

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	Daniel Borrás Morales
Contributor name(s) (+ ORCID) & roles	Abhishek D. Garg (Supervisor) (Head of CSI lab) Sabine Tejpar (Co-Supervisor) (Head of Digestive Oncology)
Project number ¹ & title	Deep multi-omics dissection of the CD8+T-cell immune-landscape in colorectal cancer (CRC) to reveal tumor-specific vs. microbial immune- drivers for immuno-oncological application
Funder(s) GrantID ²	1279223N
Affiliation(s)	<input checked="" type="checkbox"/> KU Leuven
Please provide a short project description	The main goal of this project is to comprehensively delineate novel, unanticipated, quantitative and qualitative features of colorectal cancer (CRC) patients' CD8+T-cells, responding to both tumoral antigens or gut microbiome depending on CRC tumor-subtypes, and differentially modulating immunotherapy-response. With this project I will 1) Characterize the functional signaling drivers of CD8+T-cells activation and potential impact on CRC patient response to immunotherapy; 2) Explore TCR annotation associated with immune cell activation and effector inflammatory pathways; 3) Characterize tumor-antigenic/microbiome associations with immune activation responses in CRC patients; 4) Delineate prognostic machine learning-classifiers of CRC response rates to different immunotherapies. We will integrate these characteristics to create uniform patient immunostratification classifiers informing rational designs of clinical trials with novel anti-CRC immunotherapy approaches.

2. Research Data Summary

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.

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 data³.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
scRNAseq	Single-Cell RNA-seq of CRC patients	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Compiled/aggregated data	<input checked="" type="checkbox"/> other: .h5ad	<input checked="" type="checkbox"/> < 1 TB	
TCGA	Expression set of TCGA CRC RNA-seq	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Compiled/aggregated data	<input checked="" type="checkbox"/> .csv	<input checked="" type="checkbox"/> < 100 GB	
scTCR-eq	Single-Cell TCR-seuencing of CRC patients' samples	<input checked="" type="checkbox"/> Reuse existing data	<input checked="" type="checkbox"/> Digital	<input checked="" type="checkbox"/> Compiled/aggregated data	<input checked="" type="checkbox"/> other: .h5ad	<input checked="" type="checkbox"/> < 100 GB	

³ 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 DATA⁴ (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).

⁴ These data are generated by combining multiple existing datasets.

<p>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.</p>	<p>All relevant code and data for this project is available on Synapse (Syn ID: syn41687327). Including novel resources created for this study such as single-cell data integrated with original datasets created for this study (scRNA-seq and scTCR), and public datasets, as well as associated metadata.</p> <p>TCGA dataset: TCGA data, publicly available from Xena - http://xena.ucsc.edu, FireBrowse portal (a Broad Institute GDAC Firehose analyses pipeline: http://firebrowse.org/), TCGA PanCancerAtlas Immune Response Working Group's Cancer Research Institute (CRI) iAtlas Explorer (https://gdc.cancer.gov/about-data/publications/panimmune).</p> <p>Single cell and TCR datasets: (2018-2795 and 2018-2376) for CRC-SG1 and CRC-SG2, Samsung Medical Center (approval no. SMC2017-07-131) for the SMC, and Commissie Medische Ethiek UZ KU Leuven/Onderzoek (approval no. S50887-ML4707) for the KUL3 and KUL5 datasets, respectively. OR accessed from these publications: Nat Genet 54, 963–975. 10.1038/s41588-022-01100-4.; Nat Genet 52, 594–603. 10.1038/s41588-020-0636-z.</p>
<p>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.</p>	<p> <input type="checkbox"/> Yes, human subject data <input type="checkbox"/> Yes, animal data <input type="checkbox"/> Yes, dual use <input checked="" type="checkbox"/> No If yes, please describe: </p>

<p>Will you process personal data⁵? 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.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes:</p> <ul style="list-style-type: none"> - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
<p>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.</p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes, please comment:</p> <p>Possible diagnostics and biomarker tool value.</p>
<p>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.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:</p>
<p>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.</p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No If yes, please explain:</p>

⁵ See Glossary Flemish Standard Data Management Plan

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 and datasets have been stored in a repository including data files (.h5ad), and metadata files and tables (.csv) as well as Jupyter python notebooks (.ipynb) for the proper reproducibility of scripting analysis approaches.
<p>Will a metadata standard be used to make it easier to find and reuse the data?</p> <p>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.</p> <p><i>REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.</i></p>	<p><input type="checkbox"/> Yes <input checked="" type="checkbox"/> No</p> <p>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:</p> <p>If no, please specify (where appropriate per dataset or data type) which metadata will be created:</p> <p>Each dataset contains all metadata and descriptions associated in csv files including all IDs to remap the data accordingly.</p>

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	Local RAID system initially and then transferred to the KU-Leuven LTS
--------------------------------	--

<p>How will the data be backed up?</p> <p><i>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.⁶</i></p> <p><i>REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.</i></p>	<p>Backed up with redundancy on the RAID 50 system stored locally and the duplicated in the KUL LTS</p>
<p>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.</p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If yes, please specify concisely: The CSI has access to a 36Tb local RAID50 system and also access to the KUL storage services including LTS</p> <p>If no, please specify:</p>
<p>How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p> <p><i>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</i></p>	<p>Data will be stored inside the CSI lab in a local RAID system not accessible remotely. Additionally, the remote accessible versions will be made only available as per user after a proper agreement has been set into place for the collaboration and only in a per collaboration project-based request (Synapse).</p>

⁶ 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?	<p>Data storage for the high number of samples and results obtained requires us to contract KU Leuven ICT services for data storage that charges between 300 euros per TB per year for the long-term storage and 800 for the high-performance storage requirements.</p> <p>We expect to use a 10TB of data storage which can be up to 4000 eur for the fast HPC capabilities or 6000 eur of long-term storage maintain it during at least the 5 years term of data keep.</p> <p>As part of the Senior PostDoc grant from FWO, these expenses were considered and specific budget to cover those was requested and granted.</p>
---	---

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 datasets and results as raw data and processed data as well as analysis output and metadata will be stored and retained for at least 5 years after the end of the project, either locally (RAID systems) or remotely (Synapse or KUL LTS services).
Where will these data be archived (stored and curated for the long-term)?	For LTS this data will be stores in both local systems such as RAID50 from CSI lab as well as the KUL LTS services.

<p>What are the expected costs for data preservation during the expected retention period? How will these costs be covered?</p>	<p>We expect to use a 10TB of LTS which can be between 4000 eur and 6000 eur of LTS maintained during at least the 5 years of long-term of data storage.</p>
---	---

6. Data Sharing and Reuse

<p>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.</p> <p><i>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/#INFOEU-REPO-ACCESSRIGHTS</i></p>	<p> <input type="checkbox"/> Yes, in an Open Access repository <input checked="" type="checkbox"/> Yes, in a restricted access repository (after approval, institutional access only, ...) <input type="checkbox"/> No (closed access) <input type="checkbox"/> Other, please specify: </p>
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>Data will be stored as restricted Synapse repository and will be only granted accessed per-project basis and upon requests prior to publication of data.</p>
<p>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.</p>	<p> <input type="checkbox"/> Yes, privacy aspects <input type="checkbox"/> Yes, intellectual property rights <input type="checkbox"/> Yes, ethical aspects <input type="checkbox"/> Yes, aspects of dual use <input type="checkbox"/> Yes, other <input checked="" type="checkbox"/> No </p> <p>If yes, please specify:</p>
<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<p>Synapse (Syn ID: syn41687327)</p>

<p>When will the data be made available?</p> <p><i>THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.</i></p>	<p>Data will only be made available upon article publication acceptance (Currently under submission).</p>
<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>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.</i></p> <p><i>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." ⁷</i></p>	<p>Not applicable: We have not produced any new data or dataset</p>
<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p><input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes:</p> <p>Current repository of Synapse (Syn ID: syn41687327) and PID or DOI will be provided after publication (currently under submission)</p>

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

What are the expected costs for data sharing? How will these costs be covered?	None
---	-------------

7. Responsibilities

Who will manage data documentation and metadata during the research project?	Dr. Stefan Naulaerts and Prof. Abhishek D. Garg
Who will manage data storage and backup during the research project?	Dr. Stefan Naulaerts and Prof. Abhishek D. Garg
Who will manage data preservation and sharing?	Dr. Stefan Naulaerts and Prof. Abhishek D. Garg
Who will update and implement this DMP?	Dr. Stefan Naulaerts and Prof. Abhishek D. Garg