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	**Prof. GIANLUCA MATTEOLI 0000-0002-2902-4976**
Contributor name(s) (+ ORCID) & roles	Co-promoter: Dr**.** Marlene Hao 0000-0002-9701-8252 Co-promoter: Dr. Lincon Stamp 0000-0002-8925-7894
Project number & title	GPUM/22/020 Glial-immune interactions in the gastrointestinal tract: Protection against inflammation and cancer.
Funder(s) GrantID[2]	BOF - Project ID: 3M220072
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☒ University of Melbourne Provide ROR[3] identifier when possible:

These data are generated by combining multiple existing datasets.
See Glossary Flemish Standard Data Management Plan
Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/
Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

[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	Intestinal immune cells are frequently exposed to antigens from the diet, commensal microbiota, and pathogens, therefore tight regulation of the balance between immune activation and tolerance is essential to maintain intestinal homeostasis. Failure to maintain this equilibrium may lead to inflammatory bowel disease (IBD), a chronic inflammatory disease of the gastrointestinal tract. Recent studies show that enteric glial cells (EGCs) are potent modulators of immune cell functions, but the numerous factors and molecular pathways involved in the EGC-immune interaction and its relevance in IBD are not yet well investigated. In this project, we aim to identify EGC-derived factors, that modulate immune functions and thereby contribute to resolving inflammation and maintaining homeostasis in the gut. Moreover, EGCs are also known to play an important role in colorectal cancer (CRC), which is a major cause of death in IBD patients and one of the most common malignancies, ranking third in the world with respect to the number of patients affected in their lifetime. Studies have reported that apart from epithelial malignant cells, the colonic tumor microenvironment is populated by resident tissue cells including stromal and endothelial cells together with recruited immune cells such as tumor-associated macrophages (TAMs). This peculiarity of the tumor microenvironment supports tumor progression, and could lead metastasis and resistance to antitumor therapies. Thus, we also aim to investigate the molecular mechanisms involved in the crosstalk between EGCs and other cells in the CRC tumor microenvironment. To this end, we will use bioinformatic approaches, novel multicellular culture techniques, molecular and cell biology techniques, and EGC-specific transgenic mouse models. Overall, identification of factors and mechanisms underlying EGC-immune crosstalk during homeostasis and intestinal inflammation will provide new insights on novel therapeutic targets for treating intestinal immune-mediated diseases. Moreover, identifying novel factors and mechanisms that enhance drug resistance by contributing to CRC progression will help design novel strategies for superior therapeutic interventions in CRC patients.

1. 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[1].

				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
RNA-sequencing raw data, processing and analysis files	RNA-Sequencing data from murine and human samples	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .tab ☒ .csv ☒ .pdf ☒ .txt ☒ fastq ☐ .dwg ☐ .tab ☒ .bam ☒ other: .doc, .jp, .tiff, .rds ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Spreadsheets	Created from observational data collected during murine experiments, patient clinical information	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .xml ☒ .csv ☒ .pdf ☒ .txt	☒ < 1 GB	

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

Flow cytometry	Raw data, processing and analysis files	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff ☒ .fcs ☒ .xls ☒ .pdf	☒ < 100 GB	
PCR	PCR raw data, processing and analysis files of murine and human samples	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff	☒ < 2 GB	
Immunohisto chemistry images and histology	High resolution images obtained from immunohistoch emistry and histology experiments	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff	☒ < 2 GB	
DSS colitis (murine experiments)	Pictures result fecal occult blood and disease parameters	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff ☒ .xls	☒ < 2 GB	
Ussing system	Experimental data and analysis files	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .xls	☒ < 2 GB	
ELISA raw data and analysis	All experimental data and analysis files	☒ Generate new data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental	☒ .tiff ☒ .xls	☒ < 2 GB	

		☐ Reuse existing data		☒ Compiled/ aggregated data			
Manuscripts	text files	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	Text file	☒ .doc ☒ .pdf ☒ .txt	☒ < 2 GB	

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 (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.	The data RNA-seq data that are already published will be downloaded from Gene Expression Omnibus (GEO) database https://www.ncbi.nlm.nih.gov/geo/

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: - **Medical Ethics Committee UZ Leuven** Human patient samples from full-thickness biopsies from the resected ileum of Crohn's disease (CD) patients undergoing curative intent surgery for fibro-stenotic ileal strictures and from patients undergoing curative intent right hemicolectomy for colon carcinoma (CRC) will be gathered after informed consent by the IBD- UZ Leuven group under supervision of prof. Dr. Severine Vermeire. Tissue samples representative of non-affected mucosa (resected margins), inflamed and stenotic areas will be identified under the supervision of a trained IBD pathologist (UZ Leuven) and digested to generate single cell suspensions for sc-RNA seq and for isolation of stromal and immune cells. Healthy and ileal full thickness biopsies will be collected during right hemicolectomy for CRC and will be used as control. S-number: S53684 Human CRC samples taken in this project form an integral part of the project entitle: Aanleggen van een colorectaal carcinoom biobank voor wetenschappelijk Onderzoek; Main supervisor: prof. Dr Sabine Tejpar, for which ethical approval has already been granted (S-number: S50887) and frequent renewals are submitted. Overall, health data such as disease severity and disease status of the collected samples will be recorded. - **Animal Ethics Committee (ECD) KU Leuven** Animal experiments will be performed as part of this project. They have been approved by the Ethical committee for Animal Experimentation (ECD) at KU Leuven. Ethical committee number: P213-2018

Will you process personal data? 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: - Short description of the kind of personal data that will be used: We start from non-anonymized patient data. After inclusion (signing of informed consent), data are anonymized and each patient is from then on only identifiable by a unique number. Only the treating physician holds the code and is able to go back to the patient chart, if particular results obtained in the project would be of such medical importance that the patient's health or disease status could be affected by not sharing this information. This is also clearly stipulated in the informed consent form. We will further stay in contact with Toon Boon for optimizing the strategy to deal with these personal data. - Privacy Registry Reference: Ethical committee numbers are: S50887, S53684
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: There might be IP depending on the obtained results. This may involve identification of biomarkers which may predict treatment response or molecules which may have a therapeutic role. In that case, LRD (KU Leuven) will be contacted.
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

2. 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).	Protocols and details related to data collection and processing will be recorded in Word or Excel files. Data folders containing raw and processed data will be hierarchically organized and labelled based on the source of the data, the type of experiment, the date of data generation, and the different experimental conditions analysed. Data analysis methods and particularities (including metadata) will be described in text documents and Excel files included in these folders. All files will be stored in the J-Drive or L drive (KU Leven Storage space). A read-me file containing lab notes, SOPs of animal models and protocols to process the tissue, SOPs of human data and processing of human tissue will always be kept together with the dataset.
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: Metadata standards will be used for proteomics (http://www.dcc.ac.uk/resources/implementations/pride-proteomics-identifications-database), genomics (http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata) . For transcriptomics, metadata will be created using the Dublin core (http://www.dcc.ac.uk/resources/metadata standards/dublin-core). Text documents and Excel files stored within each experiment folder in the J-Drive or L-Drive will respectively contain guidelines describing data collection/analysis methods and all relevant metadata (including experimental conditions, sample keys, computational analysis pipelines and their parameters) to ensure the reusability of the data and the reproducibility of any further data generation.

3. Data Storage & Back-up during the Research Project	
Where will the data be stored?	Upon data collection/pre-processing, data will be stored in the J-Drive or L-Drive of our research unit. These servers are centrally managed by ICTS KU Leuven and have back-up capacities (KU Leuven enterprise box, Large volume-storage). Temporary copies of the data will be made and kept on personal hard drives if necessary.
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.* *Refer to institution-specific policies regarding backup procedures when appropriate.*	Data stored on the KU Leuven J-Drive and L-Drive is managed, maintained, and backed up by KU Leuven IT Services (ICTS). Specifically, mirror copies of the stored data are made immediately upon upload for safety backup purposes.
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: Yes, the KU Leuven J-Drive and L-Drive has sufficient storage capacity for the outlined project. 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 stored on KU Leuven-managed personal computers are protected via password access to the computers, as set up by the KU Leuven IT Department. Off-site access to J-Drive and L-Drive data is available from KU Leuven personal computers and data access points and is password protected. Upon request, access to the shared drive will be given only to authorized researchers.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The annual cost of J-Drive storage is €519 and L-Drive is €113.84 per 1TB of storage space per year. This cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. We expect that 5 TB will be sufficient to store all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Gianluca Matteoli).

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 raw data will be retained for at least 5 years on the K-Drive storage space. Publication data will be further organized and catalogued on a figure-by-figure basis for future reference to raw datasets used for figure generation. As stated in the ICF, all data are anonymized. These will be also preserved for the 5-year period after end of this project. The generated data will be stored on designated KUL servers. Only in case of findings which have direct implications for the health or disease status of the patient, the principal investigator has the code to go back to the original patient, as also stipulated in the ICF. This code is stored on the hospital UZ server which is password protected and which also allows to consult the electronic medical chart of the patient stored on UZ Leuven Hospital servers.
Where will these data be archived (stored and curated for the long-term)?	Long term data archives will be maintained in specific archive folders on the K-Drive. With the server centrally managed by the ICTS, we will use the back-up possibilities as proposed by KU Leuven ICTS.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The annual cost of K-Drive storage is €5.69 per 100GB of storage space per year. We expect that 5 TB will be sufficient for long-term storage of all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Gianluca Matteoli).

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: All the data that will be published as a part of the project will be made available.
If access is restricted, please specify who will be able to access the data and under what conditions.	Data not deposited in open-access repositories will in principle only be accessible to members of Prof. Matteoli's lab. Other collaborations and sharing are possible with staff within the Inflammatory Bowel Diseases research group at TARGID, upon reasonable request. Any user can place reasonable requests data for non-commercial purposes, and these requests will be assessed on a case-by-case basis by the project lead (Prof. Gianluca Matteoli). Commercial-based requests will be discussed with the project lead as well.
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

Where will the data be made available? If already known, please provide a repository per dataset or data type.	-In an Open Access repository. -Biobank tissue samples via the Biobank -Experimental data will be made available through a data repository such as Genebank, FigShare (https://figshare.com/), Dryad (https://datadryad.org/) or https://zenodo.org/ depending on the type of data. We will explore the possibilities via online repositories and will use the website www.re3data.org.)
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 the research results, the data will be made available via the required links in the publication or upon request, and after an embargo period after publication (for example: phenotype files, genetic data).
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."* [4]	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.

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

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: Not available at the moment
What are the expected costs for data sharing? How will these costs be covered?	The annual cost of J-Drive storage is €519 and L-Drive is €113.84 per 1TB of storage space per year. The annual cost of K-Drive storage is 5.69 € per 100GB of storage space per year. Expected amount of data 5 Tb. Digital vault for private data: Windows server (KU Leuven ICTS): 1302 €/year, Linux server (KU Leuven ICTS): 1278.40 €/year. The cost for data sharing will be discussed with collaborators depending upon the data repository selected on a case-by-case basis.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Prof. Gianluca Matteoli who is the project lead will be responsible for data documentation and metadata, generation/preservation of the project.
Who will manage data storage and backup during the research project?	The project lead will be responsible for collecting/generating data and for correct documentation and upload onto the L/J/K-Drive storage space and KU Leuven enterprise box. The KU Leuven IT department will be responsible for maintenance and back up of data storage spaces.
Who will manage data preservation and sharing?	The project lead will bear responsibility for ensuring data preservation and reuse.
Who will update and implement this DMP?	The project lead bears the end responsibility of updating & implementing this DMP.