DMP title

Project Name FWO - PhD fellow Strategic basic research grant - DMP title

Grant Title 1SB5922N

Principal Investigator / Researcher Miguel Claro Bhagubai

Description AI-based automated seizure monitoring tools for epilepsy patients. More specifically, how can we detect seizures using wearable medical sensors and how to build a robust multimodal framework (EEG, ECG, and EMG) to be used in non-controlled environments. The data to be used was collected within the SeizeIT2 project.

Institution KU Leuven

1. General Information

Name applicant

Miguel Maiur Claro Bhagubai

FWO Project Number & Title

1SB5922N

PhD fellow strategic basic research

Affiliation

- KU Leuven

2. Data description

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

A large dataset is being collected in the EIT-funded SeizeIT2 project. This contains video-electroencephalography (vEEG), electrocardiography (ECG) and electromyography (EMG) data of epileptic patients. Additionally, patients are simultaneously recorded with the byteflies wearable EEG, ECG and EMG sensors. The data is annotated by neurologists and stored in EDF format (european data format). The total estimated volume is 1000 - 1500 GB.

The data is going to be used to build and test machine/deep learning models, in which features from the different modalities are extracted (either manually, with the use of standard signal processing techniques; or automatically, in the case of deep learning, in which the features are extracted via data-driven approaches). The framework will be built with the use of MatLab (The MathWorks Inc.) and Python.

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:

Short description of the kind of personal data that will be used: anonymized physiological measurements (EEG, EMG, ECG) and annotations of epileptic seizures.
The PI of the Seize IT2 project (Wim Van Paesschen), who is in charge of data collection, is currently applying for a Privacy.

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

Ethical approval from CTC under number S63631

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?

- No

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?

- Yes

The data cannot be disseminated until the clinical trial ends (June 2022), which is covered by the original ethical approval. We do not generate novel data, so no extra restrictions are in place.

4. Documentation and metadata

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

In the context of the Seize IT2 project, the data has been organized in an easily shareable format (so it can be shared after the end of the clinical trial). This data is organized in EDF format, and annotations are standardized in .csv format so they can be easily read for further analysis.

During the development of the seizure detection frameworks, the methodologies will be documented in text/report format. Additionally, the code and scripts will have documentation and commented lines for ease of reproducibility.

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

See EDF: European Data Format

5. Data storage and backup during the FWO project

Where will the data be stored?

The code and clinical datasets are stored on the ESAT servers (Department of Electrical Engineering, KU Leuven), where access is regulated by an access control list (ACL) that grants:
- read/write access to the project owner
- read-only access to specific users

The ACL is managed by the project owner. Client computers can access the data using SMB2 from specific IP ranges or NFSv4 from specific (IT-managed) systems.

How is backup of the data provided?

Data in ESAT servers is daily backed up. It is also replicated in an off-site storage system and will be stored at least 5 years after the research project ends.

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 available storage on the ESAT servers is larger than the maximum estimated volume of the dataset.

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

Costs are automatically covered by ESAT, the host department.

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

See above, answer on ACL.

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, ...).

All data without exceptions.

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

Data will be stored in a secure archiving server at ESAT, conforming to ongoing policies.

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

The ESAT servers have enough server space to store the data for the coming 10 years. Policies are in place to expand server space when it would be full.

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:

Data cannot be shared until the trial is completed. Afterward, data can be shared upon specific conditions, and will always be granted by PI Prof. Van Paesschen.

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

All data (recordings and annotations), with permission from Prof. Van Paesschen. The code will be made available as well, following approval from the developers and promotors.

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

- In a restricted access repository
- Upon request by mail

The dataset will be accessible from the ESAT servers for the parties allowed to access it. The code will be stored within GitLab and the ESAT servers, on a private repository, and shared only with the parties allowed to access it.

When will the data be made available?

- Immediately after the end of the project

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

The PI of the data collection will decide on this upon completion of the study.

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

Costs are covered by the ESAT service.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Not part of this grant.

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

ESAT, KU Leuven, the host department.

Who will be responsible for ensuring data preservation and reuse ?

The promotor (Prof. Maarten De Vos) and co-promoter/PI of the data collection (Prof. Wim Van Paesschen)

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

The PI bears the end responsibility of updating & implementing this DMP.