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	Guy Boeckxstaens 0000-0001-8267-5797
Contributor name(s) (+ ORCID) & roles	Heiko De Schepper, co-PI
Project number[1] & title	T003722N
Funder(s) GrantID[2]	
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	The irritable bowel syndrome (IBS) is a chronic gastrointestinal disorder characterized by recurrent abdominal pain and is associated with abnormal stool form or frequency. The prevalence of IBS strongly varies depending on the diagnostic criteria used and is estimated to affect 5-20% of the general population. Despite the high prevalence of IBS and its significant impact on the quality of life (QoL), insight into the underlying disease mechanisms is rather poor. According to the International Foundation for Functional GI Disorders, the economic burden of IBS is estimated at US$21 billion per year in the US. In the UK, costs were estimated in 2012-2013 to be approximately 12 billion pounds. Hence, there is a great unmet need to improve clinical management of this large patient population, in particular with drugs targeting abdominal pain. Based on our previous research, we identified histamine as a major target for treatment, a finding that was validated in a clinical proof-of-concept study showing clinical efficacy of the histamine 1 receptor (H1R) antagonist ebastine as treatment of IBS. To further validate this novel approach, we designed a placebo-controlled randomized multicenter clinical trial comparing the efficacy of ebastine with the spasmolytic mebeverine, the current first-line treatment for IBS. In addition, we will study the effect of both treatments on quality of life and quality-adjusted life years of IBS patients.

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
Symptom questionnaires	Pain scores, stool consistency, quality of life	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☒ .xml ☐ .tab ☒ .csv ☒ .pdf ☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: SAS (*.sas7bdat) ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	NA

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

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.	NA
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:

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

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: As stated in the GDPR conformity questionnaire of UZ Leuven, the following categories of personal data will be used: **Regular personal data** Name and/or address, Contact details (tel. number, e-mail address,…) , Date/year of birth and/or age, Initials, personal identification number assigned to data subjects participating in the study such as EAD number. **Special/sensitive categories of personal data** Health data (e.g. description of characteristics of physical features of the body, medical history and medical test information (such as blood samples, results from scans, and biopsies)). The medical information will include questionnaire data regarding gastrointestinal symptoms, stool consistency and defecation frequency, quality of life, and reporting of adverse events. **Privacy Registry Reference** S67029 is the project-specific number of UZ Leuven under which the GDPR conformity questionnaire has been approved.
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:

[6] See Glossary Flemish Standard Data Management Plan

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:
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).	All data will be collected via an electronic CRF in REDCAP. REDCap allows for following output formats: Microsoft Excel, SAS, Stata, R, or SPSS for 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: REDCap allows for following output formats: Microsoft Excel, SAS, Stata, R, or SPSS for analysis 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?	All patient data are encoded, i.e. each patient receives a unique code in order to pseudo-anonymize the collected information (code table is managed by prof Boeckxstaens, no access is granted to other staff). Files are stored on the J drive of KU Leuven which is protected via a central login for KUL personnel. The file containing the link between the CRF number and the patient's identity is saved in a password protected file on J-drive and which is only accessible by the PI and trial nurse. In case data are stored on managed laptops, the hard drive is encrypted.

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.*	"J-drive" servers at KU Leuven are accessible only by laboratory members, and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.
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: There is sufficient storage and backup capacity on all KU Leuven servers: the "J-drive" is based on a cluster of NetApp FAS8040 controllers with an Ontap 9.1P9 operating system. If no, please specify:

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

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]	To ensure secure, efficient and clean data collection, REDCap is used. For randomized study drug allocation, Randomize.net will be used. These software packages both are user ID and password protected; access to the software can be configured on a per-study and per-user basis. Validation checks in REDCap (e.g. range checks) minimize data entry errors and ensure data consistency. All data crucial for verification of research results are kept in a secured online lab notebook. All patient data are encoded, i.e. each patient receives a unique code in order to pseudo-anonymize the collected information (code table is managed by prof Boeckxstaens, no access is granted to other staff). Files are stored on the J drive of KU Leuven which is protected via a central login for KUL personnel. The file containing the link between the CRF number and the patient's identity is saved in a password protected file on J-drive and which is only accessible by the PI and trial nurse. In case data are stored on managed laptops, the hard drive is encrypted. The questionnaires will collect information on demographics, previous history, psychological factors, and symptoms. The strategy to protect the privacy of individuals has been approved by the Ethics Committee (EC).
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The expected costs during the study will be €104,42 per Terabyte (TB) and per year (per blocks of 5 TB!). These costs will be covered by the FWO-TBM grant that supports this clinical trial.

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 for at least 5 years and conform the Research Data Management (RDM) policy of KU Leuven. The RDM policy could be found here: https://www.kuleuven.be/rdm/en/policy
Where will these data be archived (stored and curated for the long-term)?	Data will be archived after 5 years on the K drive of KU Leuven devoted for long term storage. Permission to access the K drive is limited to the PI and only for read-only use. Moreover, the data will then be hard-locked thus introducing changes to the data is not allowed anymore.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	These costs for archiving are expected to be €52,21 per TB and per year. These costs will be covered by the FWO-TBM grant that supports this clinical trial.

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

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]	None
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: The Digital Object Identifier (DOI) is unknown at the moment of writing and might be added to the study upon eventual publication.
What are the expected costs for data sharing? How will these costs be covered?	NA

[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?	G. Boeckxstaens
Who will manage data storage and backup during the research project?	G. Boeckxstaens
Who will manage data preservation and sharing?	G. Boeckxstaens
Who will update and implement this DMP?	G. Boeckxstaens