Jürgen Duchenne – FWO Data Management Plan

1. General information

"Name applicant"

Jürgen Duchenne

"FWO Project Number & Title"

12ZZN22N

"Affiliation"

KU Leuven

2. Data description

"Will you generate/collect new data and/or make use of existing data?"

Generate new 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)."

WP1 (Method validation): Data from this work package (WP) will consist of programming files, echo imaging files and data exports. The programming files include Matlab and Labview scripts, whereas imaging data will be stored in the default DICOM file type (circa 5-10 GB). Data exports encompass XML and TXT files (circa 100 MB). Data will be analysed in Microsoft Excel and statistical software such as SPSS (XLS and SAV files; estimated data volume: 200-500 MB). Data will be presented in Microsoft PowerPoint format (circa 0.5-1 GB in total).

WP2 (Animal experiments): Data concerning the animal experiments will consist of peri-procedural notes, data exports, echo and cardiac magnetic resonance (CMR) imaging files, as well as files for data analysis and presentation. Peri-procedural notes include written notes of the hemodynamic status of the animal during the experiments. After each experiment, these notes will be digitized in Microsoft Word/Excel files (estimated data volume 10-50 MB). Data exports come from pacemaker programming and pressure-volume catheterisation, and are stored in PDF and/or XML/TXT files (estimated data volume 50-100 MB). Imaging files will be stored in DICOM format (estimated data volume for echo: 260 GB; for CMR: 100 GB). Data will be analysed in Excel and SPSS (XLS and SAV files; estimated data volume: 200-500 MB). Data will be presented in PowerPoint format (circa 0.5-1 GB in total).

WP3 (Computer simulations): Data analysis files from WP2 will be used as reference for the computer simulations in WP3. WP3 data will consist of programming files and data exports/analysis/presentation files. Programming files will include Matlab scripts (circa 5-10 GB). Data exports include CSV and TXT files, analysis will be done with the help of XLS and SAV files (circa 100 MB). Data will be presented in PowerPoint format (total file size circa 0.5-1 GB in total).

WP4 (Clinical study): Data concerning patient inclusion and follow-up include XLS files with patient characteristics (circa 1-2 GB), PDF files of ECG's and pacemaker exports (circa 0.5-1 GB) and imaging files (in DICOM format; echo circa: 1 TB; CMR circa: 100 GB). Data will be analysed in Excel and SPSS (XLS and SAV files; estimated data volume: 200-500 MB). Data will be presented in PowerPoint format (circa 0.5-1 GB in total).

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 the file in KU Leuven's Privacy & ethical review tool (PRET). Be aware that registering the fact that you process personal data is a legal obligation."

Yes

Patient data will inherently be personal when they will be retrieved from in-hospital records. For study purposes, however, they will only be stored in pseudonymized manner.

"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. We will collect both data from animal experiments and patients. Ethical approval for the clinical part of the project is already obtained (UZ Leuven clinical ethical committee study nr S60439). The animal experiments have not been initiated yet, and an application to the KU Leuven animal ethical committee is currently being prepared.

"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?"

No

4. Documentation and metadata

Documentation and metadata should be created to describe and provide contextual information about the data, and thus aid their discovery, understanding and reuse.

Metadata ('data about data') refer to a highly structured way of describing data. Standards refer to standardised ways of metadating.

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

Digital data: All raw digital files will be labelled in a structured manner. Data from specific animals will be labelled with their unique Santinel number. Data from specific patients will be labelled with an

unique identifier: e.g. patient 1 will be labelled SWE_CRT_001, with SWE_CRT acting as a study identifier and 001 as a patient identifier. A TXT or Microsoft Word file with a clear description of what a specific folder contains will be saved in each specific folder.

Written data: Any written documentation on study design, data analysis, variable details and all information necessary for a secondary analyst to use the data accurately and effectively will gathered in notebooks which will be labelled with the investigators' name, title of the project and book number.

"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."

Medical images: All images used in our study will be saved in the conventionally-used DICOM format which contains metadata in the file header. This header contains key technical attributes of each specific file. In addition, each animal/patient will have a specific folder carrying the specific animal/patient code with all images.

Other medical data: Metadata of other medical data include e.g. naming and headers of PDF files with the ECG and pacemaker read-outs will be labelled with the animals'/patients' specific code.

Numeric data: Metadata of numeric data (both raw and processed) include e.g. any headers of CSV, XLS, TXT, SAV files used throughout the study. These will be created manually in a descriptive and structured manner.

5. Data storage and back up during FWO project

"Where will the data be stored?"

Medical images: All raw imaging data from patients is stored automatically on the hospitals' clinical servers. These are automatically back-upped. In addition, a copy will be stored on the server of our research group, which are also located within the hospital in Leuven and which are automatically back-upped on a daily basis. The server access subject to the access management of the university/hospital.

Medical images from the animal experiments and other medical and numeric data will also be stored on the server of our research group.

Everyday working files: Processed data, such as Excel sheets, CSV files etcetera will be stored on the personal virtual hard disks of the respective co-worker which are automatically back-upped on a daily basis.

Data on Paper: All written data on paper will be digitized and stored on the server described above. Originals will be kept in a locked cabinet for future reference only when required for legal reasons.

"How is back up of the data provided"

The server system of the hospital provides an automatic back-up.

"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. Sufficient back-up capacity on the hospital servers is currently available for the project. If needed, back-up capacity can always be increased.

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

Data storage capacity including backup on hospital servers is available for a yearly fee of 200 EUR/TB. We expect a total of close to 2 TB, in the end. Given the gradually increasing demands over the three years of the project, we foresee ca. 1000 EUR in total.

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

Access to data stored on the hospital servers is subject to the access rights management of the hospital. Access will be restricted to persons directly related to the study. Paper files containing study related information bill be locked in a separate cabinet within the Medical Imaging research Center with access for the PI and Co-PIs only.

6. Data preservation after the FWO project

FWO and KU Leuven expect that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.

"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 generated during the project will be retained for minimum 5 years after publication of the study findings.

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

For the first five years, digital data remain on the servers and paper data locked away at the research center as described above. After that period, a decision will be taken if a selection of data will be retained and if yes, which.

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

Data storage capacity including backup on hospital servers is available for a yearly fee of 200 EUR/TB. The expected costs are therefore 400 EUR/year, i.e. ca. 2000 EUR it total.

7. Data sharing and reuse

This section aims to identify which data will be made available to others for reuse after the project, where/how this data will be made available and under which conditions.

"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)?"

No

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

After the primary publication of the main research results, the full datasets will be made available to all collaborators on request.

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

Upon reasonable request, data will be shared for re-use by other research groups.

"When will the data be made available?"

After publication of the research results.

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

Data will be made available for other research groups, upon reasonable request and after publication of research results. Appropriate credit should be given to our research group.

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

None. Potential costs for data sharing will be billed to the requesting investigator.

8. Responsibilities

Clearly identify the responsibilities for data management in your project

"Who will be responsible for data documentation & metadata?"

The responsibility for data storage lies with the academic staff linked to the study, more specifically the assigned researcher Jürgen Duchenne, under supervision of the PI. Data backup is automatically provided by the hospital.

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

The responsibility for data storage and back-up lies with the academic staff linked to the study, under supervision of the PI.

"Who will be responsible for ensuring data preservation and reuse?"

The responsibility for data preservation and re-use lies with the academic staff linked to the study, under supervision of the PI.

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

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