

DMP title

Project Name Schlenner_2021 (FWO DMP) - DMP title

Grant Title G054722N

Principal Investigator / Researcher Susan Schlenner

Institution KU Leuven

1. General Information

Name applicant

Susan Schlenner

FWO Project Number & Title

G054722N

"Functional dissection of RNA modifiers in immune cells using CROPseq"

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Type of data	Format	Volume	How created
bulk/single-cell RNAseq	quantitative tabular data: comm-separated value files (.csv), MS Excel (.xls/.xlsx)	max. 5GB per sample	experimental: RNA sequencing computational: data QC, generate gene list, statistics
flow cytometric data	flow cytometric standard (.fcs)	max. 10GB per sample	experimental: Analysis of cells on Flow cytometers

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

n/a

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, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

Approvals by the Leuven Laboratory Animal Ethics Committee:

008/2019, 150/2019, 000/(GS1/GS2)-breeding/Schlenner

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what

data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

1. omics data (RNA sequencing)

The methodology to generate the data as well as the analysis procedures will be documented (see metadata). Versioning of generation methodology does not apply. Versioning of analysis procedures will apply.

2. flow cytometric data

Upon publication, all the data and methods supporting a manuscript will be made publicly available.

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- Yes

omics data:

The final dataset as deposited in the chosen data repository will be accompanied by this information under the form of a README.txt document.

The following metadata will be provided:

Title/Creator/Date reference/Subject/project/Sample information/description (explaining the content of the data set and other contextual information needed for the correct interpretation of the data, the software(s) (including version number) used to produce and to read the data, the purpose of the experiment, etc.)/Organism/File format/Resource Type (data set, etc.)/Identifier (DOI = when applicable)/Access rights (e.g. closed access, embargoed access, restricted access, open access).

For specific datasets, additional metadata will be associated with the data file (experiment design/protocol).

5. Data storage and backup during the FWO project

Where will the data be stored?

All data will be stored centrally on storage facilities of the university at an external data center (L-drive).

This server is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end. Stored data 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.

Upon publication and depending on the data type, data will be made available via benchling.org, the public repository GEO at NCBI, the public repository FlowRepository (<https://flowrepository.org>) and the SRA at NCBI. The datasets will be cited using the Digital Object Identifier (DOI) link generated by the repository, and will also be accessible via the search page on the corresponding website.

How is backup of the data provided?

Digital files will be stored on KU Leuven servers:

the "L-drive": Stored data 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.

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

Data will be stored on dedicated drives on KU Leuven storage servers (capacity is unlimited, available in blocks of 100 GB).

What are the expected costs for data storage and back up during the project? How will these costs be covered?

<200 Euro per TB and year. Cost will be covered by grant funding.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be stored in the university's secure environment; access to the data is password-restricted.

6. Data preservation after the FWO project

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

All data.

Where will the data be archived (= stored for the longer term)?

After the research all data will be archived in permanent archival drives on KU Leuven storage servers (capacity is unlimited, available in blocks of 100 GB) for a minimum of 5 year after the publication of the results or the end of the period of the project funding.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

<200 Euro per TB and year. Cost will be covered by grant funding.

7. Data sharing and reuse

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes. Specify:

Data will be shared upon publication.

Which data will be made available after the end of the project?

Published RNAseq and flow cytometry data will be made available.

Where/how will the data be made available for reuse?

Upon publication and depending on the data type, data will be made available via benchling.org, the public repository GEO at NCBI, the public repository FlowRepository (<https://flowrepository.org>) and the SRA at NCBI.

When will the data be made available?

- Upon publication of the research results

Who will be able to access the data and under what conditions?

Anyone with access to the Repositories will have access to the published data. / Open Data.

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

none.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Susan Schlenner

Suresh Poovathingal

Who will be responsible for data storage & back up during the project?

Susan Schlenner

Who will be responsible for ensuring data preservation and reuse ?

Susan Schlenner

Who bears the end responsibility for updating & implementing this DMP?

The PI bears the end responsibility of updating & implementing this DMP.