KU LEUVEN (KUL): KU LEUVEN C1 GRANT – IGLIO (D-2024-3081)

RESEARCH DATA SUMMARY

Dataset Name	***Description***	***New or Reused***	***Digital or Physical***	***Digital Data Type***	***Digital Data Format***	***Digital Data Volume***	***Physical Volume***
FACS data	.fcs files from flow cytometers, .xml tables after FlowJo analysis	Generate new data	Digital	*Experimental*	other: fcs, .xlsx	*<100 GB*	
Plate reader measurements	Absorbance, luminescence, fluorescence readings from microplate reader	Generate new data	Digital	*Experimental*	other: .xlsx	*<100 GB*	
Tumor volume measurements	Measurements of mouse tumors (volumetric analyses via MRI scans or BLI scanning, or neuro-deficit symptoms.	Generate new data	Digital	*Experimental*	other: .xlsx	*<100 GB*	
Genomics, transcriptomics and single-cell RNAseq (patient) data	Publicly available data from databases such as TCGA, GEO datasets, and published research papers	*Reuse existing data*	Digital	*Compiled / aggregated data*	.csv, .h5ad, .xlsx	*<100 GB*	
RNAseq (mouse) data	Data from mouse tumour single-cell profiling via 10x Genomics' scRNAseq	*Generate new data*	Digital	*Experimental*	.H5SEURAT, .h5ad	*<100 GB*	
Biological samples	Cell lysates from tumours, etc.	Generate new data	*Physical*	*Experimental*	*NA*	*NA*	*<100 vials*
Digital images (blots, MILAN, Xenium)	Western blots, cytokine arrays, etc	Generate new data	Digital	*Experimental*	other: snc, .tiff, .png	*<100 GB*	
Cell lines	CRISPR-Cas knockout cells	Generate new data	*Physical*	*Experimental*	*NA*	*NA*	*<100 vials*

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.	Existing data from: https://cellxgene.cziscience.com/collections/999f2a15-3d7e-440b-96ae-2c806799c08c
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,	Yes, animal, human & oncolytic virus data. Ethical approval for human (S59804, S61081, S62248, S61081, S67312) and animal (P148/2023) relevant experiments has already been granted. This applies to

please describe these issues further and refer to specific datasets or data types when appropriate.	data resulting from human material analyses and animal experiments, including 'tumor volume measurements' and 'FACS data'. We also have official approval of the Dienst Bioveiligheid en Biotechnologie of Flanders to work with some oncolytic viruses relevant for this project. This is the relevant file number: SBB 219 2023/1191.
Will you process personal data1? 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.	No
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 expect that our project will result in research data with potential for tech transfer and valorization. KU Leuven has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in several 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 publications don't need to be delayed. We are currently already in contact with LRD and are following their guidelines.
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.	No
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.	No

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).	Documentation will consist of electronic laboratory records, that contain all of the information of the performed experiment itself (Excel sheet-based metadata files). Those notes will describe the biological samples used/generated, experimental setup and protocols used, results generated, the links to the specific computer location as well as the names of the respective datasets. We will also maintain a metadata sheet with the connection between lab samples and files on our data storage, so that data files, lab samples, and experimental notes remain properly linked. Detailed protocols will be written, including research methods and practices for each experimental initiative. This will be stored in Word or Excel files. Furthermore, a logbook will be kept in Excel containing all steps that were taken to develop the final methodology, date of implementation and name of the researcher who carried out the experiment. Algorithms, scripts and software usage will be documented and stored alongside the electronic

	laboratory records, e.g. GraphPad Prism, FlowJo. Finally, we will also keep all the information (in dedicated Excel sheets, PDFs or Word-files) about purchased antibodies, cell-lines, mouse models and other analogue data-sources. Other relevant information about these reagents and tools (e.g., proof of antibody specificity) will be derived from initial standardization and optimization experiments and will be retained along with general research documentation/meta-data files.
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.	Yes. We will use various metadata standards as applicable for different experiment/datatypes, as already established elsewhere: https://fairsharing.org/. For instance, flow cytometry (https://flowrepository.org/), microscopy imaging (https://www.openmicroscopy.org/), qRT-PCR (http://miqe.gene-quantification.info/), and publicly available TCGA patient data analyses (https://gdc.cancer.gov/aboutdata/ data-standards) have very well-defined pre-established meta-data standards. In case we do not have a metadata standard available for a technique/datatype, a metadata of the numerical datasets will be created manually (e.g. based on the Dublin core metadata standard). For most of the data, metadata will be provided as readme, word or excel files, containing all settings and technical descriptions of the experiment and data. In parallel, detailed meta-data info will be integrated within the electronic laboratory records linked to each experiment (as described above)

DATA STORAGE & BACK-UP DURING THE RESEARCH PROJECT

Where will the data be stored?	Digital files will be stored on KU Leuven data storage servers. All data generated during the project will be stored on the local KU Leuven servers, PI computers, and backup hard drives, as well as on a local RAID storage available in the office. This will be initially located in the real-time folders (on lab provided laptops/PCs of the students or employees and local KU Leuven servers) and later only in the archive folders (archive is mirrored; on local KU Leuven servers, backup hard-drives as well as PI computers). Any algorithms, scripts or softwares originally generated during the project will be stored in private online git repositories of the PIs. As soon as the manuscript is publicly available, the repository will be changed to a public repository. Specific biological samples (e.g. cell lysates, protein/nucleic acid samples) will be stored in a freezer (-20°C or -80°C) while cell lines will be stored in liquid nitrogen.
How will the data be backed up?	Digital data will be stored on the university's secure network drives with automatic daily back-up procedures (e.g. J: Drive for confidential data and KUL Enterprise Box/OneDrive for non-confidential data. Data is also stored as a backup on lab computers, external hard drives, and on a local RAID 50 system in the office.
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. In the various KUL storage services (currently we have > 1 TB free data-space at our disposal), as well as the 36Tb local RAID 50 system present in the office.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	The “J-drive” and “KUL Enterprise Box or KUL OneDrive” servers are accessible only by laboratory members and PIs. Local storage systems like the RAID system and external hard drives are not accessible remotely and password protected.
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 storage during the project is 1000 EUR. This estimation is based on the following costs: - The costs of digital data storage are as follows: approximately 100 EUR/100 GB/Year for the “J-drive” and approximately 20 EUR/100 GB/Year for the “KUL Enterprise Box/OneDrive”.

DATA PRESERVATION AFTER THE END OF THE RESEARCH PROJECT

Which data will be retained for 10 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 relevant data. Cell lines will also be stored for at least or more than 10 years. Cell lysates and certain biological samples will be stored for at least 5 years (beyond 5 years, the degradation of these samples makes their storage “useless” and would require repeating the experiments to re-generate the data).
Where will these data be archived (stored and curated for the long-term)?	Data will be stored as described in section 4, as well as archived using KU Leuven’s longterm storage/large volume storage service.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	KU Leuven long term storage/large volume storage should cost around 100€/TB/year. The total volume of relevant data to be stored is expected to be significantly under that. The costs for archival data will be paid upfront from the C1 project funds.

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.	Yes, as open data and as restricted data (after approval, institutional access only)
If access is restricted, please specify who will be able to access the data and under what conditions.	Access to data will be granted upon request.
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)?	No, unless IP is still under filing.
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Digital data will be made publicly available as per the journals’ data availability policy, or available upon email request to the PI. Data that do not support publication will be made available upon reasonable request by email.
When will the data be made available?	Upon publication of research results
Which data usage licenses are you going to provide?	Creative Commons Licenses (CC BY) will be attached to the data deposited.

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.	Yes, a PID will be added upon deposit in a data repository
What are the expected costs for data sharing? How will these costs be covered?	No costs are expected for data sharing.

RESPONSIBILITIES

Who will manage data documentation and metadata during the research project?	The personnels associated with this project in each of the participating labs will be responsible for data documentation & metadata, under supervision of the PIs of each of the participating labs.
Who will manage data storage and backup during the research project?	Data management, storage and back up will be performed by the respective personnels associated with this project in each of the labs, under the supervision of the PI of each of the labs.
Who will manage data preservation and sharing?	The PIs of each of the respective participating labs (Prof. Abhishek D Garg, Prof. An Coosemans, Prof. Dirk Daelemans, Prof. Steven De Vleeschouwer) will be individually responsible for ensuring data preservation and reuse, for their specific labs.
Who will update and implement this DMP?	The coordinator/head PI of this consortium (Prof. Garg) bears the end responsibility of updating & implementing this DMP.