FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Guido Gagliardi, 0000-0003-2020-6439**
Contributor name(s) (+ ORCID) & roles	**Prof. Maarten De Vos, supervisor, 0000-0002-3482-5145**
Project number [1] & title	1SH4Z24N, eXplainable Artificial Intelligence to embed new-generation Artificial Intelligence architectures for brain signal analysis in clinical scenarios
Funder(s) GrantID [2]	FWO (1SH4Z24N)
Affiliation(s)	KU Leuven ROR identifier KU Leuven: 05f950310
Please provide a short project description	New-generation Artificial Intelligence architectures are powerful tools when applied to brain signal analysis in multiple clinical use cases, e.g. seizure detection or sleep staging. The clinical context would greatly benefit from introducing these architectures, using them as a diagnostic aid tool, or automating costly and time-consuming processes. Even though these architectures show great performance in recognition tasks, they are not yet implemented in a clinical setting because they are known to suffer from what is known as black box scepticism, i.e. clinical end users do not trust them because they recognize their results to be "too good to be true" and their decision-making mechanism to be too opaque to be easily understood. In this project, we will propose a novel eXplainable Artificial Intelligence framework to solve the trust problem of new-generation Artificial Intelligence architectures for brain signal analysis and we will systematically test and validate it in two clinical use cases, i.e. seizure detection and sleep staging. The new framework will address the problem of trust by proposing new algorithms to "open the black box" and by providing explanations of its decision-making mechanisms in line with the clinical context and by allowing clinical end users to personalize and adapt these explanations to their knowledge and experience over time.

2. Research Data Summary

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data [3].

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
MAHNOB-HCI	EEG benchmark dataset for a multiclass emotion recognition task	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.edf	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
DEAP	EEG benchmark dataset for a multiclass emotion recognition task	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 100 GB	
TUSZ	EEG dataset for Seizure detection	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 1 TB	
Epilepisae	EEG dataset for Seizure detection	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 100 GB	
SHHS	EEG benchmark dataset for	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 1 TB	

[3] Add rows for each dataset you want to describe.

	sleep staging						
MASS	EEG benchmark dataset for sleep staging	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 1 TB	
DREEM	EEG benchmark dataset for sleep staging	☒ Reuse existing data	☒ Digital	☒ Numerical	. edf	☒ < 100 GB	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.

MAHNOB-HCI	https://mahnob-db.eu/
DEAP	https://www.eecs.qmul.ac.uk/mmv/datasets/deap/
TUSZ	https://isip.piconepress.com/projects/tuh_eeg/index.shtml
Epilepisae	https://epilepsy-database.eu/
SHHS	https://sleepdata.org/datasets/shhs
MASS	http://ceams-carsm.ca/mass/
DREEM	https://github.com/Dreem-Organization/dreem-learning-open

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

☐ Yes, human subject data; provide SMEC or EC approval number:
☐ Yes, animal data; provide ECD reference number:
☐ Yes, dual use; provide approval number:
☒ No
Additional information:
Publicly shared anonymized datasets

Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☒ No If yes, please comment:
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. Documentation and Metadata

[4] See Glossary Flemish Standard Data Management Plan

Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded). *RDM guidance on documentation and metadata*.	**To allow for replication of the experiments developed in the project the data will be kept in its original format as benchmark data, as it is acquired by the provider.** **Since they are quite popular benchmark datasets for EEG classification in the state of the art literature they already came with all the information to keep them understandable and usable.**
Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☐ Yes ☒ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: If no, please specify (where appropriate per dataset or data type) which metadata will be created: No additional metadata will be created.

4. Data Storage & Back-up during the Research Project

Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☐ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☐ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☒ Other (specify) The data is already stored in the ESAT (KU Leuven) servers. The backups are performed daily. Furthermore, the data is replicated to a remote storage system located at the ICTS data centre.
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, all the data is already available and safe in the ESAT (KU Leuven) servers. ☐ No If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	The information will be kept on the ESAT servers, whose access is controlled by an ACL that authorizes: Access to the project owner with read-write privileges. - read-only access for a select group of users. The project owner oversees the ACL. Client computers can use to access the data. - SMB2 (or higher) from particular IP ranges. - NFSv4 from particular (IT-managed) systems.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	No additional costs for data storage are required.

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☐ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specifiy): **The data is replicated to a remote storage system located at the ICTS data centre.**
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	No additional costs for data storage are required.

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	
Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☐ KU Leuven RDR ☒ Other data repository (specify): All the code will be available on my github account https://github.com/guidogagl ☐ Other (specify)

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☐ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☒ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☐ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☒ No
What are the expected costs for data sharing? How will these costs be covered?	**No additional costs expected for data sharing.**

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Me, (Guido Gagliardi)**

Who will manage data storage and backup during the research project?	**ESAT department**
Who will manage data preservation and sharing?	**Me, (Guido Gagliardi)**
Who will update and implement this DMP?	**Me, (Guido Gagliardi)**