**Data management plan (DMP)**:

**Project Identifier:** EOS MetaNiche
**Grant Title** G0I2922N (EOS ID: 40007532)
**Principal Investigator / Researcher** Patrizia Agositinis, Gabriele Bergers, Colinda Scheele
**Description** Understanding and tackling the spatio-temporal changes of the metastatic lymph node epicenter
**Institution** KU Leuven

## 1. General Information

**Name of project lead (PI)** Patrizia Agostinis

Project number and Title: G0I2922N (EOS ID: 40007532)

## 2. Data description

**2.1 Will you generate/collect new data and/or make use of existing data?**
Generate new data

**2.2 What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.**

| Type of data | Format of data | Volume | How created |
|---|---|---|---|
| Flow cytometry | .fcs files | 25 GB | Using diverse flow cytometers/sorters |
| Microscopy images | .lif, .tiff, .zem | 150 TB | Using diverse microscopes |
| Sequencing files | .raw, .fastq | 2 TB | Using 10x genomics or other platforms |
| Spatial transcriptomics files | .fastq, .gem, gef | 15 TB | Stereo-seq platfom or equivalent |
| Data analysis files | .xls | 50 GB | Using image analysis software, FIJI, Matlab |
| Statistical data | .pzfx, .xls, .txt | 5 GB | Graphpad, Rstudio |
| Cell lines | .docx, .xls | 500MB | On L-drive |
| Plasmids and vectors | .DNA | 5 GB | On L-drive |
| R/Python scripts | .txt, .py | 1 GB | Seurat, GSEA, ScVelo |
| Electronic lab notebooks | .enl | 10 GB | On L-drive |
| Publication reports | .pdf, .docx | 3 GB | On L-drive |
| Data presentation | .pptx | 15 GB | On L-drive |
| Computational data | .dat, .txt | 10 GB | On L-drive |
| Mouse strains | LAIS database | / | / |
| Photoshop files of data | .psd | 20 GB | On L-drive |
| Illustrator files of data figures | .ai | 20 GB | On L-drive |
| Scans of westernblots | .tiff, .jpeg, .gel | 5 GB | Typhoon NIR or A1600 scanner |

| | | | |
|---|---|---|---|
| Experimental readouts | .pda, .txt, .xpt, .asyr, .xlsx | 1 GB | Flex station, plate reader, Seahorse XFe24 analyzer |
| Metabolomics | .bmp, .jpg, .xlsx | 1 GB | Mass spectrometer |

## 3. Ethical and legal issues

### 3.1 Use of personal data?

No personal data will be used in this project. All patient samples provided by our clinical collaborators are completely anonymized to us.

### 3.2 Are there any ethical issues concerning the creation and/or use of the data?

**Cell lines**: Established cell lines were obtained from commercial supplier, and MSDS sheets have been published and can be consulted online.

**Mouse experiments:**

**P035/2022** – for cancer cell transplantation (melanoma and breast cancer cell lines), ex vivo imaging and profiling of tissues using microscopy/FACS/Single-cell analysis/spatial transcriptomics, and in vivo microscopy.

**P052/2021**-related to the role of lymphatic endothelial cell autophagy in anti-tumor immunity; it involves subcutaneous transplantation and growth of melanoma cells, lymph node isolation imaging and immunoprofiling using microscopy/FACS and treatment with immune checkpoint inhibitors.

**P068/2022**-related to the role of autophagy in the lymph node as modulator of lymphocyte trafficking, it involves subcutaneous transplantation and growth of melanoma cells, lymph node isolation imaging and profiling using microscopy/FACS/Single-cell analysis/spatial transcriptomics, and in vivo intravital microscopy.

**LA1210604-** related to HEV autophagy and inflammation in lymph nodes as well as HEV alterations in sentinel lymph nodes, focusing on the cell-biological regulation mechanisms of the LN vasculature**;** it involves LN HEV investigation in different genetic mouse models in which metabolic and autophagic pathways are altered and includes orthotopic melanoma injection and investigation of lymph node metastases.

**Human tissue donation:**

All human biomaterial will be obtained through our clinical collaborators following the ethical principles under informed consent, protection of privacy, and upon voluntary donation. Lymph nodes from breast cancer patients will be obtained through the post-mortem tissue donation program UPTIDER (study number S64410/NCT04531696) or biobanked sentinel lymph nodes taken after surgery under study number S64813.

All patient-derived material will be used following the standard operating procedures for handling human biomaterial in our center, and in accordance with the European and national regulations.

### 3.3. Does your research possibly result in research data with potential for tech-transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

All employees have signed a contract, and the IP rights are implemented in this employment contract. The proposed work could result in data with potential for technology transfer or valorization. All data generated within this project belongs to KU Leuven and VIB, in accordance with the agreement between both institutes.
The project will be actively monitored by the tech-transfer offices of both institutes to scout for research data with valorization potential. Each invention will be thoroughly assessed and if desired, the invention will be IP protected (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 in such a way that publication of the data is not delayed.

**3.4. Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions regarding reuse and sharing are in place?**

This project can result in data with the potential for technology transfer or valorization. The need for 3rd party agreements will be evaluated case by case in consultation with KU Leuven and VIB. We do not exclude that the work could become restricted due to 3rd party agreements.

## 4. Documentation and metadata

**4.1. What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?**

Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) and in hard copy lab notebooks that refer to specific datasets.
All protocols and necessary details related to data collection as well as methods of analysis will be recorded in licensed E-lab journal containing **word (.docx), endnote (.enl)**, and **excel (.xlsx) files** stored at a shared KU Leuven Large Volume Storage drive, which is backed up by KU Leuven IT services. The raw files will be segregated in separate folders according to the Work Packages and experiments within the Work Packages itself.
All standard operating procedures, protocols, lists of materials, lists of cell lines (either commercially available or generated for the project) will be stored in a shared folder on the KU Leuven Large Volume Storage server. The names of files will include date of the experiment, experiment number, type of experiment and different experimental conditions to make the data findable. All biological material will be labelled and stored according to good scientific practice. Mouse data will be kept in the LAIS mouse database, including all the data on the procedures and surgeries that were performed for each mouse.
All data generated in this project will be available to the wider scientific community upon publication. Raw data will be deposited in data repositories, sequencing raw data will be made publicly available using GEO or equivalent, data quantification will be shared using .xslx files.

**4.2 Will a metadata standard be used?**

No, each folder containing a separate experiment will also contain information in a Word (.docx) and Excel (.xlsx) file explaining data methods and all relevant metadata, which include but are not limited to experimental conditions, genetic models used, all sample

identification numbers and computational analysis pipelines. Metadata files with detailed explanations will be stored in a shared folder on the KU Leuven Large Volume Storage server. This will ensure the reusability of the data and the reproducibility of any further data generation.
Metadata will include the following elements:
• Title: free text
• Creator: Last name, first name, organization
• Date and time reference
• Subject: Choice of keywords and classifications
• Description: Text explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.
• Format: Details of the file format.
• Resource Type: data set, image, audio, etc.
• Identifier: DOI (when applicable)
• Access rights: closed access, embargoed access, restricted access, open access.
Additionally, we will closely monitor MIBBI (Minimum Information for Biological and Biomedical Investigations) for metadata standards more specific to our data type.
For specific datasets, additional metadata will be associated with the data file as appropriate.
The final dataset will be accompanied by this information under the form of a README.txt document. This file will be located in the top level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will allow the data to be understood by other members of the laboratory and add contextual value to the dataset for future reuse.

## 5. Data storage and backup during the project

### 5.1. Where will the data be stored?

All data (except for the large imaging files) will be stored on the L-drive (Large Volume Data Storage) in a dedicated folder for this project. Only the PI's and the researchers involved in this project will have access to these folders. This project will generate extremely large image files. All processed image files will be stored on the L-drive. To accommodate the raw imaging data, we have purchased an additional 180TB of network storage hosted by the KU Leuven ICT. Upon publication the data will be moved to the data archive (K-drive), which is designed for long-term storage of archived data. The data on this drive cannot be moved, modified, or deleted by the researchers, nor the PI's (only ICT service can modify these data).
In particular;
- Omics and single cell RNAseq data: omics and scRNAseq data generated during the project will either be stored on KU Leuven servers or on the Flemish Supercomputer Centre (VSC), initially in the staging area and subsequently in the archive area.
- Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (- 80°C). All published vectors and the associated sequences will be sent to the non-profit plasmid repository Addgene, which will take care of vector storage and shipping upon request.

- Cell lines: Newly created human cell lines will be stored locally in the laboratory in liquid nitrogen storage and will be deposited in the UZ Leuven-KU Leuven Biobank. Other human cell lines will be stored locally in liquid nitrogen cryostorage of the laboratory when actively used for experiments. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory.
- Bacterial and yeast strains will be stored in a -80°C freezer in the lab of Patrizia Agostinis, Gabriele Bergers and Colinda Scheele. Costs are covered by general lab expenses.
- Genetically modified organisms: Mice will be maintained in facilities of the Laboratory Animal Center of KU Leuven, which applies Standard Operation Procedures concerning housing, feeding, health monitoring to assure consistent care in accordance with European and national regulations and guidelines. All animals will be registered in the Leuven Animal Information System (LAIS) database, along with corresponding genotyping information, ethical approval documents and animal provider receipts.
- Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.
- Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository from the GitHub account of the department (https://github.com/vibcbd).

**5.2 How will the data be backed up?**

Data stored on the KU Leuven L-Drive is managed, maintained, and backed up by KU Leuven IT services. Specifically, mirror copies of the stored data are made immediately upon upload, for safety backup purposes. Raw imaging data stored on the extra network storage will be backed up once every few weeks, and only 1 backup will be kept.

KU Leuven drives are backed-up according to the following scheme:
- data stored on the “L-drive” is backed up daily using snapshot technology, where all incremental changes in respect of the previous version are kept online; the last 14 backups are kept.
- data stored on the “J-drive” is backed up hourly, daily (every day at midnight) and weekly (at midnight between Saturday and Sunday); in each case the last 6 backups are kept.
- data stored on the digital vault is backed up using snapshot technology, where all incremental changes in respect of the previous version are kept online. As standard, 10% of the requested storage is reserved for backups using the following backup regime: an hourly backup (at 8 a.m., 12 p.m., 4 p.m. and 8 p.m.), the last 6 of which are kept; a daily backup (every day) at midnight, the last 6 of which are kept; and a weekly backup (every week) at midnight between Saturday and Sunday, the last 2 of which are kept.
- All single cells/omics data stored on the Flemish Supercomputer Centre (VSC) will be transferred on a weekly basis to the archive area which is backed up.
Incremental backups are done daily from one 20 TB QNAP NAS to a second 20 TB QNAP NAS.

**5.3. 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.**

There is sufficient storage and back-up capacity on all KU Leuven servers:

- 180 TB of storage is already foreseen for raw image data storage – in our budget we have requested for another cloud base storage system to allow for easy sharing of data between external partners of the consortium (20.000 euro over 4 years) with an estimated size of 500TB.
- The "L-drive" is an easily scalable system and is expandable in blocks of 5TB, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.
- the "J-drive" is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system.
- The archive storage is done on the "K-drive" and this drive is expandable in blocks of 100GB.

**5.4. What are the expected costs for data storage and backup during the project? How will these costs be covered?**

The estimated cost for the KU Leuven Large Volume Storage (L-)drive per 5TB per year is 569,20 euro. Total estimated size of the generated data on the L-drive within this project is 100TB, which reflects 11.384 euro per year. The performance of mirror copies of the stored data for safety backup purposes is included in the prize. These costs will be jointly covered by the project leaders at KU Leuven.
An additional 180TB is already available, paid by Prof. Scheele. The total size of the images and sequencing datafiles generated within this project will be around 500 TB which will come to about EUR 20,000 Euro for the entire project.

For the K-drive (data archive) storage space of 1 TB is foreseen and will cost €128 each year, this is also expandable in blocks of 100 GB. These costs were foreseen in the application and if more the lab budget will be used to cover these expenses.

**5.5. Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

Data will only be accessible by authorized members, i.e., PIs and researchers working on the project. Folders will be managed by the ICT service of KU Leuven, and all data will be stored on drives and servers managed by KU Leuven. Both the "L-drive" and "J-drive" servers are accessible only by laboratory members, and are mirrored in the second ICTS datacenter for business continuity and disaster recovery so that a copy of the data can be recovered within an hour.

Access to the servers is only possible through a KU Leuven user-ID and password, and will only grant access to data made accessible to the specific user-ID. KU Leuven works with a multi-factor authentication mechanism to increase the security. Sensitive data transfer will be performed according to the best practices for "Copying data to the secure environment" defined by KU Leuven. The operating system of the vault is maintained on a monthly basis, including the application of upgrades and security patches. The server in the vault is managed by ICTS, and only ICTS personnel (bound by the ICT code of conduct for staff) have administrator/root rights. A security service monitors the technical installations

continuously, even outside working hours. All private data will be rendered anonymous before processing outside the digital vault. Only the PI will be granted access to the server to deposit private data. The PI will be the only responsible for linking patient information, survey data and/or tissue samples, and will strictly respect confidentiality. All de-identified data will be exported from the database by the PI, and stored on KU Leuven servers from where it can be accessed by the research and technical staff from the laboratory. Together, these measures ensure that non-authorized persons can't access or modify the data.

## 6. Data preservation after the end of the project

### 6.1. Which data will be retained for the expected 10-year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

The minimum preservation term of 5 years after the end of the project will be applied to all datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 5 years, conform the KU Leuven RDM policy. The costs (€156 per TB per year for "Large volume-storage" ) will be covered by the lab.

### 6.2. Where will these data be archived (= stored for the long term)?

As a general rule, datasets will be made openly accessible, whenever possible via existing platforms that support FAIR data sharing (www.fairsharing.org), at the latest at the time of publication.
For all other datasets, long term storage will be ensured as follows:
-Digital datasets: files will be stored on the "L-drive".
-Tissue samples: Tissues will be stored locally in the laboratory.
-Omics and scRNASeq data: datasets will be stored on the "L-drive" or, for larger datasets, on the Vlaams Supercomputer Centrum.
-Vectors: As a general rule at least two independently obtained clones will be preserved for each vector, both under the form of purified DNA (in -20°C freezer) and as a bacteria glycerol stock (- 80°C).
-Cell lines: human cell lines will be stored in the UZ Leuven Biobank (-80°C). Human pluripotent stem cell lines generated during this project will be deposited in hPSCreg. Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory.
-Other biological and chemical samples: storage at 4°C and/or as frozen samples in cryovials as appropriate.
- Following publication, the results associated with each study will also be deposited in the Dryad repository, where they will be preserved indefinitely.

### 6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Each year €128 will be charged from our ICT service for the use of 1 TB on the k-drive (long term storage), back-up service is included in the price. These costs were foreseen in the budget request of the application and if more, the lab budget will be used to cover these expenses.

**7. Data sharing and re-use**

**7.1. Are there any factors restricting or preventing the sharing of (some of) the data(e.g. as defined in an agreement with a 3rd party, legal restrictions or because of IP potential)?**

This project can result in data with the potential for technology transfer or valorization. The need for 3rd party agreements will be evaluated case by case in consultation with KU Leuven and VIB. We do not exclude that the work could become restricted due to 3rd party agreements.

**7.2. Which data will be made available after the end of the project?**

Participants to the present project are committed to publish research results to communicate them to peers and to a wide audience. All research outputs supporting publications will be made openly accessible. Depending on their nature, some data may be made available prior to publication, either on an individual basis to interested researchers and/or potential new collaborators, or publicly via repositories (e.g. negative data).
We aim at communicating our results in top journals that require full disclosure upon publication of all included data, either in the main text, in supplementary material or in a data repository if requested by the journal and following deposit advice given by the journal. Depending on the journal, accessibility restrictions may apply.
Biological material will be distributed to other parties if requested.

**7.3. Where/how will the data be made available for reuse?**

In an Open Access repository
Other (specify): Upon request by email

**7.4. When will the data be made available?**

Upon publication of the research results. As a general rule all research outputs will be made openly accessible at the latest at the time of publication. No embargo will be foreseen unless imposed e.g. by pending publications, potential IP requirements – note that patent application filing will be planned so that publications need not be delayed - or ongoing projects requiring confidential data. In those cases, datasets will be made publicly available as soon as the embargo date is reached.

**7.5. Who will be able to access the data and under what conditions?**

Whenever possible, datasets and the appropriate metadata will be made publicly available through repositories that support FAIR data sharing. As detailed above, metadata will contain sufficient information to support data interpretation and reuse, and will be conform to community norms. These repositories clearly describe their conditions of use (typically under a Creative Commons CC0 1.0 Universal (CC0 1.0) Public Domain Dedication, a Creative Commons Attribution (CC-BY) or an ODC Public Domain Dedication and Licence, with a material transfer agreement when applicable). Interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. For data shared directly by the PI, a material transfer agreement (and a

non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted.

**7.6. What are the expected costs for data sharing? How will these costs be covered?**

It is the intention to minimize data management costs by implementing standard procedures e.g. for metadata collection and file storage and organization from the start of the project, and by using free-to-use data repositories and dissemination facilities whenever possible. Data management costs will be covered by the laboratory budget. A budget for publication costs has been requested in this project.

**8. Responsibilities**

**8.1. Who will be responsible for the data documentation & metadata?**

Metadata will be documented by the research and technical staff at the time of data collection and analysis, by taking careful notes in the electronic laboratory notebook (E-notebook) that refer to specific datasets. The data will be reviewed by the principal investigators/PIs.

**8.2. Who will be responsible for data storage & back up during the project?**

Data storage and back-up: VIB IT-manager (Urbain Schepereel) and the KULeuven ICTS-IT department (Raf De Coster).

**8.3. Who will be responsible for ensuring data preservation and sharing?**

All three PIs are responsible for data preservation and sharing, supported by the research and technical staff involved in the project, and the VIB IT-manager (Urbain Schepereel) and the KULeuven ICTS-IT department (Raf De Coster).

**8.4. Who bears the end responsibility for updating & implementing this DMP?**

All three PIs bear the end responsibility of updating & implementing this DMP.