Advanced modelling of cerebrovascular autoregulation lab animal data to build a patient autoregulation monitor

A Data Management Plan created using DMPonline.be

Creators: Bavo Kempen, n.n. n.n., Maarten Devos https://orcid.org/0000-0002-3482-5145

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n., Maarten Devos https://orcid.org/0000-0002-3482-5145

Project Administrator: Bavo Kempen

Grant number / **URL:** 1S12523N

ID: 198677

Start date: 01-11-2022

End date: 01-11-2026

Project abstract:

Experiencing traumatic brain injury (TBI) can lead to life-changing consequences, ranging from mild impairments to devastating disabilities and premature death. Research has shown that following the initial injury, subsequent secondary injuries can occur due to abnormal brain functioning. These secondary injuries are a known important threat to good outcome and rehabilitation.
Cerebrovascular autoregulation (CA) is the mechanism that ensures adequate cerebral blood flow, carrying essential nutrients and oxygen. TBI may disturb the protective CA mechanism and exacerbate the probability of permanent brain damage. If it would be possible to monitor the CA status, this could be a valuable help for clinicians to guide blood pressure management. To date, no validated CA monitor exists that
can adequately determine and predict the status of CA in real-time.
The current project aims at developing a dynamic CA monitor for clinical use. To do so, we must first perform an in depth investigation on CA and its dynamic nature to gain insight into the underlying interacting physiological processes over time using the data generated by the Leuven porcine cranial window model experiments, which is reflective of human physiology. Based on the acquired information a mathematical model will be developed using machine learning techniques to detect and describe the CA status dynamically. The CA monitor will be validated in further studies on the experimental animal model and extended to humans.

Last modified: 21-04-2023

Advanced modelling of cerebrovascular autoregulation lab animal data to build a patient autoregulation monitor
Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects …) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

1) Reuse existing animal (Piglet model) and human data and generate novel data.
2) Human data can be linked to an identifiable natural person within the context of the hospital, but is always provided to me with an unidentifiable pseudonymized ID.
3) Raw intensive care unit monitoring data saved in .dta and transformed into .csv and .hdf5 format. Preprocessed ICU data saved in .hdf5 and .tfrecords. Developed code in .py, .R, .ipynb format. Developed machine learning models as .h5, .pth or .ckpt. Genetics data in .BAM, .VCF.

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

1. Dr. MD Bart Depreitere
2. Sufficient storage capacity is guaranteed:
    - During and after the research the utilized data, code and models is maintained on a local server (Synology), an additional local server (NIKON PC), L - drive (KUL), potentially expanded to Vlaamse Super Computer server. Code and non - medical data are also retained on online hosting services in private repositories (e.g. Github) until project completion / paper publishing.
    - Access will be limited to the members of the research group, aside from model and accompanying code, and should remain available 5 years beyond the end of the project.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

NA - Should be able to retain all processed data during the project.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

For side projects, human medical data was anonymized prior to access and handled according to GDPR legislation and hospital policy, of which the latter is more severe in terms of security.

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

The same data sets are preprocessed in numerous distinct ways for their respective purposes (e.g. downsampling). While the code should be sufficient to recreate them from the raw data sets, given allocated seed for pseudorandomization, the preprocessed data sets will be retained too.

Advanced modelling of cerebrovascular autoregulation lab animal data to build a patient autoregulation monitor
DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Advanced modelling of cerebrovascular autoregulation lab animal data to build a patient autoregulation monitor
GDPR

GDPR

Have you registered personal data processing activities for this project?

- Yes

Advanced modelling of cerebrovascular autoregulation lab animal data to build a patient autoregulation monitor
FWO DMP (Flemish Standard DMP)

1. Research Data Summary

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.

				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
		Please choose from the following options: • Generate new data • Reuse existing data	*Please choose from the following options:* • Digital • Physical	*Please choose from the following options:* • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA	*Please choose from the following options:* • .por, .xml, .tab, .cvs,.pdf, .txt, .rtf, .dwg, .gml, … • NA	*Please choose from the following options:* • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA	
Baseline - Cranial Window Piglet data set	Experimental manipulation limited to gradually induced hypo - or hypertension.	Reuse existing data	Digital	Experimental, observational, and compiled/aggregated data	.csv, .dta, .hdf5	< 100 GB	
CO2 - Cranial Window Piglet data set	Experimental manipulation pertained 1) gradually induced hypo -or hypertension, and 2) induced hypo - or hypercarbia.	Reuse existing data	Digital	Experimental, observational, and compiled/aggregated data	.csv, .dta, .hdf5	< 100 GB	
Ketamine - Cranial Window Piglet data set	Experimental manipulation pertained 1) gradually induced hypo -or hypertension, and 2) level of Ketamine used.	Reuse existing data	Digital	Experimental, observational, and compiled/aggregated data	.csv, .dta, .hdf5	< 100 GB	
KidsBrainIT data set	Paediatric ICU data set gathered in function of the KidsBrainIT project in the context of severe TBI.	Reuse existing data	Digital	Observational and compiled/aggregated data	.csv	< 100 GB	
Edinburgh - Newcastle data set	Paediatric ICU data set gathered in the context of severe TBI.	Reuse existing data	Digital	Observational and compiled/aggregated data	.csv	< 100 GB	
BrainIT, Center-TBI and other adult TBI data sets.	Adult ICU data sets gathered in the context of severe TBI. Will be used to translate the developed models from animal to human population.	Reuse existing data	Digital	Observational and compiled/aggregated data	.csv, .dta, .hdf5	< 5 TB	
Genetics side project data sets and code	Adult ICU data set gathered in the context of severe TBI and in addition genetic profile derived from bloodsamples.	Reuse existing data	Digital	Observational and compiled/aggregated data	.csv, .BAM, .VCF	< 1 TB	
Project code and data -ground truth animal experiments (WP1)	Scripts, data and reports used to determine and generate ground truth.	Generate new data	Digital	Software, compiled/aggregated data and other	.R, .xlsx, .csv, .pdf, .docx	< 1 GB	
Project code and data - anomaly detection code (WP2)	Preprocessing, experimental frameworks, model code, evaluation code	Generate new data	Digital	Software, experimental, compiled/aggregated data and other	.py, .sh, .ipynb, .R, .txt, .tfrecords, .hdf5, .csv, .git, .png	< 100 GB	
Project code and data - CBF prediction and Lassen curve simulation code (WP3)	Preprocessing, experimental frameworks, model code, evaluation code	Generate new data	Digital	Software, experimental, simulation, compiled/aggregated data and other	.py, .sh, .ipynb, .R, .txt, .tfrecords, .hdf5, .csv, .git, .png	< 100 GB	
Project code and generated data KidsBrainIT	Preprocessing, data generation and analyses, statistical analyses	Generate new data	Digital	Software, observational, compiled/aggregated data and other	.py, .ipynb, .R, .Rmd, .jpeg, .tiff, .eps	< 100 GB	
Project code and generated data Edinburgh - Newcastle data set	Preprocessing, data generation and analyses, statistical analyses	Generate new data	Digital	Software, observational, compiled/aggregated data and other	.py, .ipynb, .R, .Rmd, .jpeg, .tiff, .eps	< 100 GB	
Project code and generated data - genetics project	Preprocessing, data generation and analyses, statistical analyses	Generate new data	Digital	Software, observational, compiled/aggregated data and other	.py, .ipynb, .R, .Rmd, .jpeg, .tiff, .eps, .VCF, .csv, .xlsx	< 100 GB	

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:

1) Animal data set (Porcine cranial window model) - retrieved from local repositories provided by colleague researchers whom gathered the latter data as part of their experimental work. This data is not public, but is described in papers (e.g., https://doi.org/10.1038/s41598-019-50046-x)

2) Edinburgh - Newcastle and KidsBrainIT data sets are not public but received due to participation of promotor in the multinational network of the consortium.

3) Genetics data set, gathered locally as part of other projects. Not public.
4) BrainIT and Center-TBI databases, among others, will be received under a data transfer agreement

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Human data that was and may be used (KidsBrainIT, Edinburg - Newcastle, Center-TBI, BrainIT) are anonymised, i.e. impossible for me to determine identity based on the data, and have been gathered under conditions which allows secondary analyses of the data. Genetic data and accompagnying pseudonymised ICU data is processed and analysed within the scope of the original study objective and its respective ethical approval.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

Genetics side project data and accompagnying ICU data.

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

Models (+ its trained weights) and software source code developed during the PhD could potentially lead to commercial valorization if the PhD candidate is successfull in his objective.

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 in the comment section to what data they relate and what restrictions are in place.

- Yes

Human data used throughout the different projects are restricted, i.e. I cannot disseminate the data, solely analyze and report results.

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 in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

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).

1) Developed code will be well documented using inline comments, function/class input - output explanations and general README.txt files to describe the general scope of the project.
2) Raw data that is transformed into new data is well documented within its file in a standardised manner (e.g. .hdf5 meta data on sampling rate, variable explanations, comments, ...). The processing of the latter data and reasoning behind it is clarified in accompanying codebase.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Currently no plans exist to make the reused data repositories public.
Code can be made public upon publication with respective metadata (i.e. README, code annotation).

3. Data storage & back-up during the research project

Where will the data be stored?

1) Local server owned by the research group (NAS - Synology)
2) Local (server) computer owned by the research group
3) KU Leuven file storage and shared network drive
4) UZ Leuven server (genetics and accompanying ICU data)
5) Online private repositories for codebases (e.g. Github)

How will the data be backed up?

Raw digital data is saved in multiple online and local servers at the same time. The code developed to process the data is constructed in such a way (i.e. pseudorandom with fixed seed value) that it will reproduce the original processed data set in case of loss. Code is maintained and backed up in online private repositories after changes are made locally.

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

Yes, a distributed system of local servers allow for storage and backup of all the projects' data.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data is protected with 2FA. Servers are also within a locked office space.

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

The local servers were purchased prior to this research project and are paid.

4. 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...).

All data relevant to research outcomes will be preserved on local servers and internal KU Leuven data storage facilities to ensure reproducibility of attained results from published research.

Where will these data be archived (stored and curated for the long-term)?

KU Leuven storage solution in addition to local server copies.

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

Given the relatively small data set sizes, especially when compressed, no extra costs are expected for data preservation during the expected retention period.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Other, please specify:

1. Software can be made (openly) available after patent and proper licensing.
2. Data sets (animal model, human) used to train and develop the models: I do not have permission nor ownership over the used data sets and thus cannot claim / decide whether or not the data will be made publicly available in the future.

If access is restricted, please specify who will be able to access the data and under what conditions.

Raw biomedical data and their derivatives are not mine to disseminate. Published code and respective model weights may be openly shared upon a to be agreed upon license scheme and potential patent.

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 in the comment section per dataset or data type where appropriate.

- Yes, Privacy aspects
- Yes, Intellectual Property Rights
- Yes, Ethical aspects

Data sets are gathered and disseminated by a third party to me in a pseudonymised/anonymised manner, I cannot share this data nor derived data.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

KU Leuven RDR

When will the data be made available?

Upon publication of research results and proper protection to software.

Which data usage licenses are you going to provide? If none, please explain why.

Depending on the progress of the project at the end of the PhD. An MIT Lincense would be ideal for the developed software if much future work is to be done, otherwise if the project already developed in a finalized product a more restrictive license would be more appropriate (e.g. no commercial use allowed by other parties, ...).

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

The DOI will belong to the respectively published paper and linked to the online software repository.

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

0

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Bavo Kempen

Who will manage data storage and backup during the research project?

Bavo Kempen

Who will manage data preservation and sharing?

Bart Depreitere

Who will update and implement this DMP?

Bavo Kempen