FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Dirk Daelemans (supervisor-spokesman)**
Contributor name(s) (+ ORCID) & roles	**Jim Baggen (co-supervisor)**
Project number [1] & title	G067423N - Identification of proviral host factor combinations as drug targets for viruses with pandemic potential
Funder(s) GrantID [2]	FWO
Affiliation(s)	☒ **KU Leuven** ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible: **Rega Institute** (https://ror.org/03w5j8p12)
Please provide a short project description	The ongoing COVID-19 pandemic illustrates the devastating effects that a new pathogen can have on human health and economies worldwide. To limit the consequences of future viral pandemics, it is crucial to have therapeutic strategies available to counter viruses with pandemic potential. One such strategy is the combined use of host-directed drugs that have synergistic antiviral activity, which could limit resistance development and cytotoxicity. Targeting host factors also has the benefit that these therapies may be effective against a broader range of viruses than virus-targeted therapies. In this project, we aim to identify targets for synergistic antiviral drugs against coronaviruses, influenza viruses, and filoviruses, which are important pathogens with pandemic potential. Therefore, we will perform genetic screens with SARS-CoV-2, Influenza A virus, and Ebola virus, to identify pairs of host genes that are required for their replication cycle. These screens involve simultaneous knockout of a massive number of unique gene pairs, a novel strategy that was never applied before in virus research. The identified genes will be potential targets for inhibitors that have synergistic antiviral effects, without compromising the viability of the host cell. Our findings will provide new insights into virus biology and may uncover combinations of host-directed drugs that could improve the efficacy of chemotherapy against these viruses with pandemic potential.

[1] "Project number" refers to the institutional project number. This question is optional. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

[3] Research Organization Registry Community. https://ror.org/

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
Cell lines	Calu-3 and Huh7 cell lines will be engineered to stably express Cas9. Cells will be stored in liquid nitrogen (A3.A340, Rega Institute).	☒ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	100*10^6 cells per cell line
Virus stocks	Three viruses will be used in this project: biologically-contained Ebola virus (EboV-ΔVP30), Influenza A virus (IAV), and SARS-CoV-2.	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Stock of 60X0.5ml = 30ml per virus
Plasmid stocks	A plasmid pool containing the Druggable paired CRISPR library	☒ Generate new data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images		☐ < 1 GB ☐ < 100 GB	500 microgram

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

	was newly made for this project, plasmids containing the Human Paralog Knockout Library will be reused. Plasmids containing single or dual sgRNAs behind human promoters will be prepared for screen validation.	☒ Reuse existing data		☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	s per plasmid
Lentiviral vectors	Vectors for sgRNA expression in human cells lines will be made for screen validation	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Stock of 10 x 1 mL per lentiviral vector
Deep sequencing data	In the final step of CRISPR screens, sgRNA sequences present in the cellular genomes will be determined by Illumina sequencing	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☒ Other:	.fastq	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Cytotoxicity data	To determine the cytopathic effect of each virus, cell viability will be measured using MTS assays. Absorbance will be measured and exported to Excel	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

Cell imaging data	Cell will be imaged using an Incucyte system. Confluence of cells can be measured, as well as red/green fluorescence objects. Output of this system consists of images and data that can are exported to Excel.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.tiff and .csv	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ > 5 TB ☐ NA	
qPCR data	Viral RNA production in wildtype or knockout cell lines will be determined by qPCR. Ct values are exported to Excel	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.csv	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Screen analysis scripts	New pipelines will be made to analyze sequencing data from combinatorial CRISPR screens	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	R scripts	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Data overviews	Results from different experiments will be collected in overview files to keep track of progress	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software	.pptx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

				☒ Other:			
Plasmid maps	DNA sequences of all generated plasmids will be stored digitally	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☒ Other:	.geneious	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Electronic lab notes	All experimental protocols and short descriptions of results and conclusions will be documented on the eLabNext platform.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☒ Other:	eLabNext	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Manuscripts	During all stages of manuscript preparation, text and figures will be stored in a shared folder. Text will be written in Word and figures are prepared using Adobe Illustrator	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.ai and .docx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

GUIDANCE: *The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.* *RDM Guidance on data*	
If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	The Human Paralog Knockout Library, which is available from Addgene (#171172), will be reused. This library was described in this publication: doi: 10.1016/j.celrep.2021.109597. Calu-3 cells were obtained from ATCC (HTB-55) Huh7 cells were obtained from Cell Line Service (https://cls.shop/HuH7/300156)
Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:
Will you process personal data[5]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
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:

[5] 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:
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:

3. Documentation and Metadata	
Clearly describe what approach will be followed to capture the accompanying information necessary to keep **data understandable and usable**, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded). *RDM guidance on documentation and metadata*.	All experiments will be described in an Electronic Lab Notebook (ELN, https://www.elabjournal.com/). Whenever possible, raw data will be uploaded to this platform and attached to the corresponding experiment. If datasets are too large or contain too many files for storage in ELN, these data will be stored on a shared drive (provided by KULeuven) in a folder that is clearly linked to the corresponding experiment in ELN. All modifications in ELN are tracked and a back-up of the shared drive is performed automatically. The following sections will be included for experiments in the ELN platform: - Background/aim - Protocol/Methodology and a description of all materials used in the experiment - Raw data and processed data (Excel) - Graphs (Graphpad) - A short report of the result with the most important graphs/images Conclusion Overviews of results from different experiments related to the same project will be prepared in Powerpoint. These data overviews will be stored on the shared drive.

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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: Metadata will be generated by instruments (SpectraMax Microplate Reader, Quantstudio II, confocal microscope etc.) and will be present in raw data files stored in the ELN per experiment. These metadata consist of the date/time of read-out and conditions/settings of measurements such as wavelengths, duration, temperature, or exposure time. In ELN, a search function is present which makes it easy to find and reuse data. Furthermore, each experiment is categorized within project groups, projects, and studies, which are made according to predefined rules in the lab.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☒ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☐ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☒ Other: Electronic Labjournal (DD group) www.elabjournal.com

How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☐ Other (specify)
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	☒ Yes ☐ No the J-drive provided by KU Leuven has enough capacity for storage of all data that will be gathered during the project. Next to this, for large data files (in this case deep sequencing data), a large volume storage drive (L drive) is provided. Automatic backups of the research unit servers are made daily and sufficient capacity is provided for these backups.
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* *Guidance on security for research data*	The research unit servers are only accessible to members of the research unit (password protection). The registered documents in ELN are only accessible to members of the research unit (password protection). Every modification is registered in ELN.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The internal storage drive is provided by the KU Leuven. Seats in the ELN are paid for yearly and are covered by the research unit.

5. Data Preservation after the end of the Research Project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☐ KU Leuven RDR ☒ Large Volume Storage (longterm for large volumes) ☒ Shared network drive (J-drive) ☒ Other (specifiy): Electronic Labjournal (DD group) www.elabjournal.com
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The internal storage drive is provided by the KU Leuven, costs are covered by the KU Leuven, also after the end of the project. The ELN seats of persons involved in this project may not be paid for after the end of the project, but data in ELN will remain accessible to all members of the research unit, so no extra costs are applicable.

6. Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *Note that 'available' does not necessarily mean that the data set becomes openly available, conditions for access and use may apply. Availability in this question thus entails both open & restricted access. For more information: https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☒ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: Articles will be published in Open Access. Certain experimental data (such as read count tables resulting from CRISPR screens) will be published together with the corresponding article. All other data will be stored in the J-drive of the institution and in ELN, which are both available to members of the research team.
If access is restricted, please specify who will be able to access the data and under what conditions.	Data used for publication in journals will have Open Access due to the Open Access policy. Data can be requested via email to Prof. Daelemans, following the signing of a data-sharing agreement. All data will remain available to the members of the research unit (on the J-drive or in ELN).
Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:

Where will the data be made available? If already known, please provide a repository per dataset or data type.	☐ KU Leuven RDR ☐ Other data repository (specify) ☒ Other (specify) Raw data belonging to a published manuscript will be available as Supplementary files that are published online with the manuscript.
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☐ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☒ No

What are the expected costs for data sharing? How will these costs be covered?	No costs are expected

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Jim Baggen will manage data documentation and metadata during the project. The head of the research group, Prof. Daelemans, bears the overall responsibility.
Who will manage data storage and backup during the research project?	Jim Baggen will manage data storage and backup during the project. After the end of the research project, Prof. Daelemans will have the responsibility.
Who will manage data preservation and sharing?	Jim Baggen will manage data preservation and sharing during the project. After the end of the research project, Prof. Daelemans will have the responsibility.
Who will update and implement this DMP?	Jim Baggen will update and implement this DMP during the project. Prof. Daelemans has the final responsibility.