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: Prof. Sven Wehner 0000-0002-8632-7631 Full Professor for Immunepathophysiology, Department of Surgery, University Hospital Bonn, Germany
Project number[1] & title	G049425N Unveiling the Synergy of Enteric Glia and Immune Cells in Resolving Intestinal Inflammation.
Funder(s) GrantID[2]	FWO-DFG Research project Weave (FWO acts as Lead Agency) - senior
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel z ☐ Other: Provide ROR[3] identifier when possible:
Please provide a short project description	Inflammation is essential for tissue repair but must be tightly regulated—especially in the intestine, where immune cells are constantly exposed to foreign antigens. Disruption of this balance can lead to diseases such as IBD and food allergies. Our labs have identified a role for the enteric nervous system in modulating intestinal immunity. Preliminary data using co-culture systems and glia-specific mouse models suggest that enteric glial cells (EGCs) influence immune responses and promote inflammation resolution. We hypothesize that enteric glia contribute to the control and resolution of intestinal inflammation by educating mucosal immune cells towards immunological tolerance. This project will explore glia–immune communication to uncover new therapeutic strategies for gut immune-mediated diseases.

[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/

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
snRNA-sequencing raw data, processing and analysis files	snRNA-Sequencing data from murine samples. Data will be processed using open source Bioinformatic tools	☒ 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, .mtx, .tsv ☒ other: .doc, .jp, .tiff, .rds ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Spatial transcriptomic raw data, processing	Spatial transcriptomic data from murine and human samples	☒ Generate new data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☐ .por ☐ .tab ☒ .csv ☒ .pdf ☒ .txt	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB	

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

and analysis files				☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ fastq ☐ .dwg ☐ .tab ☒ .bam, .mtx, .tsv ☒ other: .doc, .jp, .tiff, .rds ☐ NA	☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Spreadsheets	Created from observational data collected during murine experiments, patient clinical information, calculations of flow cytometry and qPCR data	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .xml ☒ .csv ☒ .pdf ☒ .txt	☒ < 1 GB	
Flow cytometry	Raw data, processing and analysis files. Flow Jo, Excel, GraphPad Prism will be used for analysis	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff ☒ .fcs ☒ .xlsx ☒ .pdf ☒ .wsp ☒ .prism	☒ < 100 GB	
qPCR	qPCR raw data, processing and analysis files of	☒ Generate new data	☒ Digital ☐ Physical	☒ Observational ☒ Experimental	☒ .jpg ☒ .tiff ☒ .xlsx	☒ < 2 GB	

	murine and human samples. Tissue stored in -80°C.	☐ Reuse existing data		☒ Compiled/ aggregated data	☒ .txt ☒ .LC96P ☒ .prism		
Immunohistochemistry and immunofluorescent images and histology, Confocal/Multiphoton microscopy images	High-resolution images obtained from immunohistochemistry and immunofluorescence and histology experiments. These will be analyzed using tools such as FIJI, Zeiss Zen Software, Excel, Graphpad Prism. Slides/tissue stored in 4°C	☒ 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 activity parameters. GraphPad Prism and excel will be used for analysis	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Observational ☒ Experimental ☒ Compiled/ aggregated data	☒ .jpg ☒ .tiff ☒ .xlsx ☒ .pdf ☒ .prism	☒ < 2 GB	

Cells and tissue samples from mouse and human	Intestinal glial cells, fibroblasts, epithelial cells, bone marrow-derived macrophages, colon tissue samples from patients or mice. Cultured in 2D or 3D at 37° C, 5% CO2 incubator, and freezers/liquid nitrogen will be used for storage. The respective centre responsible for collection and storage of human body material (UZ Leuven Biobank) and next generation sequencing (Genomics Core	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data			

	UZ Leuven) have implemented quality manuals and quality control procedures available.						
Patient Data and personal information	The patient's name or other identifying information will be stored separately (site file) from their study data and replaced with a unique code to keep the patient's identity secret. All personally identifiable information is replaced by a unique study identifier for each patient and each biological sample. The	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☒ Observational ☐ Experimental ☒ Compiled/ aggregated data	☒ .xlsx ☒ .pdf	☒ < 10 GB	

	subject identification log, linking the identifiers to the participant's name, will be only kept in the site's regulatory binder and only accessible by the treating physicians.						
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[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.

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

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 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** The KUL lab has already received approval by the Ethics Committee Research UZ / KU Leuven (reference number S64914 and S69568) for collection and processing of tissue samples and personal data. These biological samples are being collected under the umbrella biobanking project s53684 (CCare PI: Prof. S; Vermeire). 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 lysed to generate single nuclei suspensions for snRNA-seq or embedded to perform immunohistochemistry, immunofluorescence or spatial transcriptomics. In addition, healthy and ileal full thickness biopsies will be collected during right hemicolectomy for CRC and will be used as control. 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. These experiments are included in an already approved project by the Ethical Committee for Animal Experimentation (ECD) at KU Leuven: project license number P250/2024 and 249/2024. We will ensure this is valid throughout the project and it will be renewed if necessary/upon expiry.
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: - Short description of the kind of personal data that will be used: We start with non-anonymized patient data. After inclusion (signing of informed consent), data are pseudanonymized, and each patient is from then on only identifiable by a unique number. The patient's name or other identifying information will be stored separately (site file) from their study data and replaced with a unique code to keep the patient's identity secret. All personally identifiable information is replaced by a unique study identifier for each patient and each biological sample. The subject identification log, linking the identifiers to the participant's name, will be only kept in the site's regulatory binder and only accessible by the treating physicians. Only the treating physician holds the code and can go back to the patient chart if 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: S53684 (frequent renewals will be submitted)
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.

[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).	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 readme 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. The respective centre responsible for collection and storage of human body material (UZ Leuven Biobank) and next-generation sequencing (Genomics Core UZ Leuven) have implemented quality manuals and quality control procedures available. Readme files with information on methodology, an overview of the dataset, data sources, data collection methods, codebooks, data cleaning, analysis, variable definitions, and, units of measurement will be provided. If any data transformation steps or workflows were utilized during the analysis, these would be also documented and shared. Detailed documentation of the analysis methodology would be provided, including the statistical techniques, models, or algorithms employed. We will use repositories in GitLab and GitHub which will also ensure reuse and version control. The provenance of the

	data be thoroughly documented using the appropriate standards. The data will be accompanied by context on how they are captured, processed, analysed, and validated and other information that enables interpretation and re-use. All personally identifiable information is replaced by a unique study identifier/ Persistent Identifier for each patient and each biological sample. The subject identification log, linking the identifiers to the participant's name, will be only kept in the site's regulatory binder and only accessible by the treating physicians (Prof. Séverine Vermeire, UZ Leuven). This log will allow proper monitoring of the clinical records and by the appointed monitors to assure completeness, accuracy and correctness of the data collected at the site. Clinical reports necessary for remote source data verification will be de-identified, coded with the personal identifiers and electronically transmitted to the partners. No personally identifiable information will be at any time shared. Deposited datasets will be identified with a persistent identifier when applicable, such as a DOI, both in preprint and publication. Metadata that will be created includes but is not limited to image information, protocols, device information, data units, species taxon identifiers and gene identifiers. We will use the Research Data Repository (RDR) of KU Leuven (https://www.kuleuven.be/rdm/en/rdr) or European Genome-phenome Archive (EGA)(https://ega-archive.org/). A metadata standard is automatically applied upon depositing the data. The metadata model will include fields that are required, recommended and optional. Using this data repository, the data sets will be findable and reusable. We will follow the applicable standard FAIR principles, Findable-Accessible-Interoperable-Reusable, mentioned in fairsharing.org. Keywords are provided through the RDR repository system and EGA repository system. The assignation of keywords to each dataset and subsets will facilitate their re-use. We will try to balance the trade-off between keywords that are very specific to increase findability, but also more generalized so that it can be found in broader searches. After uploading a file, the RDR repository gives the option to add a ReadMe file and add 'file metadata' to each file. The metadata in RDR is always publicly available after the reviewer has approved the dataset, even with restricted or embargoed files. Metadata makes our dataset findable and is therefore essential to make our data FAIR. The metadata deposited in EGA can also be harvested and indexed. Upon the completion of the project, the data will be made freely available in the public domain with restricted access for data from patients and human tissues. For storage and sharing, we will ensure compliance with GDPR, and will use trusted European repositories such as RDR or EGA. Metadata will be

	made openly available and licensed under a public domain dedication open licenses (Creative Commons Attribution International Public License (CC BY)).
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: We will use the RDR (https://www.kuleuven.be/rdm/en/rdr) data repository of KU Leuven or European Genome-phenome Archive (EGA; (https://ega-archive.org/). /) and we will restrict access for data from patients and human tissue. We will contact the data access committee (dac@uzleuven.be) before depositing. At the end of the project, the data may be also deposited in domain specific domain-specific repositories such as Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). RDR uses digital object identifier (DOI). EGA also provides the submitters with stable, globally unique identifiers to enable reference of datasets in publications and across genomics infrastructures. A metadata standard is automatically applied upon depositing the data. We will follow community-endorsed interoperability best practices that are relevant to our discipline. The FAIR (Findable, Accessible, Interoperable, and Reusable) principles will be followed to make data findable using persistent identifiers, which will ensure data accessibility and reusability with clear and standardized metadata, through open licenses and data formats (https://fairsharing.org/) Example: Sequencing: MINISEQE (https://fairsharing.org/MINSEQE 2) Flow cytometry: MIFlowCyt (https://fairsharing.org/MIFlowCyt) Immunohistochemistry: OME (https://fairsharing.org/OME) Spatial trancriptomics: Matrix Market (.mtx) format The metadata model will include fields that are required, recommended, and optional. Using this data repository, the data sets will be findable and reusable. All descriptors will be given in a language that can be understood across multiple disciplines to reach the broader scientific community. Data will be deposited in a standard format that is commonly accepted and able which can be accessible to everyone. We will use shared vocabularies and ontologies to promote data interoperability. In the case of usage of uncommon or generated project-specific ontologies or vocabularies, clear descriptions and easily accessible mappings to more commonly used ontologies will be provided. The

	generated ontologies or vocabularies will be openly published to allow reusing, refining or extending them. The references will also be included wherever applicable in a standardized and acceptable manner. 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.

4. 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. Metadata will be made openly available and licensed under a public domain dedication open licenses (Creative Commons Attribution International Public License (CC BY)). Both EGA and RDR ensures long term storage. KU Leuven RDR ensures storage for at least 10 years. However, it is possible to request that the data be kept for a longer period of time (such as 20-25 years for medical fields). FAIR principle A2 states that metadata should persist even when the data are no longer sustained. In addition, data will be available in easily accessible form, e.g., pdf, excel files. Relevant software and coding details will made available via the use of repository such as GitLab and GitHub, which will also ensure reuse and version control.
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.*	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. Both EGA and RDR ensures long term storage. KU Leuven RDR ensures storage for at least 10 years. However, it is possible to request that the data be kept for a longer period of time.
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?	The EGA provides a web-based portal and an API to enable data controllers to manage data access and transfer permissions. Authorization infrastructure (AAI) management is vital for operating the EGA, where a user's identity is verified before confirming access rights to particular information. When a researcher logs

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

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]	into the EGA website to access sensitive metadata, their credentials are authenticated. They can then browse a dataset of interest and request the download of sensitive data. If the researcher is permitted to access the dataset after the EGA validates their request, the request is approved, and they can retrieve the data. Regarding the data of patients or human tissue samples, the patient's name or other identifying information will be stored separately (site file) from their study data and replaced with a unique code to keep the patient's identity secret. The subject identification log, linking the identifiers to the participant's name, will be kept in the site's regulatory binder and only accessible by the treating physician. Clinical reports necessary for remote source data verification will be de-identified, coded with the personal identifiers and electronically transmitted to the partners. No personally identifiable information will be at any time shared. For personal data deposited in EGA and RDR that are not publicly accessible, access can be requested via the designated data access committee. 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	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.

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...).	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. Also, Both EGA and RDR ensures long term storage. KU Leuven RDR ensures storage for at least 10 years. However, it is possible to request that the data be kept for a longer period of time.
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. For storage and sharing, we will ensure compliance with GDPR, and will use trusted European repositories such as RDR (https://www.kuleuven.be/rdm/en/rdr) or EGA (https://ega-archive.org/). Metadata will be made openly available and licensed under a public domain dedication open licenses (Creative Commons Attribution International Public License (CC BY)). The access for data from human tissue will be restricted.

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 Disease research group and Center of Neuro-immune interactions 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. Furthermore, the EGA offers a web-based portal and an API to allow data controllers to manage data access and transfer permissions. Authorization infrastructure (AAI) management is a key for operating the EGA where the identity of a user is verified before confirming the access rights to specific information. When a researcher signs in to the EGA website to access sensitive metadata, their credentials are authenticated to verify their identity. Then they can navigate to a dataset of interest and request download of the sensitive data. After the EGA validate their request against the permissions assigned to their identity, if the researcher has permission to access the dataset, the request is authorised, and they can download the data.
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 All personally identifiable information is replaced by a unique study identifier for each patient and each biological sample. The subject identification log, linking the identifiers to the participant's name, will be only kept in the site's regulatory binder and only accessible by the treating physicians (Prof. Séverine Vermeire, UZ Leuven). This log will allow proper monitoring of the clinical records and by the appointed monitors to assure completeness, accuracy and correctness of the data collected at the site. Clinical reports necessary for remote source data verification will be de-identified, coded with the personal identifiers and electronically transmitted to the partners. No personally identifiable information will be at any time shared.

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 -To ensure compliance with GDRP, please use European repositories such as RDR (https://www.kuleuven.be/rdm/en/rdr) or EGA (https://ega-archive.org/) and restrict access for data from human tissue. We will contact the data access committee (dac@uzleuven.be) before depositing. Metadata will be made openly available and licensed under a public domain dedication open licenses (Creative Commons Attribution International Public License (CC BY)). -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."* [8]	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] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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.