

DMP title

Project Name FWO Data Management Plan - DMP title

Project Identifier 1S31522N

Grant Title 1S31522N

Principal Investigator / Researcher Nicolas Heintz

Description In this research project, we will use

Institution KU Leuven

1. General Information

Name applicant

Nicolas Heintz

FWO Project Number & Title

1S31522N: Fast, adaptive and wearable auditory attention decoding: towards practical neuro-steered hearing devices

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Pseudonymised EEG signals, channel names, stimulus gains, experiment parameters	.mat, .bdf, .csv	10 GB	Recorded during experiment with EEG sensors.
Stimulus	.wav, .mat	1GB	Extracted from public databases (e.g. universiteit van Vlaanderen)
Behavioural data	.xls, .docx...	<100MB	Interview of the subjects: questions of
Informed consent forms	.pdf	<100MB	Signed by subject
Name, age, gender and hearing capability of participant	.xls, .docx...	<100MB	Filled in by subject

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Privacy Registry Reference: G-2022-5194

Short description of the kind of personal data that will be used:

The methods developed during the course of this PhD will be validated on several data sets consisting of:

- Pseudonymized EEG recordings in auditory attention decoding tasks
- Results from behavioral data and the speech intelligibility tests
- Time-varying volume of competing speakers
- Informed Consent Forms
- Subject selection results and strictly necessary personal information (name, age, gender...)

These datasets have been and will be collected in the ExpORL division, with informed consent forms

from all involved subjects/patients, covering the research proposed in this project proposal. For collection of new data, ethical approval will be obtained first.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

If new data are collected, ethical approval will be obtained first at the Social and Societal Ethics Committee (SMEC). There are nevertheless currently no specific plans for data collection and thus no ethical approvals submitted.

We will mostly use readily available datasets that are either publicly available or recorded at the ExpORL lab for other research projects.

The PRET application for using available datasets is file G-2022-5194.

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

Any IP restrictions for valorisation will be aimed at specific algorithms and the results of these algorithms on used data. The results would include classification accuracy, correlations and other metrics. The datasets themselves will not be subject to any IP restrictions for this project specifically.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

Each collected dataset will have:

- A ReadMe file with the most important aspects of the data: what are the different conditions, what was the goal of the project...
- A clear reference to the used auditory stimuli
- All parameters relevant to the experiment (sampling rate, recording device, number of channels, channel names, gains of the stimuli, class label (eg to which speaker was the subject listening)...))
- Preprocessing parameters if relevant
- The preprocessing file used in the study if relevant
- Preprocessed data if relevant
- Information of additional data (eg. eye gaze...) if relevant

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

All datasets are stored according to the BIDS standard.

5. Data storage and backup during the FWO project

Where will the data be stored?

- The EEG data set is pseudonymized after acquisition and stored on encrypted KU Leuven (ESAT) servers.
- The informed consent forms and identifiers containing the personal data are each stored separately and will only be accessible to a small number of responsible people.
- We will temporarily store the pseudonymized data on a local, encrypted hard drive during the development of new models. It would be computationally cumbersome to train and test them on a data set stored on a non-local server. The raw data are removed from the local hard drive when the research is completed.
- We may grant other researchers of the lab access to the pseudonymized EEG data when they need it for their research.

How is backup of the data provided?

All datasets will be stored on encrypted KU Leuven (ESAT) servers, which are automatically backed up.

All algorithm implementations and research results are stored on a local computer and are backed up at least once every two weeks on both an external HD disk and on KU Leuven OneDrive. Note that these disks only contain research results, algorithms and other non-personal information. No personal data are stored on these disks.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available then explain how this will be taken care of.

- Yes

The datasets used in this research project are stored on KU Leuven servers with sufficient storage. Most datasets used in the project are currently already stored on these servers for other projects or are publicly available.

The research output (results, code...) generated in the project require relatively little space compared to these datasets (<10GB) and can thus easily be stored on any server/disk.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

This project is not expected to generate additional costs for data storage and back up. We will use readily available storage facilities provided by (departments of) KU Leuven (ESAT servers, KU Leuven Drives, OneDrive...), which are sufficiently large.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All datasets will be stored on secure servers, which are not accessible for unauthorized persons. The access to these servers is monitored by the supervisors of this project and other authorised personnel of KU Leuven. The security is provided by the ICTS of KU Leuven (ESAT).

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

Any collected datasets will be retained for longer than 5 years after the end of the project, since these datasets are also used by other research projects.

Where will the data be archived (= stored for the longer term)?

The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

Given the relatively small data sizes (<100GB), we don't expect any additional costs for data preservation. Any user of the KU Leuven storage facilities can use 2TB of data (and up to 5 TB without additional costs).

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

The pseudonymised EEG data may be shared given explicit consent of the participants. However, no personal identifiers (name, age...) will be shared publicly.

Which data will be made available after the end of the project?

The pseudonymised data will be made available on the K-drive of the KU Leuven Drives. Given consent, the data will also be made publicly available to external research groups through e.g. Zenodo.

Where/how will the data be made available for reuse?

- In an Open Access repository
- In a restricted access repository

Pseudonymised data will be made available on Zenodo given permission from the participants. The source code of the developed algorithms in this research project will be made available on GitHub after (submission for) publication.

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Any researcher from KU Leuven will have access to the pseudonymised data. Given consent of the participants and the ethical committee, this access can also be extended to external researchers through an open-access database such as Zenodo.

What are the expected costs for data sharing? How will the costs be covered?

No additional costs are expected for data sharing.

8. Responsibilities

Who will be responsible for data documentation & metadata?

The researcher himself, Nicolas Heintz, will be responsible for data documentation and metadata.

Who will be responsible for data storage & back up during the project?

The researcher himself, Nicolas Heintz, will be responsible for data storage & back up during the project.

Who will be responsible for ensuring data preservation and reuse ?

The project supervisors, Alexander Bertrand and Tom Francart, will ensure data preservation and reuse, also after the project has finished. Nicolas Heintz will nevertheless be stay available for any questions regarding potential datasets after project termination.

Who bears the end responsibility for updating & implementing this DMP?

Nicolas Heintz is responsible for updating & implementing this DMP during the project. However, Alexander Bertrand and Tom Francart will be responsible for long term data management.