Exploring the under-exploited bacterial purine biosynthesis pathway towards developing novel antibiotics and biochemical tools.

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.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		*Indicate: **N**(ew data) or **E**(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
Synthetic procedures	Experimental conditions leading to molecules	N	D	T	Online ELN (can be exported as .pdf)	<1GB	
Structured synthetic details	A fully written out version of the experiment, with an interpretation of the main analytical results	N	D	Compiled data	.docx and .pdf	<1GB	
NMR data	Raw and processed NMR data	N	D	Experimental	.fid	<100GB	
Mass data	Processed Mass data	N	D	I and T	.pdf	<1GB	
Flash column	Processed Flash column purification profiles	N	D	I and T	.pdf	<1GB	
Stardrop analyses	Prediction of physicochemical properties and hypotesis formulations	N	D	Simulation Data (N and T)	.sdproj	<1GB	
Monthly progress meetings	Slides used during monthly individual progress meetings	N	D	Compiled Data	.pptx	<100GB	
Compound Index	Index of available intermediate and final molecules with indications of physical locations	N	D	Compiled Data	.dwar	<1GB	
Compound Index	Physical storage of final compounds synthesized in this project	N	P				<10 storage sample boxes
In vitro enzymatic inhibition	Processed IC50 data	N	D	N	.xlsx and .dwar	<1GB	
DEL screening results	Processed results of NGS analyses and annoted processed results of DEL screen	N	D	N and T	.csv, .xlsx and .dwar	<100GB	
Bacterial assays	Processed results of bacterial susceptibility assays (MIC, BIC50)	N	D	N	.xlsx	<1GB	
Mamalian cell toxicity	Processed results of mamalian toxicity (cell viability)	N	D	N	.xlsx	<1GB	

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:

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, 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 below)

Ethical clearance for cell-based viability assays on muman cell lines will be applied for. As indicated in the CMT, this is planned for the final year of the project. If the project advances faster than anticipated the approval application will be moved forward.

Will you process personal data? 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).

- 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

The data could become a patentable invention with potential for commercial exploitation: novel antibiotics will be developed.
As such only data related to published work will be shared with the scientific community. All non-published results will remain confidential, but subject to regular review in function of valorization potential.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

Yes, in the framework of this project a collaboration agreement with a commercial enterprise is being considered, after some initial exploratory discussions that indicate aligned interests. For the moment only a CDA is in place, which puts no restraints on the data dissemination of current data (being) obtained at KU Leuven.
All data recorded in the framework of the collaboration agreement will be subject to confidentiality unless agreed by both parties to disseminate results. In the drafting of the collaboration agreement, data dissemination will be a key point of attention.

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 in the comment section to what data they relate and which restrictions will be asserted.

- Yes

As mentioned above, all data currently obtained by KU Leuven are owned by KU Leuven. To allow for valorisation, all data of non-published work will remain confidential, only data related to published work can be shared.
The exact ownership of data obtained within the potential collaboration agreement will be subject to stipulations still to be negotiated.

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).

Datasets will be accompanied with a README file, clearly identifying the compound for which data has been recorded. This will initially link to internal unique compound codes.
All information required to identify the data, will be added where appropriate. For example the equipment used to record the data, definitions of variables, units of measurement, the identity of the researcher who obtained the data. All identifiers will be amended/replaced with their identifiers in the associated publications, where appropriate when data is available to be shared.

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.

- No

Data will be accompanied with a README file and file names will include a unique molecule identifier code

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Shared network drive (J-drive)
- Other (specify below)

Hard drives of (PhD) students and researchers, regularly to be uploaded to the shared J drive

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution
- Personal back-ups I make (specify below)

Data will additionally be backed up on external hard disks.

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

For the J drive, data is not accessible for unauthorized persons. For each folder a "green list" of users can and will be defined. Physical backups (e.g. external hard drives) will be stored in locked cabinets.

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

No additional costs, the currently available storage space is sufficient.

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...).

- All data will be preserved for 10 years according to KU Leuven RDM policy
- Certain data cannot be kept for 10 years (explain below)

In case of a clause in an agreement with a commercial company which would ask certain data (owned and generated at an external partner) to be removed at the end of the contract, these data will not be retained. This will be subject to contract negotiations.

Where will these data be archived (stored and curated for the long-term)?

- Large Volume Storage (longterm for large volumes)
- KU Leuven RDR
- Other (specify below)

External hard drive

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

10 Years storage, up to 5TB (Large volume storage) at KU Leuven server: approx. 5000 euro*
2 External Hard drives up to 1TB (SSD): approx. 400 euro.
Costs for this storage are covered by the research grant.
*At the end of the project, all options will be assessed once more.

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

As mentioned above, only for data related to publised results

If access is restricted, please specify who will be able to access the data and under what conditions.

Data of unpublished results will not be made available. Access will only be authorised to the primary researchers and the PI.

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, intellectual property rights

only published results will be shared, to not compromise the project valorization strategy

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)

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?

The size of shared data will not exceed the yearly allowance of 50 GB/year at the KU Leuven RDR.
For compound samples, when available, costs will be paid for by the requesting party.

Responsibilities

Who will manage data documentation and metadata during the research project?

Primary researchers

Who will manage data storage and backup during the research project?

Primary researchers and PI

Who will manage data preservation and sharing?

PI

Who will update and implement this DMP?

PI