DATA MANAGEMENT PLAN: BICEPS

A Data Management Plan created using DMPonline.be

Creator: Hans Dierckx

Affiliation: KU Leuven (KUL)

Template: KU Leuven BOF-IOF

Grant number: ID-N/22/006

ID: 3E220600

Start date: 1/10/2022

End date: 30/9/2026

Project abstract:
Life-threatening arrhythmias often occur with pathologies of the heart. As medical examinations only give indirect information on the underlying cause, it remains, even for experts, often difficult to assess with certainty which would be the ideal treatment. Computer modeling of cardiac arrhythmias has the potential to reveal underlying mechanics, if only the model can be tuned to specific patients. Computer modeling can also enhance the identification of locations for ablation. To be of practical use in the clinic, more insight is needed in the reliability of such identification. Both modeling errors and limited amounts of measured data lead to uncertainty on computer simulations. In this project, we will increase the relevance of computer simulations of cardiac arrhythmias for clinical practice, by developing a computational framework for uncertainty quantification. This step comes with a large computational cost for which we will introduce multiscale modeling for cardiac excitation simulations. We apply our findings to the characterization of ventricular tachycardia substrates from non-contact endocardial catheters, and predict virtual ablation outcome, with uncertainty quantification on the forecast.

Last modified: 16-05-2023

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.

- Image-based geometries of animal hearts obtained via MRI or CT scan at Gasthuisberg hospital in the group of Piet Claus. We will use max. 10 geometries here, which are 1 GB each. **Estimated volume for storage: 10GB.**

- Functional measurements of animal hearts in new animal experiments. The output is provided in standard medical formats. **Estimated volume for storage: 100 GB**.

- Simulations of cardiac excitation using different software (ithildin, openCARP, …), performed at the Dierckx group (Math, Kulak) and Samaey group (NUMA, Leuven): newly generated and digital. Output are the parameter set-up and version of the code, and the output of the different physical fields, e.g., transmembrane voltage (custom .npy-files) and electrogram traces (as .txt files or numpy arrays).
  3D heart simulations can be few 100 MB per frame, leading to 1-10 GB per simulation. The use of multiscale hierarchical approaches and explorations of parameter space can lead to thousands of simulations.
  Estimated volume: 10TB.

- Code generated for scripting and analysis of the simulation loop, with version control. **Estimated volume: 10GB**

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:

In addition to the above, historical data can be used form animal experiments obtained within C1 project (C14/18/079) and FWO project (G097021N) of which PC is co-PI, according to their resp. DMP's.

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.

The animal experiments will be subjected to the ethical committee of KU Leuven and the approval number will be inserted here when available.

Will you process personal data? 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).

No personal data will be processed, within this project we only work with in silico data and animal experiments.

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.

Within the scope of this project, the research is fundamental, and no commercial exploitation is foreseen.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g., Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate to and what restrictions are in place.

No.

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.

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).

For simulations, there are standardized log-files, whose description is provided at the wiki page of the software. See https://gitlab.kuleuven.be/heartkor/ithildin (internal page) and https://opencarp.org/documentation. These logfiles are kept together with the simulation output. For ithildin, a git hash is given referring to the software version (commit).

For hierarchical bayesian simulations, the set-up script is saved together with the folder structure.

For the medical images of animal hearts, a standardized headerfile contains the details of the acquisition. A manual comment is added to describe the subject and reason for imaging.

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.

For the experiments, metadata standards are provided by the hardware.
For our own software (ithildin), we implement automatic upload to the KU Leuven ManGO platform. A *json template file has been created and is distributed among the groups at KU Leuven using the software. For openCARP, the question of metadata standardisation will be discussed at the upcoming openCARP user meeting in May 2023.

If not, please specify which metadata will be created to make the data easier to find and reuse.

N/A.

DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?

- Python scripts and source code are stored at the GitLab repository: gitlab.kuleuven.be. This includes both the ithildin source files and the pre- and post-processing scripts.
- The output of simulations is stored in ManGO.
- The experimental recordings and images are stored at internal servers of UZ Leuven.
- Project documents, reports and meeting reports are stored in the SharePoint folder of the HeartKOR group.

How will the data be backed up?

All the repositories above are equipped with a back-up system.

Is there currently sufficient storage & backup capacity during the project?

The only concern is the possible 10 TB of simulation data in ManGO. We are in contact with the support team, who is currently implementing the upload of large files (of 5-10 GB each). When that is successful, we will discuss which storage capacity can be acquired for our collaboration.

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

We will discuss with the MANGO team if they can provide 10 TB. This will be needed by the end of 2024, so there is time to resolve this potential issue.

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

The servers (GitLab, ManGO, SharePoint) are protected via an authentication procedure.

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

At the current price, we will require 11 TB over 4 years, at a price of 40€/TB/year, this becomes 1760€, which is foreseen in the project budget.

DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for 10 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...).

At the end of the project, we will select which datasets should be kept for a longer period. Estimated that about 50% needs to be kept, for 10 years, a budget of 2200€ is foreseen for long-term data preservation.

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

After the project, the ‘cold’ data will be transferred from ManGO to appropriate servers, such as KU Leuven RDR or cold storage at KU Leuven data centers. Parts of the results will be shared together with the publications (e.g., via Zenodo or journal supplements)

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

See above, 2200€ can be used from the project budget.

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.

In agreement with the open science policy, the relevant part of the data with each publication can be shared, either via the ManGO platform, international alternatives (e.g., Zenodo) or journal supplements.
Scripts and simulation code can be shared as we did before, see e.g. https://gitlab.com/heartkor .
Animal experiments in the Claus lab carried out for BICEPS will be published in open access, via the KU Leuven RDR.

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

There will be no restriction on the sharing of the experimental datasets within the project.

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.

We see no such restrictions.

Where will the data be made available?

See above, on data sharing platforms. We would like also to publish versions of software at gitlab.kuleuven.be but this is currently not possible, so we need to export it to gitlab.com.

If already known, please provide a repository per dataset or data type.

Software: https://gitlab.com/heartkor
A public link to the ManGO or RDR folder is not yet available.

When will the data be made available?

At the time of publication of the accompanying paper.

Which data usage licenses are you going to provide?

For the experimental data: CC-BY-NC-ND-4.0 (no changes or commercial use possible)
For the software development: GNU General Public Licence 3.0 (further use remains free software)
Depending on the specific outcome, modifications will be possible and if appropriated discussed with LRD.

If none, please explain why.

N/A.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g., a DOI or accession number? If already available, please provide it here.

Yes, this will be done for datasets linked to publications.

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

Currently, gitlab.kuleuven.be, gitlab.com and zenodo.org are free services to the user.
The sharing via journal's servers is paid at the time of submitting the paper.

RESPONSIBILITIES

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

Project PI Prof. Hans Dierckx

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

Project PI Prof. Hans Dierckx

Who will manage data preservation and sharing?

Project PI Prof. Hans Dierckx

Who will update and implement this DMP?

Project PI Prof. Hans Dierckx