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	**Peter Carmeliet**, 0000-0001-7961-1821
Contributor name(s) (+ ORCID) & roles	**Mieke Dewerchin**, project supervisor, ORCID: 0000-0002-0382-9346 **Hui-Chao Zhao**, responsible for all omics data analyses, assistance with project supervision, ORCID: 0000-0001-5032-1062
Project number[1] & title	**MEPIcephaly** Metabolic and epigenetic interplay in neural progenitor cells: investigating neurodevelopmental disorders associated with impaired neural progenitor cell expansion
Funder(s) GrantID[2]	
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	This project focuses on microcephaly-associated neurodevelopmental disorders that are characterized by a critical decrease in brain size, primarily caused by impaired neural stem and progenitor cell expansion in the fetal neocortex. We will collect/generate transcriptomic, metabolomic and histological data to explore the progenitor cell metabolism in normal and diseased conditions, seeking to better understand the causal relationship between abnormal neural progenitor cell metabolism and microcephaly. This project will provide new insights into the causes of microcephaly, which is key for the development of improved diagnostic tools, treatment options and preventive care.

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
Published single cell/nucleus RNA sequencing data sets of fetal human neocortex (syn26346373, dbGaP: phs001836, etc.)	Processed data files (expression matrix and metadata) are retrieved and analysed using R; generated files are saved as .csv, .png or R object.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .png ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Metabolomic data of whole neocortex samples of control and Mcph1 mouse brain	Unprocessed data files are processed for picking peaks and metabolite annotation; generated results are	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB	

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

	saved as csv or png files.			☐ NA	☐ .tab ☐ .gml ☒ other: .png ☐ NA	☐ NA	
Histological staining data	Scanning images from histological staining: digital images processed by ImageJ (.tiff/.jpeg), analyzed in excel (.csv)	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .tiff, .jpeg ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
In vitro functional assay data	Metabolic flux assay (.xlsx / .docx), use of spheroid sprouting model (.xlsx / .docx) digital images processed by the ImageJ/FiJi JAVA package (.tiff / .jpeg), etc.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .lif, .tiff, .png	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

					☐ NA		

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).

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 Raw count data and associated metadata from project dbGaP: phs001836 were downloaded from the CoDEx viewer (http://solo.bmap.ucla.edu/shiny/webapp/), while the data from project ID: syn26346373 were retrieved from the Single Cell Portal: https://singlecell.broadinstitute.org/single_cell/study/SCP1756. The information will be updated if new published data is included in the future.

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

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: Mouse tissues will be assessed for metabolomic assay and histological staining. All the mouse samples for these analyses (frozen tissue and paraformaldehyde-fixed paraffin-embedded sections) are generated by Partner Paris of the Mepicephaly consortium. No local ethical approval is needed. In the case human tissue would also become available (collection by the Paris partner), these will also be , paraformaldehyde-fixed paraffin-embedded sections which will be used in Leuven for histological staining, the relevant data will be handled according to the principles of the General Data Protection Regulation (GDPR) 2016/679 and the Belgian privacy law, and approval by the Ethics Committee Research UZ/KU Leuven will be applied for.
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:
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:

[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: Access to pre-existing background of consortium partners, joint ownership of results, and policies about dissemination of own and jointly owned results, are stipulated In the Consortium Agreement, signed by principal investigators and the Legal representatives of the institute of each consortium partner.
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: the study is part of an international consortium project (4 partners). Agreements about legal issues are stipulated In the Consortium Agreement, signed by principal investigators and the Legal representatives of the institute of each consortium partner.

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).	Metadata will be collected from all experiments performed for clarification. We follow standard metadata schemes, provided and required by the host institution. All metadata will be in English language. Any abbreviations used will be defined. Standard vocabulary, taxonomy and ontology formats are followed. Metadata will provide information about the design of the study, individual samples, experimental approaches and protocols, and references to processed and raw data file names will be provided upon data deposition in repositories. This ensures enough information for data interpretation, supports findability, citation and reuse. We also adhere to standard formats of the data, using only standard programs that are available either for free or for a fee and are commonly used in the scientific community. Excel documents will always be saved in the csv format, so that it can be read as American csv/tab-delimited text or European csv/tab-delimited text. For publication, the standards of the journal in which the data will be published, will be used. For all stored data, a readme-file is provided, which includes a short description of the filename, definitions of column headings and row labels, data processing steps, storage information and contact information.

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 will be kept on file in accordance with the standards for naming required by the host institution, which includes date generated, project ID, and numbering of experiments. In case human fetal samples become available, metadata will be created for the human fetal brain tissues, including from which we obtain the tissues and their gestational week. In the case of animal data, metadata will record all relevant data, which is associated with respective animal label. For all stored data, a readme-file is provided, which includes a short description of the filename, definitions of column headings and row labels, data processing steps, storage information and contact information. All data receives a digital object identifier to ensure persistent identification and easy searchability and discoverability. Search keywords will be provided, so that data will easily be located on the server and cloud storage for other researchers in the lab. 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?	All data will be stored on the "large storage network L-drive" - KU Leuven LUNA, centrally managed by the central computer, IT department of KU Leuven: ICTS (Informatie en Communicatie: Technologie en Systemen). Additionally, a cloud-based KU Leuven Enterprise BOX is available for the secure storage, management and sharing documents between the research group. Moreover, data can also be stored at the institutional research data repository: RDR from KU Leuven (after completion of the project).

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.*	The data will be backed up in a double way. Automatic back-up (every 24 hours) of the network L-drive is controlled by the ICTS KU Leuven department. In addition, every researcher's computer has installed the Druva Cloud Platform. Druva Cloud protects and manages data across all devices, and allows to perform the backup operations even every 5 minutes (managed individually - depends on the user).
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: The project will generate data less than 100 GB. An unlimited storage space is already available and maintained by the ICTS KU Leuven department. 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]	Research data are stored and managed by the local IT-manager (Urbain Schepereel) and the ICTS KU Leuven department, and are accessible only by the researchers working on the project (access right management & password protection). Moreover, data will be backed up in two ways as stated above.

[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?	Yearly storage costs of 1TB data on large storage servers of the host lab are estimated at 130 €/year. Costs will be covered by internal lab fundings.

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 generated data will be retained for at least 10 years after the end of the project by the IT department of the host institution. Moreover, the data will be publicly available and should therefore be assessable for re-usage with no time limitation. Detailed documentation will be kept in secure storage.
Where will these data be archived (stored and curated for the long-term)?	All the generated data will be stored and archived on the "large storage network L-drive" - KU Leuven LUNA, centrally managed by the central computer, IT department of KU Leuven. All data is backed-up daily to the cloud-storage to ensure safe storage.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Yearly storage costs of 1TB data on large storage servers of the host lab are estimated at 130 €/year. Costs will be covered by internal lab fundings.

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.	All relevant data will made publicly available upon publication. However, before publication, the data will be accessible only by the researchers working on the project. The identity of the person who accesses the data will be verified using institutional account system. If for any reason of accession is needed at an earlier time-point, this can be arranged through collaborations and in cooperation with the host institution's guidance.
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: All relevant data will made publicly available upon publication. The generated materials, such as mouse tissues will be available for lab members and collaborators upon request.

Where will the data be made available? If already known, please provide a repository per dataset or data type.	For publication purposes, our data will be publicly available on data repositories and published articles have an open access status. The generated metabolomic data will be deposited to a public database of MetabolomeXchange.
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 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]*	Not applicable to this project. All relevant data will be publicly available in line with the journal's guidelines. Other access is managed by the lab manager (via verified using institutional account system).
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: Accession number will be added to the generated metabolomic data upon its deposit to MetabolomeXchange.

[8] 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?	We do not expect any costs for data sharing to publicly available repositories.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Peter Carmeliet, lab manager (Luc Schoonjans)
Who will manage data storage and backup during the research project?	local IT-manager (Urbain Schepereel) and the ICTS KU Leuven department
Who will manage data preservation and sharing?	local IT-manager (Urbain Schepereel) and the ICTS KU Leuven department
Who will update and implement this DMP?	Peter Carmeliet, lab manager (Luc Schoonjans)