FWO DMP Template - Flemish Standard Data Management Plan

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	**Robin Lemmens 0000-0002-4948-5956**
Contributor name(s) (+ ORCID) & roles	
Project number[1] & title	1841923N
Funder(s) GrantID[2]	FWO
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	Stroke is a devastating disease and a leading cause of mortality and handicap over the world. In patients with ischemic stroke the obstruction of a blood vessel will result in neurological symptoms which can be temporarily if blood flow is restored in time, or persistent if this will not occur in time, resulting in brain ischemia. Identifying early ischemic changes and vessel occlusions is critical to select patients for endovascular therapy. Automated analyses of neuroimaging can assist physicians in the diagnostic pathway. The development of brain ischemia is time dependent with large individual variation in infarct growth. The size of the baseline infarct and the growth rate are predictors of the final infarct volume and clinical outcomes. To increase the number of patients eligible for endovascular stroke therapy and to improve the outcomes of patients undergoing this therapy, it is critical to develop treatments that slow infarct growth before blood flow can be restored. Individual prediction of baseline infarct and growth rate as proposed in this project will enable precision-based management of stroke patients for both reperfusion therapy and neuroprotective drugs. In addition, development of models beyond prediction of infarct growth, but focusing on functional outcome is needed. We aim to establish both tissue- and functional outcome clocks which can be implemented in clinical practice.

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[4].

				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
Objective 1	Validation of RAPID ASPECTS	☐ Generate new data ☐ Reuse existing data - Retrospective collection of 150 NCCT scans from UZ LEUVEN patients (S62526)	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv (for clinical data) ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM (Brain imaging) ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Objective 2	Congruence between NCCT and CTP	☒ Generate new data - Retrospective dataset UZLEUVEN 87 patients (S65359)	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software	☐ .por ☐ .xml ☐ .tab ☒ .csv (clinical data) ☐ .pdf ☐ .txt	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB	

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

		☒ Reuse existing data - DEFUSE 3 *DOI:10.1056/NEJMoa1713973*		☐ Other ☐ NA	☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM ☐ NA	☐ > 50 TB ☐ NA	
Objective 3	rCBF time	☐ Generate new data ☒ Reuse existing data - CRISP *doi:10.1002/ana.24953* - EVAS registry (Karolinska University) *Swedish ethical committee No. 2019-0048;2020-01391*	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv (clinical) ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM (imaging) ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Objective 4	MEVO detection	☒ Generate new data ☐ Reuse existing data - Retrospective collection of data from UZ	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software	☐ .por ☐ .xml ☐ .tab ☒ .csv (clinical) ☐ .pdf ☐ .txt	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB	

		LEUVEN, 400 patients. (S64846)		☐ Other ☐ NA	☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM (imaging) ☐ NA	☐ < 50 TB ☐ > 50 TB ☐ NA	
Objective 5-7-8	Machine learning	☐ Generate new data ☒ Reuse existing data - MRCLEAN *DOI:10.1056/NEJMoa1411587* - CRISP *doi:10.1002/ana.24953* - EVAS registry *Swedish ethical committee No. 2019-0048;2020-01391* - BASICS *DOI:10.1056/NEJMoa2030297* - MRCLEAN registry *doi:10.1136/bmj.k949*	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv (clinical) ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM (imaging) ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

		- BATMAN *DOI:10.1161/STROKEAHA.116.015492*					
Objective 6	Infarct growth prediction	☒ Generate new data ☐ Reuse existing data Prospective data collection of patients transferred to undergo thrombectomy in Stanford and UZLeuven. (S64634)	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv (for clinical data) ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: DICOM (Brain imaging) ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

Guidance: *Data can be digital or physical (for example biobank, biological samples, …). Data type: Data are often grouped by type (observational, experimental etc.), format and/or collection/generation method.* *Examples of data types: observational (e.g. survey results, sensor readings, sensory observations); experimental (e.g. microscopy, spectroscopy, chromatograms, gene sequences); compiled/aggregated data[5] (e.g. text & data mining, derived variables, 3D modelling); simulation data (e.g. climate models); software, etc.* *Examples of data formats: tabular data (.por,. spss, structured text or mark-up file XML, .tab, .csv), textual data (.rtf, .xml, .txt), geospatial data (.dwg,. GML, ..), image data, audio data, video data, documentation & computational script.* *digital data volume: Please estimate the upper limit of the volume of the data per dataset or data type.* *physical volume: Please estimate the physical volume of the research materials (for example the number of relevant biological samples that need to be stored and preserved during the project and/or after).*	
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.	Ok. (cfr previous table)

[5] These data are generated by combining multiple existing datasets.

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, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: These projects involve clinical and imaging data from patients. All projects are approved by the corresponding ethical committees. - Existing data (ethical approvals can be found in the original publications) - New data from UZ Leuven: Objective 1: S62526 Objective 2: S65359 Objective 4: S64846 Objective 6: S64634 + approval from the Stanford ethical committee.

Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes ☐ No If yes: Personal data - Short description of the kind of personal data that will be used: We will collect clinical data (basic variables, neurological examination, vital parameters, medical history, functional outcome) and brain imaging details (initial and follow-up). Personal data relating to study participants including UZ Leuven patient number will be collected for ID purposes during data collection. This information will only be available to researchers directly involved in recruitment of patients (only for the prospective CRISP2 study). All further derivative data, including imaging data, will be coded, and thus pseudonymised. The file linking the code and personal identifiers will only be accessible to authorized individuals and stored in a restricted access, secure environment managed by the KU Leuven/UZ Leuven ICT facility. - Privacy Registry Reference: S-nrs above
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:
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: For existing data from the CRISP, MRCLEAN, BATMAN, BASICS, MRCLEAN registry and EVAS registry, the data is only intended for the described study protocol.

[6] See Glossary Flemish Standard Data Management Plan

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:

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).	As mentioned earlier for many of the projects we will use already existing data that is saved in secured and locked .csv files that come with a code explanation file (readmefile) that can be found in the same folder as the dataset. The abbreviations that are used are as much as possible standard ones that are used in the literature as well. For the CRISP2 prospective study, study data will be collected in REDCAP (a database software that is metadata driven). This information will be transferred to .csv files that are locked when the inclusion of the study is finished. Also for this dataset we will use standardized abbreviations and will keep a code explanation file in the same folder. Working copies from the datasets can be created for processing and analysis.

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: For the CRISP2 study (objective 6) we will use REDCAP (research electronic data capture) that is already metadata driven. Radiological data for this project will be stored on passport protected encrypted computers at the Stanford Stroke Center. For the retrospective data collected at UZLeuven there is no existing metadata available. For the other mentioned research objectives we will use already existing datasets. If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	UZ Leuven servers and for project 2 also protected servers at Stanford Stroke Center. Copies can be made and kept on personal devices in accordance with the level of authorization of the user and the data security level of their device. Sensitive personal data concerning the study participants will be stored in a KUL/UZ secure environment.

How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.*[7] *REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.*	UZ Leuven datacenter policy: redundancy in storage + backups.
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 ☐ No If yes, please specify concisely: Currently we have a back-up capacity for the data on the server of UZ Leuven which is expected to be large enough. If necessary, in the future, additional storage will be obtained, for which we will have budget available. 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. 7	Data will be handled in accordance with KU Leuven information security guidelines, involving measures proportionate to their nature and the risks involved. Lab computers and external drives will be password-protected, and the rooms in which they are kept will be locked when no lab members are present. Read-only permissions will be implemented for raw data files so that they cannot be edited.

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	We estimate around 1500 euros per year for 5-10TB. These costs will be covered by the research budgets for every project.

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...).	All data will be retained on the servers for a period of at least 5 years.
Where will these data be archived (stored and curated for the long-term)?	UZ Leuven servers and for project 2 also protected servers at Stanford Stroke Center.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Around 1500 euros per year. Payed from FWO funding.

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, in an Open Access repository ☒ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	- For the new datasets from UZLeuven we will share data with external collaborators on up reasonable request.
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: For the already existing datasets: CRISP, MRCLEAN, BATMAN, BASICS, MRCLEAN registry and EVAS registry we cannot share data due to contractual restrictions.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Publications will be made available through Lirias. We will share the data with scientific collaborators upon reasonable request. After signing of a data sharing agreement.

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	Upon publication of research results.
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.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [8]	For the moment there is no license, but this will be further explored with our collaborators
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 ☒ No If yes:
What are the expected costs for data sharing? How will these costs be covered?	Those costs are not expected to be high, since most data cannot be shared. Possible costs will be payed by FWO funding.

Met opmerkingen [A1]: Weet jij wat er kan voor de UZLeuven en CRISP2 data?

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Anke Wouters, Jeroen Bertels.**
Who will manage data storage and backup during the research project?	**Robin Lemmens for UZ Leuven and Soren Christensen at Stanford university.**
Who will manage data preservation and sharing?	**Robin Lemmens**
Who will update and implement this DMP?	**Robin Lemmens**