1. General Project Information	
Name Grant Holder & ORCID	**Xing Yang 0000-0002-8403-8254 – Promoter**
Contributor name(s) (+ ORCID) & roles	**Rik Gijsbers 0000-0003-0191-3904 – Copromoter**
Project number [1] & title	**3E240676 – Development of novel membrane chromatography platform for viral vector purification**
Funder(s) GrantID [2]	**Industrial Research Fund – C3**
Affiliation(s)	**KU Leuven (KUL)**
Please provide a short project description	Traditionally recombinant adeno-associated viral vectors (rAAVs) are purified using ultracentrifugation for small scale use. Increased demand for rAAVs to fuel novel gene therapy developments has pushed the field toward resin chromatography, which however face challenges due to low throughput and efficiency. Membrane chromatography is emerging as a promising alternative, but its development is still at infancy stage. Thus, we aim to develop a new, generic membrane chromatography platform for purifying the rAAV particles guided by a combined experimental-modeling approach. Novel high-throughput membranes will be fabricated with fast mass transfer & improved selectivity towards full-genome rAAV particles. Machine learning approach will assist this major leap towards fast optimization and hence the establishment of a generic rAAV downstream purification platform.

[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.

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
		☐ 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	
Lab notes	Description of the practical execution of experiments	New	Digital and if physical, they will be digitalized	Observational and experimental	NA	NA	4-5 note Books & OneNote
Electrospinning parameters	Conditions for membrane fabrication	New	Digital	Experimental	.xlsx	< 1 GB	NA
Fourier-transformed Infrared Spectroscopy	Characterization of different chemistries	New	Digital	Experimental	.xlsx .spc .0	< 1 GB	NA

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

(FTIR)							
X-ray Diffraction (XRD)	Structural characterization	New	Digital	Experimental	.raw ,dif	< 1 GB	NA
Scanning Electron Microscopy (SEM)	Surface morphology	New	Digital	Experimental	.tif .png .jpg	< 100 GB	NA
Transmission Electron Microscopy (TEM)	Morphology and composition of materials at nanometer scales.	New	Digital	Experimental	.tif .png .jpg	< 100 GB	NA
Binding Capacity and Selectivity Data	Experimental results of binding tests under various conditions	New	Digital	Experimental	.xlsx	< 1 GB	NA
rAAV Binding & Separation Results	Membrane performance for rAAV full/empty capsid separation	New	Digital	Experimental	.xlsx	< 1 GB	NA
Dynamic Elution Profiles	Data showing elution conditions for optimized rAAV recovery	New	Digital	Experimental	.xlsx .csv	< 1 GB	NA
Purity and Yield Results	Quantitative analysis of rAAV samples after separation	New	Digital	Experimental	.xlsx	< 1 GB	NA
Machine Learning Models	Models predicting membrane performance and design optimization	New	Digital	Modeling	.m .docx .xlsx .py	< 1 TB	NA
AAV preps	Use of existing AAV vectors & Production of new AAV vectors	New and existing	Physical/digital	Experimental	.xlsx	< 1 GB	-80 storage at LVVC ; BSL2

AAV QC	QPCR & ddPCR, for genome copies, Western Blot for VP1/2/3, SYPRORuby total protein staining, mass photometry (F/E)	new	digital	Experimental	.xlsx .ddpcr .pltd .pcrd .scn .jpg .tiff .pdf .png	< 100 GB	NA
AAV functional testing	Transduction of reporter cells (HEK293T(+AAV-R))	New	digital	Experimental	.fcs .pdf .xlsx .csv .txt	< 10 GB	
Cell culture	Seeding and culturing of producer cells	New and existing	physical	Experimental	NA	NA	3-4 Notebooks

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.	NA
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: Ethic approval is pending– S nr (S70070 - Productie van virale vectoren voor het maken van celmodellen, diermodellen en gentherapeutische toepassingen). The cells we use are laboratory cells, purchased from ATCC and anonymous.

Will you process personal data[4]? 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: The project has strong potential for commercial valorization. The novel membrane chromatography platform under development is designed to address current bottlenecks in the purification of recombinant adeno-associated viral (rAAV) vectors, a growing market driven by advancements in gene therapy. Specific datasets with commercial potential include: **Performance Data:** • These datasets will validate the scalability, efficiency, and selectivity of the membranes, essential for IPR protection and/or licensing or direct application in industrial settings. **Membrane Design and Fabrication Data:** • Parameters guiding the membrane fabrication process (e.g., ligand density, pH responsiveness) could be commercially valuable for the production of tailored membranes. Data will be considered for IPR protection. **Modeling Data:** • These datasets can assist in optimizing membrane performance and may be offered under fee-for-service agreements or through collaborations with industrial stakeholders.
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: The project includes collaborations that may involve Material Transfer Agreements (MTAs) or similar agreements with industrial stakeholders.
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: The project involves legal considerations related to intellectual property rights (IPR) and ownership, particularly regarding the novel membrane technologies being developed. These include: 1. **Ownership of Developed Data:** ○ Data generated during the project (e.g., membrane fabrication protocols, performance

[4] See Glossary Flemish Standard Data Management Plan

	metrics, and modeling results) will be owned by KU Leuven and will be considered to support IPR protection of the novel membrane technology/product. 2. **Patents and IPR:** ○ An IPR protection strategy will be developed covering project data related to membrane compositions and methodologies for rAAV purification, . The project's valorization plan includes a licensing strategy and strategic industry collaborations. 3. **Third-Party Restrictions:** ○ Some materials obtained from Third Parties used for benchmarking our novel technology in this project, provided under Material Transfer Agreements (MTAs), may have restrictions on the use or disclosure of proprietary details. However, performance data generated using these materials is free from such restrictions. 4. **Collaborator Agreements:** ○ Legal agreements with industrial collaborators and suppliers will define data ownership and usage rights, ensuring that all project outcomes comply with contractual obligations.

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*.**	**The approach to capture the accompanying information necessary to keep data understandable and usable is dependent on the types of data:** **1. Data Related to Experimental Design, Testing Conditions/Settings, and Raw Data** • Electronic Lab Notebooks (ELN): The setup design, synthesis protocols, and observations for membrane fabrication and parametric studies will be recorded in electronic lab notebooks. This includes data related to pH-responsive membrane fabrication, ligand density, and performance testing under various conditions. These notes will be securely stored on KU Leuven's drives, ensuring centralized access. • Manuscripts: Scientific results will be communicated through manuscripts created in .docx, .pdf, or .tex formats and stored in project-specific directories. • Presentations: Presentations summarizing findings will be prepared using PowerPoint and stored in .ppt or .pdf formats for dissemination. **2. Data Related to Analytical Measurements**

	• **Membrane Characterization:** Data from morphological and surface chemical characterization (e.g., SEM, TEM, FTIR) will follow standard naming conventions (e.g., sample_type_conditions). These datasets will be saved in formats such as .tif, .csv, or .xlsx. • **Performance Evaluation:** Binding capacity and separation efficiency metrics from filtration experiments will be recorded digitally in .xlsx format. Specific conditions (e.g., pH, ionic strength, and serotype variations) will be annotated in the dataset headers. • **Sample Naming:** Each experiment or sample will use a unique identifier reflecting the work package (WP) and experimental conditions (e.g., WP1_membraneType_date). **3. Data Related to Modeling or Computational Work** • **Machine Learning Models:** Data used to train and validate machine learning models for membrane optimization will be labelled systematically (e.g., WP3_training_dataset_conditions). The generated predictive models and outputs will include metadata explaining their structure and key parameters. • **Simulation Data:** Computational Fluid Dynamics (CFD) simulations and parametric studies will produce files in .mph (COMSOL) format, accompanied by .xlsx plots and .docx descriptions of operating conditions.
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: For this project, metadata will be custom-created and tailored to the specific datasets to ensure usability and reproducibility. Key elements will include: 1. **Experimental Data:** o Metadata will detail sample preparation protocols, testing conditions (e.g., pH, ionic strength, temperature), and instrument settings (e.g., for SEM, TEM, or spectroscopy). o File naming conventions will clearly link datasets to specific work packages (e.g., WP1_membraneType_conditions).

	2. **Modeling Data:** ○ Metadata will describe computational parameters, boundary conditions, and software versions (e.g., COMSOL, MATLAB) used in simulations. ○ Each modeling dataset will be accompanied by a .docx file summarizing the context, inputs, and outputs. 3. **General Practices:** ○ README files will accompany all datasets, explaining file structures, formats, and key variables. ○ Metadata will be stored alongside datasets to facilitate reuse by both internal collaborators and external researchers upon data sharing.

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 : Archive (K-drive) ☐ Digital Vault ☐ Other:
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) Data are backed up on the cloud (OneDrive) immediately. The software indicates the update status (green, blue or red) and, in case of a nonsync, action can be taken using the online version of the tool. Data are further back up regularly on an external hard drive. After completion of (sub)WPs, data will be additionally backed up on the KUL service servers.
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 Maximum data storage should not exceed 2 TB per project.

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 accesses to OneDrive and SharePoint are only for researchers with permission. All users need to use a two-factor Authenticator (2FA app used at KUL). Furthermore, a log-out is always performed when leaving Lab PCs (where data is generated) to prevent modification of parameters by unauthorized people.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The estimated costs for data storage and back up during the project will not exceed 5000€. These costs will be covered and shared by both PIs of the project.

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) All the data will be retained for the expected 5 years period after the end of the project with no exceptions.
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): All the data will be stored for the long-term on the KUL service servers (with automatic back-up procedures), conforming with KUL RDM policy.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The estimated costs for data storage and archiving after the project will not exceed 5000€. These costs will be covered and shared by both PIs of the project.

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: Data relevant for publication will be made available in an Open Access repository (i.e.,Lirias). Full datasets will only be made available upon request. In any case, IPR protection and/or licensing will be prioritized which might delay or impede data sharing with the wider research community.
If access is restricted, please specify who will be able to access the data and under what conditions.	If data restriction occurs in light of the IPR/licensing strategy developed in the project, data access will be guided by the (exclusive) licensing agreements with third parties which will be installed with the support of LRD.

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: Some restrictions exist due to intellectual property considerations and potential agreements with collaborators or industrial partners. Specifically: 1. **Patentable Data:** o Data related to the novel membrane compositions, fabrication methods, and downstream processing performance may be restricted until the intellectual property (IP) is secured through patent filings. This includes datasets demonstrating: ▪ The pH-responsive characteristics of the membranes. ▪ Binding efficiency and selectivity of the membranes for full vs. empty rAAV capsids. 2. **Collaborative Restrictions:** o Data generated in collaboration with industrial stakeholders, including companies that may supply materials or funding, could be subject to agreements limiting early dissemination. Specific terms will depend on the agreements established during the project. 3. **Proprietary Testing Data:** o Data involving the use of proprietary commercial membranes or resins for benchmarking purposes may have restrictions on sharing the material-specific results, as governed by Material Transfer Agreements (MTAs). However, general performance comparisons and analysis will remain shareable.
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)
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) This is mainly applicable to data generated by non-open source software. For MATLAB Codes: MATLAB License will be provided.
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?	The expected cost for data sharing is 0€. Free tools like Belnet FileSender (KUL account) will be used for data sharing. In the unlikely event that there would be costs, these costs will be covered by project budget, or both PIs of the project.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Prof. Xing Yang Prof. Rik Gijsbers
Who will manage data storage and backup during the research project?	Prof. Xing Yang Prof. Rik Gijsbers
Who will manage data preservation and sharing?	Prof. Xing Yang Prof. Rik Gijsbers
Who will update and implement this DMP?	Prof. Xing Yang and Prof. Rik Gijsbers bear the end responsibility of updating and implementing this DMP in the long term.