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	Raquel Salvador Laureano, orcid ID: 0000-0002-6204-627X
Contributor name(s) (+ ORCID) & roles	Abhishek Garg, PI/promotor. orcid ID: 0000-0002-9976-9922
Project number[1] & title	Revealing novel immunomodulatory myeloid niches for improving colorectal cancer immunotherapy
Funder(s) GrantID[2]	1S44123N
Affiliation(s)	☒ 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	Colorectal cancer (CRC) remains one of the leading causes of cancer-related death in the world. While immune-checkpoint blockers (ICBs) have revolutionized the oncology field, there is still a large subset of patients that do not respond to ICBs, a disparity that is well captured in CRC. CRC has two subtypes with opposing ICB-responses: microsatellite instable (MSI; highly responsive to anti-PD1 ICBs) vs. microsatellite stable (MSS; resistant to anti-PD1 ICBs). Surprisingly, a recent clinical trial showed that anti-CTLA4+anti-PD1 ICB induces positive responses in MSS-CRC. While linked to increased tumoral T cell/myeloid infiltrates, these responses couldn't be distinguished by CTLA4 relevant T cell biomarkers, raising the question: could CTLA4 have unknown myeloid immunomodulatory functions, accounting for this success? In this project we will combine clinical immunogenomics with state-of-the-art experimental techniques and novel CRC mouse models, together with various novel immunotherapy approaches, to comprehensively delineate novel immunomodulatory functions of CTLA4 on the level of myeloid cells, which may be playing an important role in CRC tumour immunology. Our project may lead to novel immunology concepts and novel myeloid-based biomarkers to inform patient selection for, and improve, CTLA4-modulatory immunotherapy for efficient CRC treatment.

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
FACS data	.fcs files from flow cytometers, .xml tables after FlowJo analysis	☒ Generate new data	☒ Digital	☒ Experimental	☒ other: .fcs, .xlsx	☒ < 100 GB	
Plate reader measurement	Absorbance, luminescence, fluorescence readings from microplate reader	☒ Generate new data	☒ Digital	☒ Experimental	☒ other: .xlsx	☒ < 100 MB	
Tumour volume measurements	Measurements of mouse tumours (height x width x length) using a digital calliper.	☒ Generate new data	☒ Digital	☒ Experimental	☒ other: .xlsx	☒ < 100 MB	
Genomics, transcriptomics and single-cell RNASeq	Publicly available data from databases such as TCGA, GEO datasets, and	☒ Reuse existing data	☒ Digital	☒ Compiled/ aggregated data	☒ .csv ☒ other: .h5ad, .xlsx	☒ < 100 GB	

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

(patient) data	published research papers						
Biological samples	Cell lysates from tumours, etc	☒ Generate new data	☒ Physical	☒ Experimental	☒ NA	☒ NA	<100 vials
Digital images (blots, MILAN)	Western blots, cytokine arrays, etc	☒ Generate new data	☒ Digital	☒ Experimental	☒ other: .snc, .tiff, . png	☒ < 100 GB	
Cell lines	CRISPR-Cas knockout cells	☒ Generate new data	☒ Physical	☒ Experimental	☒ NA	☒ NA	<100 vials

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).

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

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.	scRNAseq data from doi: 10.1038/s41588-020-0636-z (GEO database: GSE132465, GSE132257 and GSE144735)
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: Ethical approval for animal experiments has already been granted (P159/2021). This applies to data resulting from animal experiments, including 'tumour volume measurements' and 'FACS data'.
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: - Privacy Registry Reference:

[6] See Glossary Flemish Standard Data Management Plan

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: We don't exclude that the proposed work could result in research data with potential for tech transfer and valorization. KU Leuven has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in several cases the invention will be IP protected (mostly patent protection or copyright protection). As such the IP protection does not withhold the research data from being made public. In the case a decision is taken to file a patent application it will be planned so that publications don't need to be delayed.
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).	Documentation will consist of electronic laboratory records, that contain all of the information of the performed experiment itself (Excel sheetbased metadata files). Those notes will describe the biological samples used/generated, experimental setup and protocols used, results generated, the links to the specific computer location as well as the names of the respective datasets. We will also maintain a metadata sheet with the connection between lab samples and files on our data storage, so that data files, lab samples, and experimental notes remain properly linked. Detailed protocols will be written, including research methods and practices for each experimental initiative. This will be stored in Word or Excel files. Furthermore, a logbook will be kept in Excel containing all steps that were taken to develop the final methodology, date of implementation and name of the researcher who carried out the experiment. Algorithms, scripts and software usage will be documented and stored alongside the electronic laboratory records, e.g. GraphPad Prism, FlowJo. Finally, we will also keep all the information (in dedicated Excel sheets, PDFs or Word-files) about purchased antibodies, cell-lines, mouse models and other analogue data-sources. Other relevant information about these reagents and tools (e.g., proof of antibody specificity) will be derived from initial standardization and optimization experiments and will be retained along with general research documentation/meta-data files.

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 various metadata standards as applicable for different experiment/datatypes, as already established elsewhere: https://fairsharing.org/. For instance, flow cytometry (https://flowrepository.org/), microscopy imaging (https://www.openmicroscopy.org/), qRT-PCR (http://miqe.gene-quantification.info/), and publicly available TCGA patient data analyses (https://gdc.cancer.gov/aboutdata/ data-standards) have very well-defined pre-established meta-data standards. In case we do not have a metadata standard available for a technique/datatype, a metadata of the numerical datasets will be created manually (e.g. based on the Dublin core metadata standard). For most of the data, metadata will be provided as readme, word or excel files, containing all settings and technical descriptions of the experiment and data. In parallel, detailed meta-data info will be integrated within the electronic laboratory records linked to each experiment (as described above) 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?	Digital files will be stored on KU Leuven data storage servers. All data generated during the project will be stored on the local KU Leuven servers, PI computers, and backup hard drives, as well as on a local RAID storage available in the office. This will be initially located in the real-time folders (on lab provided laptops/PCs of the students or employees and local KU Leuven servers) and later only in the archive folders (archive is mirrored; on local KU Leuven servers, backup hard-drives as well as PI computers). Any

	algorithms, scripts or softwares originally generated during the project will be stored in private online git repositories of the PIs. As soon as the manuscript is publicly available, the repository will be changed to a public repository. Specific biological samples (e.g. cell lysates, protein/nucleic acid samples) will be stored in a freezer (-20°C or -80°C) while cell lines will be stored in liquid nitrogen.
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.*	Digital data will be stored on the university's secure network drives with automatic daily back-up procedures (e.g. JDrive for confidential data and KUL Enterprise Box for non-confidential data. Data is also stored as a backup on lab computers, external hard drives, and on a local RAID 50 system in the office.
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, in the various KUL storage services, as well as the 36Tb local RAID 50 system present in the office. 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]	The "J-drive" and "KUL Enterprise Box" servers are accessible only by laboratory members and PIs. Local storage systems like the RAID system and external hard drives are not accessible remotely and password protected.

[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?	The total estimated cost of data storage during the project is 500 EUR. This estimation is based on the following costs: - The costs of digital data storage are as follows: approximately 52 EUR/100 GB/Year for the “J-drive” and approximately 10 EUR/100 GB/Year for the “KUL Enterprise Box”.

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...).	The minimum preservation term of 10 years after the end of the project will be applied to all relevant data. Cell lines will also be stored for at least or more than 10 years. Cell lysates and certain biological samples will be stored for at least 5 years (beyond 5 years, the degradation of these samples makes their storage “useless” and would require repeating the experiments to re-generate the data).
Where will these data be archived (stored and curated for the long-term)?	Data will be stored as described in section 4, as well as archived using KU Leuven’s long term storage/large volume storage service.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	KU Leuven long term storage/large volume storage should cost around 100€/TB/year. The total volume of relevant data to be stored is expected to be significantly under that. The costs for archival data will be paid upfront from the FWO project funds.

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.	Access to data will be granted upon 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:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Digital data will be made publicly available as per the journals' data availability policy, or available upon email request to the PI. Data that do not support publication will be made available upon reasonable request by email.

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'.*	After 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]	Creative Commons Licenses (CC BY) will be attached to the data deposited.
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: DOI when published.
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected for data sharing.

[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?	The PhD and technician associated with this project will be responsible for data documentation & metadata, under supervision of the PI.
Who will manage data storage and backup during the research project?	Data management, storage and back up will be performed by the PhD and technician associated with this project, under supervision of the PI.
Who will manage data preservation and sharing?	The PI (Prof. Abhishek D Garg) will be responsible to ensure data preservation and reuse.
Who will update and implement this DMP?	The PI bears the end responsibility of updating & implementing this DMP.