before the interviews are recorded to check the sound quality.
  - **Pilot Testing**: Conduct a pilot test of the interviews and questionnaires to identify any issues and make necessary adjustments.
  - **Standardization**: Use standardized instruments and procedures to ensure consistency and reliability of the data.

The folders will be organized according to the following structure:

1. Project folders
    a. Project Management
        i. Proposal
        ii. Finance
        iii. reports
    b. Ethics Governance
        i. Ethical approvals
        ii. Consent forms
    c. Experiment one
        i. Readme.txt
        ii. Input
        iii. Data
            1. Database
                a. Codebook
                b. data
            2. Audio
                a. Audio lit.docx
                b. Date_Audio_PupilCode.mp3
            3. Transcript
                a. Transcription list.docx
                b. Date_Tran_PupilCode.docx
            4. Observational notes
                a. Observational notes list.docx
                b. Date_Obs_PupilCode.docx
        iv. Analysis
        v. output
    d. Experiment two
        i. idem
    e. Dissemination
        i. Presentation
        ii. Publication
        iii. vulgarization

Files naming strategy:

- Date_Audio_PupilCode.mp3
- Date_Tran_PupilCode.docx
- Date_Obs_PupilCode.docx

Versioning strategy: A table with the version numbers, the author's name, the change and the date will be added at the beginning of each transcription. The same will be done for the observational notes.

Version	Author name	Changes	date
V01	AM, post-doc	Removal of direct identifier	2024.10.10
V02	AZ, PhD	Removal of indirect identifier	2024.10.25

1.3 What documentation and metadata will you provide with the data?

We will document our project on three different levels:

- Project Level: A report will be produced, including a brief summary of the project, information on copyright, ownership, data collection etc, following the CESSDA recommendations.
- Files and folders level: A readme.txt describing the file naming convention and the file structure will be attached to the project folder
- Data level (follow the UK data service recommendations):
    - A list of interviews conducted, including *interview ID, age, gender, occupation, organization, location, place of interview, date of interview, transcript file name, recording file name*, will be attached to the audio folder.
    - A list of the transcriptions, including … will be attached to the transcription folder.
    - A codebook describing the codes used to analyze the interview, how codes were selected and generated by Nvivo.
    - A Codebook describing the variables on REDCap

Metadata: The metadata of the INHALE project (such as audio transcription and observational notes) will be available on the Lausanne University data repository SWISSUBase in a generic English version as well as in French. The repository SWISSUBase, follow the DDI metadata standard (title, authors names, institution, etc.).

2. Ethics, legal and security issues

2.1 How will ethical issues be addressed and handled?

INHALE activities are related to human subjects research. Therefore, this project is guided by:

- The Swiss Federal Human Research Act (HRA) legislation
- International research ethics principles such as the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine developed by the council of Europe in 1997 (Ovideo, 1997) and the 'Declaration of Helsinki' (a set of ethical principles regarding human experimentation developed for the medical community by the World Medical Association, last revised at the 64th WMA Meeting held in Fortalezza/Brazil in October 2013) (World Medical Association (WMA), 2024).

In accordance with the HRA legislation, INHALE has been submitted to the CER Vaud and has been validated.

Furthermore, INHALE respondents are volunteers, and the entire data collection is based on informed consent. This holds on three levels:

1. First, all potential respondents are informed about the contents of the interview:
   - Participants and their parents will receive detailed written information at the beginning of the project.
   - Then, researchers will provide an oral presentation of the project to ensure understanding and address any questions.
2. Second, at the end of the oral presentation with the pupils and parents, the consent forms are distributed to the participants. Teachers will collect the completed consent forms three weeks later in class. The consent forms will include:
   - General information: authors information, subject study, aims and hypos, description of the structured interviews etc.
   - Free Use of Data: Consent will be sought for the use of data in the study.
   - Anonymized Data: Consent for the use of anonymized data for future research and publication.
   - Data Sharing: Authorization to archive and share anonymized data with other researchers.
3. Third, during the interview, answers to all questions are voluntary. Each single question can be skipped if an individual does not want to answer a specific question.

2.2 How will data access and security be managed?

Detailed technical and organizational measures are in place to ensure the security of data processing and compliance with the national data protection laws.

- Data will be manually anonymized by removing direct and indirect identifiers throughout the transcription of the audio. Anonymization takes place before the data is analyzed, shared or published. All names and other personal information that could be used to identify teenagers are removed from these datasets. Released data never contains ID numbers but only codes.
- The correspondence table with the student's name and their correspondence code is randomly generated before transcription and then assigned to the transcription and observation note.
- Only the PI have access to the correspondence codes list.
- Only software approved by the DPO of the University will be used for this project. This includes REDCap, Nvivo and Whisper software.
- Strict access controls are implemented. Data access will be given only to researchers involved in the collection and analysis of the data as well as to the PI in charge of the INHALE project.
- Encryption: Veracrypt solution will be used for the encryption of the data.
- We will follow the IT recommendations and use the institutional storage solution Tresorit and the automated backup solution Comet Backup.

2.3 How will you handle copyright and Intellectual Property Rights issues?

- **Intellectual property rights (primary data)**
  - Data owner: University of Lausanne
  - Agreements between researchers

    - Agreement with external parties concerning the transfer of copyright and ownership rights
    - Licence selected: **CC BY-NC-SA**
    - DOI: The INHALE project will use Digital Object Identifiers (DOI) to make datasets permanently identifiable and locatable.
- **Respect of copyright (secondary data)**
    - Data owner: The school involved in the project
    - Respecting the terms of re-use
    - Citation

3. Data storage and preservation

3.1 How will your data be stored and backed-up during the research?

Storage solution

- With the exception of the correspondence table, all data is stored in the institutional cloud solution Tresorit. Tresorit is a cloud storage solution that uses end-to-end encryption to store and share folders and files with designated collaborators. It provides encryption, access logging, and file versioning that is compliant with the Swiss Federal Act on Human Research (HRA).
- UNIL has a special contract with Tresorit AG, which guarantees that data is stored in Switzerland. Tresorit uses online storage on the Microsoft Azure platform. Tresorit encrypt the data before sending it to the online repository. They require the identity of the researcher to be verified using a password and a code on the researcher's mobile phone. Only the PI and the researcher assigned to the project have access to the platform.
- Tresorit provides three physical copies of the data.
- The correspondence table is stored on a password-protected external hard drive in a separate office, which requires a key to enter.
- Research material related to the project, such as documentation, will be attached to the data on Tresorit.

Back-up solution

- Comet Backup is an application provided by UNIL's IT Center with which it is possible to back up users' computer session. This software allows users to choose the documents to back up and to restore files in case of problems, without the help of a local IT manager.
- The data is stored on the European servers of Wasabi SA and is encrypted with an encryption key (encryption key A). Encryption key A is itself encrypted with encryption key E, which is stored in the personal profile of the person registered for the service on the Comet Backup server of UNIL. Encryption key E is itself encrypted with encryption key R, which is stored by the backup software on the machine of the person registered for the service. The data centers are managed by Wasabi. Wasabi cannot access the data due to its encryption. The service administrators do not have access to the data of the people registered for the service. If necessary and upon request for support from a person registered for the service, temporary access to the requester’s data can be granted to a service administrator.

- Data backup is triggered every 3 hours. Data retention on the backup servers is 90 days. Data deleted by the person registered for the service can be recovered up to 90 days after deletion, provided it has been backed up at least once. A device not connected to the service for more than 5 months is considered obsolete and is automatically deactivated.

3.2 What is your data preservation plan?

Due to the sensitivity of the data, only anonymized raw data will be made available via an institutional data repository (e.g. SWISSUbase), at the end of the project. Data will be archived together with the relevant detailed documentation. SWISSUbase is recognized by the SNFS as a FAIR data repository.

The rest of the research material will be archived on the University's LT servers and will only be available to the PI and researchers involved in the project. The archive will include all raw data, transformed data, transformation operations, version history, documentation, consent forms, informational sheets and so on.

The comprehensive archive will be maintained for a minimum period of 10 years, after this term data will be re-evaluated in order to determine what to preserve.

When possible, data will be stored and archived in non-proprietary formats (.txt, .csv, .xml, .pdf). In case this is not possible, we will include information on the software used and its version number.

4. Data sharing and reuse

4.1 How and where will the data be shared?

Research material from this work will be deposited in SWISSUbase, the University of Lausanne's institutional data repository, and made available at the time of publication. In accordance with the informed consent and the aim of potential re-use, only anonymised raw data will be shared. To this end, we will share the anonymized versions of the transcription, observation notes, REDCap database and documentation. Access to the data is free for scientific use

Data files deposited in SWISSUbase will be assigned to a Digital Object Identifier (DOI) and the associated metadata will be listed in the projects catalogue. The initial retention period for data will be 10 years from the date of deposit, with extensions applied to datasets accessed subsequently.

The DOI assigned to datasets in the repository can be included as part of a data citation in publications, allowing the datasets underlying a publication to be identified and accessed.

Metadata about datasets held in the University Repository will be publicly searchable and discoverable via the SWISSUbase projects catalogue and will indicate how and on what terms the dataset can be accessed.

Data to keep	Strategy	Audience	Purpose	Location	Format
Transcription	Sharing	Researchers in the field	Reuse	SWISSUBase	.txt
Observation notes	Sharing	Researchers in the field	Reuse	SWISSUBase	.txt

All materials	Archiving	PI and researchers involved in the project	Future personal research PI need	LTS Unil
Etc…				

4.2 Are there any necessary limitations to protect sensitive data?

Due to the sensitivity of the project, data will only be shared for teaching or similar research purposes.

Therefore, the data will only be shared under the conditional access solution provided by SWISSUbase.

Conditional access sharing provides metadata information to unauthorized individuals and access to the entire shared material to individuals who meet the conditions.

This allows researchers to retain control over the re-use of the data.