

LatentEMG: Data management plan

0	General information	
0.1	Project number	J2-70103
0.2	Project title	LatentEMG: Analysis and Utilization of Latent Spaces in High-Density Electromyography
0.3	Project leader number	21301
0.4	Project leader name	prof. Aleš Holobar, PhD
0.5	Research organisation's data steward name	Luka Kevorkijan, data steward at University of Maribor (UM) for the field of Natural sciences and Technology
0.6	Research organisation's internal data management policy	Since October 2025 the UM adopted the University of Maribor Open Science Policy. Web page: https://www.um.si/en/research/open-science/university-of-maribor-open-science-policy/
0.7	DMP version	1.0
1	Abstract and description of research data	
1.1	Will the project reuse any existing data from previous research?	☒ Yes ☐ No We will reuse an existing collection of surface HDEMG signals that were measured during repeated dynamic contractions of the biceps brachii muscle in a preliminary study in 2018. The data will be used to test, validate, and compare the developed signal processing procedures.
1.2	What types of data will you create or reuse and in what formats will it be stored?	We will create several collections of synthetic surface HDEMG signals for different muscles, and we will also perform several experimental recordings with healthy adult volunteers, where we will capture surface HDEMG signals from different muscles. The data will be stored as 2D matrices of raw EMG values in binary files in the Matlab data format (.MAT). This format is widely used among researchers in the field of biomedical signals and allows for significantly more efficient storage than CSV files. Open source functions for reading and writing are available.
1.3	What is the purpose of creating, collecting, or reusing data and how does it relate to the project goals?	The HDEMG signals will be used for testing, validation and comparison of the HDEMG signal processing procedures developed during the project. They will also be used for demonstration and educational purposes at various workshops and courses. The data will be freely available to the wider professional public according to the FAIR principle.
1.4	What is the expected size of the data you plan to create or reuse?	☐ 0–10 GB ☒ 10–100 GB ☐ 100–1000 GB ☐ >1000 GB
2	Data storage and backup	

2.1	Where will the data be stored and backed up during the project?	During the project implementation, the data will be stored on the project data server within the RO network. The server has regulated access protection and regular data backups.
2.2	How will you select data for long-term retention?	We will follow the guidelines of the Open Science Policy of the University of Maribor. We intend to store all created datasets in the public repository of the Digital Library of the University of Maribor (DKUM) for the long term. There are no restrictions on publication for artificially generated or simulated data. We will publish all experimental data for which we obtain written consent for publication from the measured subjects, in accordance with the ethics committee approval.
2.3	Will the data be stored in a trusted repository?	☒ Yes ☐ No The data will be stored in the long-term in the Digital Library of the University of Maribor – DKUM (https://dk.um.si/info/index.php/slo/), which is a certified institutional trusted repository.
3.	FAIR data management	
3.1	Ensuring data is findable (F)	
3.1.1	Will the data be assigned a persistent identifier (PID)?	☒ Yes ☐ No All datasets will be labelled with PIDs that will be assigned by the DKUM repository.
3.1.2	What metadata will be created and what metadata standards will be followed?	We will describe the datasets with Dublin Core meta tags stored in an XML file. We will also add any additional descriptions that the DKUM repository requires when submitting data (based on the DataCite Metadata Schema standard).
3.1.3	Will the metadata contain keywords to improve findability and reusability?	☒ Yes ☐ No We will also add a set of 6-10 keywords to the metadata to improve data findability.
3.2	Ensuring data is accessible (A)	
3.2.1	Will all data be openly accessible?	☒ Yes ☐ No All simulated data and all experimental data for which we obtain consent from volunteers will be made available in open access.
3.2.2	When will the data be openly accessible and for how long?	The data will be available by default no later than 6 months after creation or possible publication in a scientific article. They will be uploaded to a trusted public repository, where they will be stored for the long term, together with all metadata. If some data will be key to possible intellectual property protection (patents), we will publish such data only after the intellectual property protection phase has been completed.

3.2.3	How will access to data be ensured during the project and after its completion in the event of restrictions on use?	We do not foresee any restrictions on the use of the published data, which will be available in open access under a CC0 or CC-BY license.
3.2.4	Will additional documentation or information about the appropriate software be required to access or read the data?	☐ Yes ☒ No There are a number of open source solutions available for reading Matlab data structures in various programming languages, e.g. the GNU Octave software package, the Python Scipy library, the MatIO library in the C programming language, etc.
3.3	Ensuring data is interoperable (I)	
3.3.1	Which ontologies or codebooks will you use when preparing data and metadata?	There are no standard ontologies or codebooks in our research field yet, so we will describe the data using general metadata standards (Dublin Core) and a detailed text description in the README.TXT file.
3.3.2	Will you be forced to use less well-known or proprietary ontologies or codebooks?	☐ Yes ☒ No We will not describe the data with our own ontologies or codebooks.
3.4	Ensuring data is reusable (R)	
3.4.1	How will you provide the documentation required for data reuse?	Documentation for data reuse will be provided with a comprehensive description in the README.TXT file. This description will be in English and will include: title, authors, date, version, funders, keywords, PID and access information, license, proposed citation format, general overview of the data, detailed description of the data and data structures, size, description and links to routines for reading the data, detailed description of the software, procedure and parameters for data generation, procedures for post-processing and quality assurance, and a list of people involved in data preparation. We will also add a screenshot showing all loaded data and their data types in the Matlab programming environment.
3.4.2	Will your data be publicly available and licensed under open licenses CC BY or CC BY-SA to enable the widest possible reuse?	☒ Yes ☐ No All datasets will be available under a CC-BY or CC0 license.
3.4.3	What data quality assurance procedures will you use?	All generated/recorded HDEMG signals will be visualized and manually reviewed by an expert to check for any errors or anomalies. Any inappropriate signals will be specifically marked.
4.	Ethical and legal aspects	
4.1	Are there any ethical or legal issues that could affect data sharing?	☒ Yes ☐ No There are no ethical restrictions on sharing simulated HDEMG signals. For experimental signals, we will obtain

		written consent from volunteers to publish and share the data. If this consent will not be given, the data will not be shared in open access.
4.2	Will you process or store personal data during the implementation of the project?	☐ Yes ☒ No We will follow the GDPR when processing data. We will obtain prior consent from the Commission of the Republic of Slovenia for Medical Ethics for the research. We do not anticipate the processing and storage of personal data. All potentially sensitive data will be pseudo-anonymized. Volunteers who will participate in experimental measurements will be assigned an anonymous identifier instead of their name, and we will also store their gender, age, and body weight.
4.3	Will special types of personal data be created or reused during the project?	☐ Yes ☒ No We will not collect/use any special types of personal data in the project.
4.4	How will you manage data ownership and any copyrights on the data you create or reuse?	The copyright and ownership of the data generated will belong to the organization that generates/captures the data.
5. Other research results		
5.1	Will you create or reuse other research outputs in addition to the data?	☒ Yes ☐ No We will also create: a) algorithms and software b) dissemination materials: posters, placards, handouts, news releases, photos, etc. c) scientific publications. To the greatest extent possible, we will also make these results publicly available under the CC0 or CC-BY license and publish them on the project website or in the DKUM repository.
6. Financial resources		
6.1	What will be the costs of handling data and other project results according to FAIR principles and how will they be covered?	The costs of data storage and processing during the project implementation and the costs of the main data server of the project within the RO network for the duration of the project and for 5 years after the end of the project will be covered by the Faculty of Electrical Engineering, Computer Science and Informatics in Maribor. For long-term data storage, we will use the DKUM repository, which has its own funding. The costs of publishing articles in open access journals will be covered by project funds.
6.2	Who will be the person responsible for handling research data in the project?	Project leader, prof. Aleš Holobar, PhD.