FWO DMP Template - Flemish Standard Data Management Plan

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

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

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

1. General Project Information	
Name Grant Holder & ORCID	**Jannes Heylen, 0000-0002-5207-2159**
Contributor name(s) (+ ORCID) & roles	**Joost Wauters, 0000-0002-5983-3897, supervisor.**
Project number[1] & title	11PBR24N: From bed to bench and back in viral-associated pulmonary aspergillosis: generation of pathophysiological insight and translation into host-based biomarkers
Funder(s) GrantID[2]	
Affiliation(s)	KU Leuven
Please provide a short project description	Influenza- and COVID-19 associated pulmonary aspergillosis (IAPA/CAPA) are fungal superinfections in critically ill patients with influenza or COVID-19 pneumonia. These infections are both hard to diagnose and hard to treat, resulting in high morbidity and mortality in affected patients. Currently, pathophysiological insights are scarce, and no host-based biomarkers for viral-associated pulmonary aspergillosis (VAPA) patients are available in the clinic. This PhD project wants to enhance pathophysiological knowledge on VAPA and discover and validate host-based biomarkers able to assess VAPA risk, to improve diagnosis and to predict outcome of VAPA in critically ill patients with severe influenza/COVID-19. I envision three research objectives: (1) to assess whether circulating autoantibodies neutralizing interferons predispose to VAPA; (2) to explore whether immune cells of VAPA patients display distinct metabolic signatures compared to immune cells of critically ill patients with influenza/COVID-19 mono-infection; (3) to develop multivariate biomarker predictive models for VAPA via a multi-omics driven integrative systems biology approach using state-of-the-art machine learning based methods. I feel confident in reaching these research objectives, given the fact that our research group has a well-established research track on VAPA, combined with extensive collaboration with international experts in the fields of infectiology, fungal immunology and immunometabolomics.

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

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

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[3].

				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
'Mechanical Ventilation'	Both retrospective collection of clinical data from electronic patient medical records as prospective data collection in online eCFR format (REDcap, Castor). Data will be used throughout the PhD project to couple experimental	☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

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

	data of the metabolomics and autoantibody work packages to the relevant clinical information.						
Metabolomics	Mass spectrometry assays of blood and bronchoalveolar lavage samples of patients with severe influenza or COVID-19.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☒ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx, .d ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Autoantibodies	Detection of serum and BAL autoantibodies in samples from patients with severe influenza or COVID-19.	☐ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB	

					☐ .dwg ☐ .tab ☐ .gml ☒ other: .xlsx ☐ NA	☐ > 50 TB ☐ NA		
Patient samples	Blood and BAL samples of patients with severe influenza or COVID-19. Directly collected from the patiend during hospitalization.	Biological samples, stored at -20°C or -80°C until use.	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☒ Observational ☐ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☐ other: .xlsx ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

Guidance: *Data can be digital or physical (for example biobank, biological samples, …). Data type: Data are often grouped by type (observational, experimental etc.), format and/or collection/generation method.* *Examples of data types: observational (e.g. survey results, sensor readings, sensory observations); experimental (e.g. microscopy, spectroscopy, chromatograms, gene sequences); compiled/aggregated data[4] (e.g. text & data mining, derived variables, 3D modelling); simulation data (e.g. climate models); software, etc.* *Examples of data formats: tabular data (.por,. spss, structured text or mark-up file XML, .tab, .csv), textual data (.rtf, .xml, .txt), geospatial data (.dwg,. GML, ..), image data, audio data, video data, documentation & computational script.* *digital data volume: Please estimate the upper limit of the volume of the data per dataset or data type.* *physical volume: Please estimate the physical volume of the research materials (for example the number of relevant biological samples that need to be stored and preserved during the project and/or after).*	
If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type.	
Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: Clinical data, as well as human samples (blood, broncho-alveolar lavage) of hospitalized patients, collected during clinical routine, will be used for metabolomics and proteomics experiments. Data will be generated from outlined experimental protocols. Methods used to generate the data will be described as standard operating procedures and included in all publications. All experimental methods are embedded within trials approved by UZ/KU Leuven Ethical Committee (S65588, S62072, S63881 and S64259).

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

Will you process personal data[5]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes ☐ No If yes: Short description of the kind of personal data that will be used: Short description of the kind of personal data that will be used: clinical data of patients admitted to ICU with severe influenza or severe COVID-19, with or without aspergillosis. Data consists of demographic data, medical history before admission, clinical parameters and disease history at ICU admission, laboratory and radiographical parameters, and clinical outcome parameters.
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:
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:

[5] See Glossary Flemish Standard Data Management Plan

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).	Raw and processed data are collected in multiple formats. Numerical data are collected in Excel, .d format, .mzXML format, FCS file format and RData. Textual data are generated in pdf, txt, and word format. Raw output data of luciferase reporter assays and enzyme-linked immunosorbent assays will be collected in Excel files. FACS assays will generate data structured in FCS file format, containing the acquired raw data and experiment-specific metadata. Mass spectrometry assays will generate raw data in. d format compatible with MassHunter software, which will be used for data acquisition and conversion in .mzXML format. Research methods and practices (SOPs) are fully documented. All data, raw and processed, will be named in a correct and clear way according to predefined folder structures, so data can be easily identified and reused. When wet lab techniques, scripts, algorithms and software tools are finalized, they are additionally described in manuscripts and/or on GitHub.
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 from in vitro analyses (autoantibody research and metabolomics) are automatically generated by the equipment used (Gyros automated ELISA and scimaX Bruker respectively. Furthermore, metadata files containing information connecting sample IDs with lab samples and patient IDs will be carefully collected and stored in digital logs (.xls)

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	Data will be stored in a secured, password protected servers of UZ Leuven, within a shared C://folder of the research group. UZ Leuven servers are backed-up automatically on a daily basis. There is sufficient storage and back-up capacity on all used servers All patient samples will be stored in UZ Leuven in labelled tubes in -20°C or -80°C freezers purchased by own funding. The samples will be registered and handled according to the UZ Leuven Biobank guidelines.
How will the data be backed up? *What storage and backup procedures will be in place to prevent data loss? Describe the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research.*[6] *Refer to institution-specific policies regarding backup procedures when appropriate.*	See question above.
Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely. If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.	☒ Yes ☐ No If yes, please specify concisely: see question 'where will the data be stored?' If no, please specify:

[6] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *Clearly describe the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.* [7]	Observational clinical patient information as well as metabolomics and proteomics data and associated pseudonymized patient/clinical information is considered sensitive information and will be handled as such. The data generated in this project will be processed and stored on the UZ Leuven IT infrastructure which are protected by a genuine user authentication system relying on username and password. Access to the data as well as the access level will be limited on a project need and individual basis. Only the researchers working on the project have access to these data. Due to the sample labelling as protective measure, the researchers are not able to decipher the identity of the donor. The coding key to patient information of linked pseudonymized data does not carry any personal identifiers and all records containing the identity of each participant will be kept private and confidential. The coding key will be kept by the PI, ICU physician prof. dr. Joost Wauters
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	No additional costs are required for data storage during the project, since existing infrastructure of UZ and KU Leuven can be used.

5. Data Preservation after the end of the Research Project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).	All electronical data collected in the two first work packages of this FWO project will be retained for the expected 5 year period after the end of the project. All (remaining) biological samples are preserved for 50 years, in accordance with the guidelines of Biobank UZ/KU Leuven. The output of the analyzed datasets and the models will also be retained. Processed datasets, including cleaned and analyzed data, as well as intermediate data generated during data processing pipelines, will be retained. Comprehensive documentation, including project plans, data dictionaries, codebooks, and metadata, will be explained on the manuscript that will be uploaded in Lirias. Source code and algorithms used for data analysis and modeling will be preserved to ensure reproducibility of the results.
Where will these data be archived (stored and curated for the long-term)?	Data will be stored in the UZ Leuven Large Volume Storage (100 GB, hourly backups, additional storage possible).
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	No additional costs are expected.

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *Note that 'available' does not necessarily mean that the data set becomes openly available, conditions for access and use may apply. Availability in this question thus entails both open & restricted access. For more information: https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights*	☐ Yes, in an Open Access repository ☒ Yes, in a restricted access repository (after approval, institutional access only, …) Institutional access only, by authorized personnel within the research team led by prof. dr. Joost Wauters. ☐ No (closed access) ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	By authorized personnel of the research team led by prof. dr. Joost Wauters upon reuse within future research projects.
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.	

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	Data will be made available upon reasonable request and upon publication of research results in peer-reviewed scientific journals.
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators." [7]*	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 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. A CC-BY license will be opted for when possible. For data shared directly by the PI (and approval of the 3rd party if necessary), 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.
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☒ Yes ☐ No If yes:

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	PhD student Jannes Heylen will manage data documentation and metadata during the project, under supervision of Prof. Dr. Joost Wauters and assisted by two clinical trial assistants: Cato Jacobs and Hanne Moon Lauwers
Who will manage data storage and backup during the research project?	PhD student Jannes Heylen will manage data storage and backup this DMP during the project, under supervision of Prof. Dr. Joost Wauters and assisted by two clinical trial assistants: Cato Jacobs and Hanne Moon Lauwers
Who will manage data preservation and sharing?	PhD student Jannes Heylen will update and implement this DMP during the project, under supervision of Prof. Dr. Joost Wauters and assisted by two clinical trial assistants: Cato Jacobs and Hanne Moon Lauwers.
Who will update and implement this DMP?	PhD student Jannes Heylen will update and implement this DMP during the project, under supervision of Prof. Dr. Joost Wauters and assisted by two clinical trial assistants: Cato Jacobs and Hanne Moon Lauwers