

iBOF/23/005
FWO DMP (Flemish Standard DMP)

1. 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.

				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
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Generate new data Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Digital Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> Observational Experimental Compiled/aggregated data Simulation data Software Other NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> <100MB <1GB <100GB <1TB <5TB <10TB <50TB >50TB NA 	
multiplex RNA-FISH	multiplex RNA-FISH (eg. CosMx, Merscope, Resolve, stereoseq)	new data	digital	experimental	.csv .tiff .fastq	<5TB	
multiplex immunohistochemistry	mIHC	new data	digital	experimental	.jpeg .tiff	<5TB	
proximal ligation assays	proximal ligation assays	new data	digital	experimental			
Mice	GEMMS, allografts, PDX	reused	physical				Mice will be housed in the animal facility. All animals will be registered in LAIS or Ticket@lab
Engineered cell-lines		new	physical				box in liquid nitrogen registered in our databases
primary cells		new	physical				box in liquid nitrogen registered in our databases
Mouse tissue samples	Fresh/frozen/fixed tissue samples from the mice	new	physical				tissue blocks will be registered in our databases
Single cell RNA sequencing data	Single cell RNA sequencing data	reused	digital	experimental			

Single cell RNA sequencing data	Single cell RNA sequencing data	new	digital	experimental	NGS raw data: .bcl .fastq read/UMI count data: .tsv, .mtx, .loom, .rds, .h5ad sequence alignment data: .bam, .cram coverage data: .bed, .bg, .bedGraph, .bw, .bigwig structural variations data: .vcf, .bcf .	<100 GB	
FACS	FACS	new	digital	experimental	.fcs	<100 MB	
microscopy images and videos	microscopy images and videos	new	digital	experimental	.svg, .eps, .svg, .ai, .TIFF, .png, .avi, .cine, .mp4	<5 TB	
human samples	Fresh/frozen/fixed tissue samples from the humans	new	physical				Tissue blocks will be stored in the biobank
scripts	scripts and software use	new	digital	software	scripts and software used will be documented in Gitlab .m		
Protein quantification	Protein quantification using BCA	new	digital	experimental	.xlsx	<100 MB	
Mass spectrometry	Tissue and plasma samples will be measured using the GC-MS and LC-MS systems.	new	digital	experimental	raw data files: .ms (GC-MS) and .d (LC) MATLAB data analysis: .m (scripts), .mat (data files) export data analysis: .xlsx, .png, .pdf	<100 GB	
Spatial metabolomics	Mass spectrometry imaging	new	digital	experimental	raw data files: .d metadata: .txt data visualization: .mis/.slx metaboscape compound annotation: .mca exported files: .png	<1TB	
CAD files	Microfluidic chip designs	new	digital	Other	.svg, .dwg, .dxf, .cif, .stl	<100 GB	
High-speed fluorescence droplet detection	Data obtained by photomultiplier tubes	new	digital	experimental	.txt	< 100 GB	
Collection and processing of primary numerical data	Collection and processing of primary numerical data	new	digital	experimental	.xls, .csv, .MAT, .txt	< 100 GB	
Mass spectrometry (proteomics)	Data-dependent/-independent analysis by LC-MS/MS of protein digests (MS and MS/MS data)	new	digital	experimental	RAW file format	1-5 GB per file (analysis)	

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:

For the bioinformatic analysis we will reuse scripts that have previously used for other projects and are available on Gitlab [Marine Lab - GitLab \(kuleuven.be\)](#)

We will reuse tumor blocks (cryo + paraffin) from mouse and human tissue that are available in the lab.

We will reanalyse spatial omics data:

- 10X visium and stereoseq are in the GEO repository: [GSE207592](#)
- Resolve data are in Zenodo: [A cellular hierarchy in melanoma uncouples growth and metastasis | Zenodo](#) (DOI: 10.5281/zenodo.6856193)

We will use mouse cell line previously generated from GEMMs used in our lab.

We will re-use CAD files of microfluidic chip designs available in the lab.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, human subject data
- Yes, animal data

ECDs related to this project: 183/2022, 006/2023, 125/2021, 127/2020, 154/2021

for WP 9: the ECD still has to be written

Ethical committee approval: S63799 (PDX mice), S67149 (MelanoMAP)

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes

We will perform spatial omics experiments on human tissues.

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

We do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. Ownership of the data generated belongs to KU Leuven, VIB, UGent, UAntwerpen in accordance with the framework agreement of the institutes. VIB has the policy to actively monitor research data for such potential. Additionally, KU Leuven IOF managers (e.g. Dr. Francesco Dal Dosso in the biosensors group) will continuously scout for IP opportunities in consultation with LRD.

If there is substantial potential, the invention will be thoroughly assessed, and in number of 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 publication need not be delayed.

The use of the mouse models will be subjected to the terms described in their respective MTAs.

Specific examples:

- potential biomarkers for tumor cell states will be claimed if this opportunity arises
- new bioinformatic software can be licensed
- new microfluidic chip concepts will be claimed

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 in the comment section to what data they relate and what restrictions are in place.

- Yes

Within the iBOF consortium, material will have to be exchanged between the different labs. Therefore, we have 1 MTA for the entire iBOF project, which will be valid for the term of the project. This will allow us to share material between the different labs without having to sign new MTAs every time material is exchanged between the partners of the consortium. The steering committee has the responsibility to keep the list of exchanged material up-to-date.

We have MTAs with addgene for several plasmids. Those plasmids can only be used for teaching and academic research purposes. We can not distribute the plasmids to other users without written consent of the provider. The full MTA can be found here: [Addgene UBMTA](#).

We have MTA with Pr A Regev for Watermelon barcoding system which we will use only for academic research purposes and we will not share with anyone else.

We have MTAs for the human cell lines either commercial cell lines or cell lines from our collaborator Ghanem Ghanam at Bordet Institute.

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 in the comment section to what data they relate and which restrictions will be asserted.

- Yes

Intellectual Property arising from this work is managed as per the Research Collaboration Agreement signed by the different iBOF parties.

2. 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).

Marine Lab

The data will be generated follow standard operating procedures (SOPs). Clear and detailed descriptions of these protocols will be stored in our lab protocol database and published along the results.

Metadata will be documented by the research and technical staff at the time of the data collection and analysis. Those metadata will be stored in the same excel file

together with the results and the SOP which is kept in the J-drive.

Large data sets will be kept on the L-drive.

For bioinformatic data the scripts associated with this project will be stored in Gitlab.

Several databases are used to keep track of the storage of physical data, eg. -80°C freezers, liquid nitrogen storage, antibodies, plasmids, primers, cryoblocks, paraffin blocks,...

Fendt Lab

Protocols and details related to data collection and processing will be recorded in electronic lab books by the lead staff scientist (Dr. Ana Margarida Campos) and PhD student (Margherita Demicco). Data folders containing raw and processed data will be hierarchically organized and labeled based on the source of the data, the type of experiment, the date of data generation, and the different experimental conditions analyzed. Data analysis methods and particularities (including metadata) will be described in README.txt and Excel files included in these folders. All files will be stored in the KU Leuven Large Volume Storage space (L-Drive), with sharing possibilities via One Drive (managed by the KU Leuven IT department).

Biosensors group (Lammertyn lab)

Protocols, the research progress and clear descriptions of obtained data, what they represent and how they were generated, will be collected in an electronic notebook (eLABJournal, Bio-ITech). Here, folders will be provided for all subtasks of the project. In each folder, a new file will be made for each experiment, named with the date and subject, and including information on the responsible person (i.e., the person who created the file) as well as version tracking. Each experimental file will contain a section on the objective, protocol, results (a description of results and observations rather than all raw and analyzed data) and conclusions. For each experiment, all raw and analyzed data files will be stored in a folder on the shared server, using the same hierarchical folder structure as the electronic lab notebook. By using the same structure on the server and in the electronic lab notebook, contextual information on the experimentally obtained data can be easily searched and used by a secondary analyst via the electronic notebook.

(Raw)data will be stored on the J-drive, or for large data-sets, the K-drive. These folders are secured and backed-up by the ICTS service of KU Leuven. The folder on the J-drive also contains all the administrative items (e.g. project proposal, project reports, update presentations, contracts).

Scott Lab

Documentation and metadata linked to each experiment will be documented by the technical and research staff in hard copy lab notebooks and in electronic lab notebooks (ELN) in this project. This includes the research design, protocol, context of data collection, data collection methods, quality control procedures, processing and analysis procedures. Cryotubes of biological samples stored at -80°C or -150°C will be labelled with a reference number that links to an entry in its corresponding database.

Gevaert Lab

Next to documentation in a standard lab notebook, all experimental designs and results will be documented on a shared drive which is centrally managed by the host center, VIB Center for Medical Biotechnology. The drive is on a secure server only accessible through the intranet. One can find several sections per experiment:

- A detailed description of the experiment including the protocol, aim of the experiment and the experimental design (eg as .docx or .xlsx files)
- The RAW data obtained during the experiment
- The analyzed data (e.g. as GraphPad files, Excel files)
- An overview of the results (e.g. as ppt file)

Berx Lab

Documentation and metadata linked to each experiment will be documented by the technical and research staff in hard copy lab notebooks and in electronic lab notebooks (ELN) in this project. This includes the research design, protocol, context of data collection, data collection methods, quality control procedures, processing and analysis procedures. Cryotubes of biological samples stored at -80°C or -150°C will be labelled with a reference number that links to an entry in its corresponding database.

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

The data are always kept together with the SOP and the relevant metadata.

For the following type of experiments international metadata standards will be used:

- qPCR: MIQE compliant RDML files will be exported from qbase+ (<https://www.ncbi.nlm.nih.gov/pubmed/19246619>)
- FACS & flow cytometry: MiFlowCyt compliant metadata files will be exported from Fortessa & Sony and Aria Fusion (<https://www.ncbi.nlm.nih.gov/pubmed/18752282>)
- Image metadata will be exported as OME-XML of TIFF-XML files using QuPath software.

If no metadata standards are available to following metadata will be stored:

- Investigator
- unique identifier
- link to related identifiers
- Project
- Keywords
- Type of experiment (in vitro, in vivo, bioinformatics)
- source of data (animal, cell lines, database, patients)
- type of material generated (DNA, RNA, protein, tumor fragment, digital)
- location of the results
- location of physical material
- device + program run
- compounds
- clearance

Biosensors group:

The Biosensors group will use the electronic lab notebook in which a number of predetermined topics have to be described for each experiment (Objective, protocol, results, and conclusion). The electronic lab notebook facilitates searching for particular metadata through a search engine. By mimicking the folder structure of the electronic lab notebook in the server-based folder with the experimental data, linking of the metadata to the actual data will be facilitated.

The electronic lab book software will also be used to manage the sample inventory (coded label, position, date) and track samples in and out of the fridges and freezers.

For each peer-reviewed article, a separate folder will be made on the server, containing the latest word version and all raw and processed data used in the article. In addition, a separate file will be made in the electronic lab notebook for each article, containing clickable links to all metadata files of data that were used in that particular article, to facilitate tracing back of protocols, results and conclusions.

Scott Lab

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) used to produce and to read the data, the purpose of the experiment, etc); Resource Type (data set, image, audio, etc); Identifier (DOI; when applicable).

Gevaert Lab

For proteomics data, prior to LC-MS/MS analysis a project is issued in an online platform (YouTrack), which allows the user to describe the project and automatically generates the corresponding metadata. Once an issue is accepted, it automatically generates a unique code. This code consists out of two parts. First, a three-letter word

corresponding to the center within the VIB (CMB in our case) and second, a number that automatically inclines per new project submitted. By systematically incorporating this code in the raw file name, the txt result files, data analysis workflows and reports, it is possible to collect and connect the different files. Data folders are created within a subfolder on the server with the department's three letter code, this allows an additional layer of control for data security. This folder can only be accessed by the user, the MS responsible within the proteomics core, the department and the IT team. From this server the database searches are performed, thereby automatically writing the results as .txt files in the same folder. In this way one folder contains the raw unprocessed and results.

Berx Lab

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) used to produce and to read the data, the purpose of the experiment, etc); Resource Type (data set, image, audio, etc); Identifier (DOI; when applicable).

3. Data storage & back-up during the research project

Where will the data be stored?

Digital files will be stored on KU Leuven servers, except for private data that will be stored on KU Leuven secure server (digital vault).

Tissue samples: Tissues will be stored locally in the laboratory. Cryoblocks of tumor tissue will be stored at -20°C, paraffin blocks of tumor tissue will be stored at room temperature. All human tissue samples will be registered with a Belgian biobank, in compliance with the Belgian law on human body material (dd 19-12-2008).

Omics data: omics 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 later 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 labs. 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 or ticket@lab, 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.

Microfluidic chips will be stored at room temperature.

Algorithms, scripts and softwares: All the relevant algorithms, scripts and software code driving the project will be stored in a private online git repository in Gitlab

Nucleic acid and protein sequences: All nucleic acid and protein sequences generated during the project will be stored on KU Leuven servers. Upon publication all sequences supporting a manuscript will be made publicly available via repositories such as the GenBank database or the European Nucleotide Archive (nucleotide sequences from primers / new genes / new genomes), NCBI Gene Expression Omnibus (microarray data / RNA-seq data / CHIPseq data), the Protein Database (for protein sequences), the EBI European Genome-phenome Archive (EGA) for personally identifiable (epi)genome and transcriptome sequences.

Scott lab

All research data are stored in personal folders on the S drive on the servers of the VIB-UGent Center for inflammation Research (IRC) according to our data policy. Raw data obtained from flow cytometry and microscopy are stored on separate drives on the IRC servers. Data are stored three times on two different locations.

Gevaert Lab

All laboratory experimental details and data will be recorded daily and dated on the CMB servers. Electronic records of the recorded data will be generated (e.g. by scanning gels or blots) and saved to the CMB servers. File names will have an appropriately descriptive title, including the date the data were generated to allow the corresponding raw data records to be easily found. Whenever possible, we will follow the FAIR guiding principle for scientific data. The host center has sufficient storage space. Data capacity is monitored and augmented when needed.

The local proteomics data will be stored and backed up as part of the VIB Proteomics Core facility data analysis pipeline. This pipeline plugs into the server architecture of the VIB-UGent Center for Medical Biotechnology and is maintained by the IT responsible of this Center, Peter Van den Hemel.

Berx lab

All research data are stored in personal folders on the S drive on the servers of the VIB-UGent Center for inflammation Research (IRC) according to our data policy. Raw data obtained from flow cytometry and microscopy are stored on separate drives on the IRC servers. Data are stored three times on two different locations.

How will the data be backed up?

KU Leuven drives are backed-up according to the following scheme:

- data stored on the "L-drive" and "K-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 omics data stored on the Flemish Supercomputer Centre (VSC) and a backup is stored on the L-drive.

We also prevent loss of physical material eg. by scanning paper documents, newly created human cell lines will be stored locally in the laboratory in liquid nitrogen storage and will also be deposited in the UZ Leuven-KU Leuven Biobank, new mouse lines will be cryopreserved, the -80°C freezers are connected to an alarm system and a backup -80°C freezer is available through campus service,...

Scott lab

Data are backed up on two different locations, one in the main building of the host institute (FSVM I building) and one in the iGent building on the Tech Lane Ghent site. The frequency of backups depends on the type of data: research data every 24 hours, emails: every 8 hours and databases every 24 hours.

Gevaert Lab

All data are backed up indefinitely on the CMB servers, and daily backups are foreseen.

Berx lab

Data are backed up on two different locations, one in the main building of the host institute (FSVM I building) and one in the iGent building on the Tech Lane Ghent site. The frequency of backups depends on the type of data: research data every 24 hours, emails: every 8 hours and databases every 24 hours.

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

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

- the "L-drive" and "K-drive" are easily scalable systems, 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 controllers with an Ontap 9.1P9 operating system.
- the sequencing data is maintained on VSC storage and backed up on the L-drive

Scott lab

Storage is scalable depending on the need of project (similar for backup capacity). To prevent bad storage policy, personal and team folders are subjected to quota.

Gevaert Lab

The VIB Proteomics Core has a 20 TB (mirrored backup) storage server to store the raw data files during data analysis. Afterwards, the raw data files will be copied on (duplicate) tapes for indefinite storage. As mentioned above, raw data files are uploaded in a public repository for proteomics data (e.g. PRIDE) upon publication.

Berx lab

Storage is scalable depending on the need of project (similar for backup capacity). To prevent bad storage policy, personal and team folders are subjected to quota.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The "L-drive", "K-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 digital vault is possible only through using a KU Leuven user-id and password, and user rights only grant access to the data in their own vault. 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. Data on VSC is only accessible by VSC username, password, and MFA, and permissions on data folders are set for specific groups in case of GDPR.

Scott Lab

The following measurements are taken to ensure secure data storage and to prevent modification by unauthorised persons: controlled physical access to the building, firewalling (on both departmental and individual server levels), switched network & encrypted communications (prevents eavesdropping), VLAN-ing (network compartmentalisation) & MAC authentication, least-known ports for well-known services (security through obscurity), brute-force intrusion detection & isolation, individual account expiry in accordance with contract of employment, ACL's (Access Control Lists; application of Principle of Least Privilege).

Gevaert Lab

Raw files generated by LC-MS/MS are automatically copied from the instruments' computer to a secure server through the intranet. None of these computers is connected to the internet and can only be accessed remotely by getting through a series of securities. Once on the server, the file is manually copied by the MS responsible to a folder bearing the same code as generated by the YouTrack platform. These folders are created within a subfolder with the center's three letter code, this allows an additional layer of control for data security. This folder can only be accessed by the user, the MS responsible within the PRC, the department and the IT team. From this server the database searches are performed, thereby automatically writing the results as .txt files in the same folder. In this way one folder contains the raw unprocessed data and results. Once the server is full, the projects that have not been opened for a certain period are copied on tape and removed from the server. However, these data are easily recovered upon request from the IT department, who can temporarily put them back on the server.

Berx Lab

The following measurements are taken to ensure secure data storage and to prevent modification by unauthorised persons: controlled physical access to the building, firewalling (on both departmental and individual server levels), switched network & encrypted communications (prevents eavesdropping), VLAN-ing (network compartmentalisation) & MAC authentication, least-known ports for well-known services (security through obscurity), brute-force intrusion detection & isolation, individual account expiry in accordance with contract of employment, ACL's (Access Control Lists; application of Principle of Least Privilege).

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 (digital + physical) storage during the project is 100,000 euro. This estimation is based on the following costs:

- The costs of digital data storage are as follows: 104 €/TB/Year for the "L-drive"/"K-drive" and 519€/TB/Year for the "J-drive".
 - the Staging on VSC has been renewed in 2021 with 3 Pb of space; and the biomed NAS is scaleable and will start with 150 Tb space. Cloud storage is scalable.
 - Maintaining a mouse colony alive costs about 2,00 euro per month, excluding the costs of genotyping. When no experiment is planned with a particular mouse strain, and in compliance with the 3R's rule (<https://www.nc3rs.org.uk>), cryopreservation will thus be used to safeguard the strain, prevent genetic drift, loss of transgene and potential infections or breeding problems. Cryopreservation of sperm/embryos costs about 500 to 700 euro per genotype, plus a minimal annual storage fee (25 euro per strain for 250 to 500 embryos). Frozen specimens are kept in two separate liquid nitrogen tanks at two different sites on campus. When necessary, the costs of revitalization from cryopreserved sperm/embryos are about 1,100/600 euro. This cost is covered by the grant.
- Electricity costs for the -80° freezers present in the labs are included in general lab costs.
Data storage and backup costs are included in general lab costs.

4. 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 datasets. All datasets will be stored on the university's central servers with automatic back-up procedures for at least 10 years, conform the KU Leuven RDM policy. The costs (€104 per TB per year for "Large volume-storage") will be covered by the Marine labs.

Datasets collected in the context of clinical research, which fall under the scope of the Belgian Law of 7 May 2004, will be archived for 25 years, in agreement with UZ Leuven policy and the European Regulation 536/2014 on clinical trials of medicinal products for human use.

Gevaert Lab

All electronic data will be stored indefinitely. The local proteomics data of the Gevaert lab will be stored indefinitely on tape.

Where will these data be archived (stored and curated 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" and "K-drive"

-Tissue samples: Tissues will be stored locally in the laboratory.

-Omics 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). Animal cell lines will be stored in liquid nitrogen cryostorage of the laboratory.

-Genetically modified organisms: will be kept in the animal facility

-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 Data repositories, where they will be preserved indefinitely.

Gevaert Lab

Data will be archived on the CMB servers. The proteomics data will be stored indefinitely on tape.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

It is the intention to minimize data management costs by implementing standard procedures (internally generated operational procedures (https://en.wikipedia.org/wiki/Standard_operating_procedure) 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.

The costs for "Large volume-storage" on the KU Leuven servers (€104 per TB per year) will be covered by the different labs involved.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- Yes, in an Open Access repository
- Yes, in a restricted access repository (after approval, institutional access only, ...)

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 if possible. The datasets that contain human data will be shared in a restricted access repository such as EGA. 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.

Data with valuable IP will be protected prior to publication. We will comply with open access regulations of KU Leuven and all other involved institutes.

Large data (e.g. microscopy images/movies microfluidics) will be, due to data volume, available on request after which access will be granted to a restricted access repository.

Proteomics data will be freely available after release of accompanying publications via the PRIDE repository.

If access is restricted, please specify 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.

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 in the comment section per dataset or data type where appropriate.

- Yes, Ethical aspects

Patient data will be made available in restricted access repositories. Requests for access to the patient data will be assessed on a case-by-case basis by the Data Access Committee, KU/UZ Leuven and the project lead.

Where will the data be made available? If already known, please provide a repository per dataset or data type.

GEO will be used to share mouse sequencing data

EGA will be used to share human sequencing data

RDR will be used to share other data sets such as eg. spatial omics data

PRIDE will be used to share proteomics data

When will the data be made available?

The data will be made available upon acceptance of the publication of the results except if data sharing has to be postponed for a certain time, for example to protect IP during patent application.

Which data usage licenses are you going to provide? If none, please explain why.

We are going to use CC-BY-NC-SA-4.0

- Free to share and adapt.
- Give appropriate credit, indicate if changes were made.
- Do not use the material for commercial purposes.
- Distribute your contributions under the same license as the original.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

A permanent identifier is added to your data upon deposit in a repository such as EGA, GEO, RDR. Whenever possible we will use those DOI link in our publications.

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

Deposition of smaller datasets in data repositories is usually covered by the repository. For example in RDR every researcher can use 50 GB/year for free. If we will deposit larger datasets the costs will be covered by the lab of Chris Marine.

The costs for sharing physical data will be paid by the researcher requesting the materials.

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Metadata will be documented by the research and technical staff at the time of data collection and analysis.

Who will manage data storage and backup during the research project?

The research and technical staff will ensure data storage and back up, with support from Pieter Joris and Urbain Scherpereel for the KU Leuven drives and from the VIB-IRC (Scott lab).

Who will manage data preservation and sharing?

The PI's of this consortium are responsible for data preservation and sharing, with support from the research and technical staff involved in the project.

Who will update and implement this DMP?

The PIs are ultimately responsible for all data management during and after data collection, including implementing and updating the DMP, with the support of the labmanagers.