FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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	Joana Rocha-Pereira (https://orcid.org/0000-0001-8299-3787)
Contributor name(s) (+ ORCID) & roles	Emma Roux (https://orcid.org/0000-0003-2629-6316), PhD student in the team
Project number [1] & title	3M220686, Deciphering human norovirus cell tropism and pathogenesis using zebrafish
Funder(s) GrantID [2]	G065623N
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310

[1] "Project number" refers to the institutional project number. This question is optional. 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.

Please provide a short project description	Human noroviruses (HuNoVs) are the most common cause of gastroenteritis, resulting in 219000 deaths/year and a societal cost of $60 billion. The lack of robust small animal models has hindered the understanding of HuNoV biology and the development of effective therapeutics. Our team established the most simple and robust HuNoV replication model available using zebrafish larvae. We will now investigate the unknown details of HuNoV pathogenesis using these small and optically transparent zebrafish larvae. We will first determine which intestinal and immune cell types are targeted by HuNoV and study the triggered innate immune response using zebrafish lines with fluorescent cell types or gene knockouts available in house. Next, we will generate a comprehensive transcriptional profile of each identified infected cell and of enteric neurons by single-cell RNA sequencing, which brings an unbiased view on host-pathogen interactions and the interplay between innate immunity and the enteric nervous system. In addition, by directly measuring the HuNoV-induced functional changes to the intestine, we aim to identify a virus-encoded enterotoxin. Taken together, this project will dramatically expand our knowledge on the multiple underlying mechanisms leading to HuNoV disease. This will greatly advance the collective understanding of this infection and will constitute an important step forward towards the development of efficient countermeasures to control and prevent this infection.

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 [3].

				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
Microscopy images	Microscopy images of whole zebrafish larvae taken with the Leica DMi8 microscope or DragonFly confocal microscope	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.ims (Imaris images) .lif (Leica images) .tif (exported images) .txt (metadata)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Whole fixated and stained zebrafish larvae are kept at 4°C protected from light in our lab.
Flow cytometry plots	Flow cytometry plots of zebrafish cell suspensions acquired with BD LSRFortessa X20 + FACS Diva software	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.fcs	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/
qPCR data	qPCR data of viral quantification	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual	.xls .eds .edt	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB	RNA extracts are kept at -80°C in the freezer in our lab until all related work is finished.

				☐ Model ☐ Software ☐ Other:		☐ NA	
RNA-Seq	Sequencing of the viral genome + expression analysis of zebrafish cells in scRNA-Seq	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.fastq	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

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.	/

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

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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☒ Yes, animal data; provide ECD reference number: P066-2022 ☐ Yes, dual use; provide approval number: ☐ No Additional information:
Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☒ Yes (provide PRET G-number or EC S-number below) ☐ No Additional information: G-2021-4376 s63536
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:
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: For some zebrafish lines that will be used in the project (including the irf8 st95 line), an MTA from the provider (European Zebrafish Research Center, EZRC) is in place which restricts the work to academic research purposes and limits the distribution with others without permission of the provider.
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:

[4] 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). *RDM guidance on documentation and metadata*.	All experiments with their details are described in an online lab notebook using OneNote which is backed up continuously by KU Leuven servers. For microscopy images, all associated documentation and metadata is stored together with the images on a computer and once opened in corresponding LAS X or Imaris software, details are shown including dimensions, bit-depth, microscopy settings etc. The same applies for other used techniques including flow cytometry and qPCR; all details on the required parameters are written down in a lab book but are visible as well once files are opened in their corresponding software programs and the files needed for this are stored on a KUL controlled and backed up server to ensure access at all times.
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 that is generated (e.g. microscopy images) will be stored together with its corresponding picture. For confocal microscopy images, the following standard is followed *(*ISO 21073:2019 Microscopes — Confocal microscopes — Optical data of fluorescence confocal microscopes for biological imaging). 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? *Consult the to find the most suitable storage solution for your data.*	☒ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☐ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☐ Other (specify)
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 All 'small' data are currently stored on the backed-up J drive of the KU Leuven on which there is sufficient storage space foreseen (800 GB) and which is constantly monitored by KUL IT services. Larger data including microscopy projects and FASTQ files will be stored on a specifically allocated L drive of the KU Leuven on which there is sufficient storage space foreseen (10 TB) and which is constantly monitored by KUL IT services. Older data of projects that are finished, will be stored on a separate K-drive for archive purposes. 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.* *Guidance on security for research data*	All data is stored on a KU Leuven backed-up server wherefore access is granted to only team members and this access is controlled and overlooked by a designated person of our research group (Head of the research group: Joana Rocha-Pereira). The -80°C containing virus and samples for further processing by qPCR quantification/sequencing/etc freezer is locked by a key only accessible to team members. For all other physical samples, the general precaution is that batch access is required to enter the building/floor.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The costs of a KUL server backend storage are; 415.2 euros/year for the J drive (800GB), 1138.4 euros/year for the L drive (10TB), and 22.768 euros/year for the K drive (200GB). The costs for data storage and backups are concerning the whole research group and are not specific for this research project and thus the costs will be divided over all funding available by our group including the bench fee available through this project.

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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)

Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☐ KU Leuven RDR ☐ Large Volume Storage (long-term for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specific): All generated data of this project will be stored in a designated folder on the network drive specifically designated for long-term storage (K drive) backed up and secured by the KU Leuven. We do not wish or plan to deviate from the plan to store this data for a minimum of 10 years.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The costs of a KUL server backend storage are; 415.2 euros/year for the J drive (800GB), 1138.4 euros/year for the L drive (10TB), 22.8 euros/year for the K drive (200GB). The costs for data storage and backups are concerning the whole research group and are not specific for this research project and thus the costs will be divided over all funding available by our group including the funding available through this project.

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, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or 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.	
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.	☒ KU Leuven RDR ☐ Other data repository (specify) ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)

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.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
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, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☒ No
What are the expected costs for data sharing? How will these costs be covered?	Every researcher can store 50 GB per year for free on the KUL research data repository. If the dataset exceeds this limit, the researcher should contact the RDR helpdesk. The storage quota applies to the first author of the dataset and will be covered by the FWO project grant.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	A PhD student still to be hired will manage data documentation during the project.
Who will manage data storage and backup during the research project?	A PhD student still to be hired will manage data storage and back-up during the project.
Who will manage data preservation and sharing?	Joana Rocha-Pereira, PI of the project will manage data preservation and sharing.
Who will update and implement this DMP?	Joana Rocha-Pereira, PI of the project will implement and update the DMP.