Causally informed development and validation of a data-driven bio-mathematical reclassification of kidney transplant rejection

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	
TEMPLATE+	All kidney patients transplanted between March 2004 and May 2021 at UZ Leuven	E	D	N	.xlsx	<1GB	1GB
BHOT	biopsy-based transcripts	E	D	N	.xlsx	<100GB	10GB

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:

TEMPLATE+: clinicaltrials.gov NCT06505200
BHOT: https://doi.org/10.1111/ajt.16059

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)

- TEMPLATE+: approved by Ethics Committee of the University Hospitals Leuven (S64006)
- BHOT: approved by Ethics Committee of the University Hospitals Leuven (S63996 and S64904)

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).

- Yes (Provide PRET G-number or EC S-number below)

TEMPLATE+ and BHOT : data are pseudonymised. Via a personal identification number, patients can still be identified by the managing hospital when in need for data cleaning or additional collection. During data analysis, the patient is however not identifiable by the researcher (S64006 for TEMPLATE+; S64904 and S63996 for BHOT).

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

Although the principal purpose of the C2M project is societal valorization, the models on the BHOT panel could become relevant for protection of IP and ultimately commercial valorization. This is included in the EC-approved project proposals of S64904 and S63996, and in the contract with the external partner performing the BHOT analyses.

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

For S63996 the agreement with the 3rd party (CareDx) restricts exploitation. While KU Leuven has broad rights, the 3rd party has some rights, as outlined in the detailed contract. These rights relate to the valorization of the models developed, but also to the data itself.

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

All of this is outlined in the contract with CareDx, in the context of S63996. This relates both to the clinical data and to the BHOT expression data.

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).

- TEMPLATE+: within the excel file, an extra tab is included, describing all variables (definition, units of measurement, values). These data are stored on the KU Leuven Onedrive servers. As our data are sensitive patient data, these will not be shared in a repository. Data can be made available to third parties via a Data Transfer Agreement. All contracts with third parties are managed by the KU Leuven legal department (LRD) or the clinical trial center of UZ Leuven.
- BHOT: for these data, files with raw gene expression data are present, and this included the expression values per sample and gene as well as the names of each gene ,and the Accession number. Demographic data related to these samples is present in a separate file. There is a common sample ID between both datasets, enabling merging them. All procedures used for the BHOT analyses are stored in a separate procedure file in the same location.

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

TEMPLATE+ and BHOT: additional excel tab with description all variables is in place.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- Other (specify below)

TEMPLATE+ and BHOT: stored on KU Leuven Onedrive (with automatic back-up procedures).

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Automatic back-up procedures are in place.

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

KUL Onedrive offers a standard of 2TB for each user, which is sufficient for this project.

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

TEMPLATE+ and BHOT: only the researchers working on the project can access the data. Only the PI (Maarten Naesens) can change the content in the mother database.

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

No extra costs are expected on top of the provided standard storage space by the KU Leuven.

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

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

- Other (specify below)

TEMPLATE+ and BHOT: KU Leuven Onedrive.

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

Additional on the standard data storage infrastructure (Onedrive) provided by the KU Leuven, no extra costs for data preservation are expected.

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.

- No (closed access)

TEMPLATE+ and BHOT: raw data are only accessible by researchers involved in the project. Data can only be made available to third parties via a Data Transfer Agreement. All contracts with third parties are managed by the KU Leuven legal department (LRD) or the clinical trial center of UZ Leuven.
Data will be available in the form of publications or other dissemination of scientific work.

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

TEMPLATE+ and BHOT: only researchers working on the project have access to the raw data. Only the PI (Maarten Naesens) can change the content in the mother database.

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

TEMPLATE+ and BHOT: sensitive patient data. Raw data can only be shared with a DTA, according to GDPR regulations.

Where will the data be made available?

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

- Other (specify below)

TEMPLATE+ and BHOT: raw data cannot be shared.

When will the data be made available?

- Upon publication of research results

Data will only be available in the form of publications or other dissemination of scientific work.

Which data usage licenses are you going to provide?

If none, please explain why.

- Data Transfer Agreement (restricted 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.

- No

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

TEMPLATE+ and BHOT:: costs of data sharing will be negotiated in the DTA.

Responsibilities

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

Maarten Naesens

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

Maarten Naesens

Who will manage data preservation and sharing?

Maarten Naesens

Who will update and implement this DMP?

Maarten Coemans