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	*Ewout Heylen (ORCID: 0000-0003-3493-1471)*
Contributor name(s) (+ ORCID) & roles	*Frederik Maes (ORCID: 0000-0003-0027-1479) - Supervisor* *Robin Lemmens (ORCID: 0000-0002-4948-5956) - Supervisor*
Project number [1] & title	*Learning-based computational strategies for multimodal image analysis of neurovascular diseases*
Funder(s) GrantID [2]	*11Q5624N*
Affiliation(s)	*KU Leuven* *ROR identifier KU Leuven: 05f950310*
Please provide a short project description	*Stroke is one of the most prevalent neurological diseases and causes of disability worldwide. Two types with different origin can be discriminated: ischemic stroke caused by a blood clot, and hemorrhagic stroke caused by bleeding. My project focuses on ischemic stroke. We want to develop reliable functional outcome prediction models that incorporate imaging data, non-imaging data and the possible treatments as input variables, which could be very useful in guiding the medical staff by assessing the benefit-risk balance of different treatment options, allowing to select an optimal treatment strategy tailored to the individual patient.*

[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.

2. 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 [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
CRISP	*Data derived from the "CT perfusion to predict response to Recanalization in Ischemic Stroke Project" (CRISP) study, containing acute CT Perfusion, acute non-contrast CT, follow-up FLAIR and metadata, acquired at the Stanford Stroke Center (USA).*	*Reuse existing data*	*Digital*	*Images Numerical Textual*	*.dcm, .nii, .csv*	*< 1 TB*	
MRCLEAN	*Data derived from the "Multicenter*	*Reuse existing data*	*Digital*	*Images Numerical Textual*	*.dcm, .nii, .csv*	*< 1 TB*	

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

	Randomized CLinical trial of Endovascular treatment for Acute ischemic stroke in the Netherlands" (MR CLEAN) trial, containing acute CT Perfusion, acute non-contrast CT, follow-up non-contrast CT and metadata, acquired in the Netherlands (multicenter).						
KAROLINSKA	*An in-house dataset acquired at the Karolinska University Hospital (Sweden) within the scope of the "NExt generation X-ray Imaging System" (NEXIS) project,*	*Reuse existing data*	*Digital*	*Images Numerical Textual*	*.dcm, .nii, .csv*	*< 1 TB*	

	containing acute CT perfusion, acute non-contrast CT, follow-up dual energy CT and metadata.						
CRISP2	*Dataset acquired at the Stanford Stroke Center (USA) and at the University Hospital of Leuven with acute CT perfusion at baseline, acute DWI at baseline, follow-up imaging and metadata.*	*Reuse existing data*	*Digital*	*Images Numerical Textual*	*.dcm, .nii, .csv*	*< 5 TB*	
Validation dataset	*Data acquired at the University Hospital of Leuven: acute CT (perfusion), follow-up imaging data (CT and/or MRI) and metadata.*	*New data*	*Digital*	*Images Numerical Textual*	*.dcm, .nii, .csv*	*< 1 TB*	

Code	*Source code to process and analyze the data, written in Python and Mevislab.*	*New data*	*Digital*	*Textual*	*.py, .mlab*	*< 1 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.	*CRISP: doi: 10.1002/ana.24953.* *MRCLEAN: doi: 10.1056/NEJMoa1411587.* *KAROLINSKA: https://www.nexisproject.eu.* *CRISP2: https://med.stanford.edu/neurology/research/clinicaltrials.html.*
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: *CRISP: Ethical approval obtained locally.* *MRCLEAN: Central medical ethics committee and research board of Erasmus MC University Medical Center (MEC-2010-041).* *KAROLINSKA: Swedish Ethical Committee approval No. 2019-0048 and 2020-01391.* *CRISP2: Ethical approval obtained locally for patients included in the US. S-number for cases included at the University Hospital of Leuven: S64634.* *Validation dataset: The request for Ethical approval will be submitted before the start of work package 4.* ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information:

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) *CRISP: Ethical approval obtained locally.* *MRCLEAN: Ethical approval obtained locally.* *KAROLINSKA: Ethical approval obtained locally.* *CRISP2: Ethical approval obtained locally for cases included in the US. S-number for cases included at the University Hospital of Leuven: S64634.* *Validation dataset: The request for Ethical approval will be submitted before the start of work package 4.* ☐ 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: *My work could possibly result in software for stroke analysis, although this is not the goal of this PhD research, nor is it feasible within the timespan of my PhD.*
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: *Data transfer agreements with the data providing institution for CRISP, MRCLEAN, KAROLINSKA and CRISP2 define the usage of the data for projects related to stroke neuroimaging analysis.*
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:

[4] See Glossary Flemish Standard Data Management Plan

3. 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). *RDM guidance on documentation and metadata*.	*Imaging data:* *The datasets will contain a README.txt file explaining the structure of the dataset on the server if ambiguities are expected. This includes information about directories, processing, data formats and data description.* *Code:* *The code will be documented within the scripts itself.*
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: *Each dataset contains a .csv file where metadata is stored. However, the column names follow the standard naming of the institution providing the data.*

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)
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: *Back-up managed by IT manager of the Medical Imaging Research Center (Dominique Delaere).* ☐ 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*	*Data is stored on the protected servers of the Medical Imaging Research Center (MIRC), only accessible via the secured network of the UZ Leuven.*

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	*Data storage is already provided by the research group of the Medical Imaging Research Center (MIRC). The data servers and back-ups are managed by Dominique Delaere, IT manager at the MIRC.*

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):

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	*Data storage for data preservation is already provided by the research group of the Medical Imaging Research Center (MIRC). The data servers and back-ups are managed by Dominique Delaere, IT manager at the Medical Imaging Research Center (MIRC).*

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: *Imaging data:* *Data of the CRISP, MRCLEAN, KAROLINSKA and CRISP2 datasets are bound by data transfer agreements and can not be made public available without approval of the institution which provided the data. By default, these datasets are only available for researchers from the Department of Electrical Engineering (ESAT) and Department of Neurosciences Division of Experimental Neurology of the KU Leuven working directly with the data.* *Data of the validation dataset collected in the UZ Leuven will be restricted to institutional access only.* *Code:* *If relevant, code will be made available on GitLab/GitHub.*

If access is restricted, please specify who will be able to access the data and under what conditions.	*Imaging data:* *Only researchers from the Department of Electrical Engineering (ESAT) and Department of Neurosciences Division of Experimental Neurology of the KU Leuven working directly with the data have access to the data.* *Code:* *Access for researchers from the Department of Electrical Engineering (ESAT) and Department of Neurosciences Division of Experimental Neurology of the KU Leuven via GitLab/GitHub. If relevant, code can be made publicly available on GitLab/GitHub.*
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: *Imaging data:* *Data of the CRISP, MRCLEAN, KAROLINSKA and CRISP2 datasets are bound by a data transfer agreement. Sharing of data is not allowed without consent of the institution which provides the data.* *The validation dataset collected in the UZ Leuven will be restricted to institutional access due to ethical and privacy aspects.*

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) ☐ Other (specify) *Imaging data:* *All data will be stored and made available for institutional access on the data server of the MIRC, located at the SHARED folder.* *Code:* *Available on GitLab (repository: "micstroke").*
When will the data be made available?	☐ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify) *Imaging data:* *Data will not automatically be made available without approval of the providing institution due to privacy and ethical concerns.* *Code:* *When relevant, code can be made available on GitLab/GitHub after a project is finished and the results are published.*

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 license (code) ☐ GNU GPL-3.0 (code) ☒ Other (specify) *Imaging data:* *Data will not automatically be made available without approval of the providing institution.* *Code:* *When relevant, code can be made publicly available on GitLab/GitHub under MIT license or Apache License 2.0.*
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?	*Imaging data:* *As no data sharing is expected, there are no expected costs.* *Code:* *Code would be shared via already existing GitLab/GitHub accounts. These costs are already covered for.*

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Ewout Heylen
Who will manage data storage and backup during the research project?	Ewout Heylen, Dominique Delaere

Who will manage data preservation and sharing?	Ewout Heylen, Dominique Delaere
Who will update and implement this DMP?	Ewout Heylen