Plan Overview

A Data Management Plan created using DMPonline

Title: Radboudumc Mycobacterial Cohort Study (MyCoS)

Creator: Lisa Kurver

Principal Investigator: Arjan van Laarhoven, Wouter Hoefsloot

Contributor: Wouter Peeters, Lisa Kurver, Arthur Lemson

Affiliation: Radboud University Medical Center (Radboudumc)

Template: Radboudumc Data Management Plan

ORCID iD: 0000-0002-6607-4075

ORCID iD: 0000-0003-1007-1075

Project abstract:

Radboudumc is a center of expertise for the diagnosis and treatment of mycobacterial infections, with approximately 150 cases yearly. Clinical diagnosis is often difficult and treatment is hampered by long and sometimes toxic antibiotic regimens. The multidisciplinary approach of Radboudumc provides an excellent setting to establish the immunological, pharmalogical, and microbiological phenotype and relate this to patiënt genetics. This will improve our knowledge on mycobacterial infections and will ultimately help our physicians in the treatment of mycobacterial infections and will therefore improve quality of care.

ID: 88801

Last modified: 02-05-2025

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Radboudumc Mycobacterial Cohort Study (MyCoS) - Full DMP (mandatory for all)

1. Project info

1.1 DMP version and date

Version 1.0 / 24-Nov-2021
Version 2.0 / 23-01-2023
Version 3.0 / 31-01-2024

1.2 Name of data management support staff consulted during the preparation of this plan and date of consultation.

Mirjam Brullemans
24-01-2023

1.3 Does the project consist of multiple (sub)projects?

- No

1.4 Project number(s)

n/a

1.5 Project leader (PI); provide contact information (Name, email address, phone number)

Arjan van Laarhoven, Arjan.vanLaarhoven@radboudumc.nl
Wouter Hoefsloot, Wouter.Hoefsloot@radboudumc.nl

1.6 Research Institute

- RIMLS

1.7 Will non human research be performed in this project?

- No

1.8 Will research involving human participants (nWMO/WMO) be performed in this project?

- Yes, WMO research (please specify which (sub)projects)

2. Planning and design

2.1 Will this research project involve collaboration with other parties? (E.g. in collecting, processing, analyzing and/or publishing the data)

- No

2.4 Who are the persons involved in data management? Mention all persons that have access to the data, and in which role they will participate.

Institute	Name	Role
Radboudumc	Arjan van Laarhoven	Data collection, data processing
Radboudumc	Wouter Hoefsloot	Data collection, data processing
Radboudumc	Arthur Lemson	Data collection, data processing
Radboudumc	Cynthia van Arkel	Data collection, data processing
Radboudumc	Lisa Kurver	Data collection, data processing data manager, key file administrator
Radboudumc	Chang van Lier	Data collection, data processing
Radboudumc	Reinout van Crevel	Data collection (clinical, treating physician)
Radboudumc	Neeltje Carpaij	Data collection (clinical, treating physician)
Radboudumc	Cecile Magis-Escurra	Data collection (clinical, treating physician)
Radboudumc	Martin Boeree	Data collection (clinical, treating physician)
Radboudumc	Wouter Peeters	Data collection, data processing
Radboudumc	Lotte Vruggink	Data collection, data processing
Radboudumc	Cedric Bosteels	Data collection, data processing

2.5 For (WMO/nWMO) studies using human participant data, will (a form of) informed consent be given by the participants to collect and process their data? *Multiple options can be selected, specify for the applicable project(s) why informed consent is needed, or not.*

- Yes (please specify)

All participants will provide informed consent: Deelbiobank Radboudumc Mycobacterial Cohort Study

2.6 If yes, has (a form of) informed consent been given by the participants to share their data for further research, according to the FAIR principles?

- Yes

3. Collect and Create

3.1 Will existing data be used for this project?

- Yes (specify which data sources)

Care data from the electronic health record, EPIC
Microbiological data from GLIMS

3.2 Do restrictions apply to the use of these existing data? Describe how the use of these data will be arranged with the owners of the data.

no

3.3 Will human participant data be used from Radboudumc's clinical archives, like Epic, GLIMS, PACS , etc.?

- Yes, data will (also) be obtained from archives other than Epic (please specify which archive(s) other than Epic)

GLIMS

3.4 If Epic data is used, how will these data be obtained?

- Other methods will be used to obtain Epic data (please specify in 3.5)

3.5 If applicable, describe which other method will be used to obtain data from Epic, and how patient privacy is safeguarded:

Clinical data will be collected for all participants enrolled in the Radboudumc Mycobacterial Cohort Study using a Smartform in Epic.
Clinical data will be extracted from EPIC to an eCRF

3.6 How will privacy of the human participants be safeguarded?

- The data will be pseudonimized

3.7 How will pseudonimization/ anonymization/other be arranged?

- The data will be pseudonimized / anonymized manually (explain in next question)

3.8 Please specify how pseudonimization or anonimization will take place. How is the key code composed, will identifying elements be omitted, where (and/or in which system) will the key be saved?

Data will be pseudonimized manually using an unique alphanumeric code consisting of at least 6 characters.
The identification log will be saved in a password protected file on the Internal Medicine H-drive:
H:\Infectieziekten\Mycobacteriën

3.9 Provide the details from the Identification Log data in PIMS.

n/a

3.10 Describe the data that will be collected / created:
(optional: the data can be described per (sub)project, for instance workpackages or chapters)

(Sub)Project *(optional)*	**Volume, N=**	**Existing / new data**	**Data source**	**Data collection tool / system**	**Data Type**	**File Format**	**Storage space**
		Existing	EPIC	eCRF	Patient data	.CSV	1-10 GB
		Existing	GLIMS	eCRF	Patient data	.CSV	1-10 GB
		New	Lab system	Excel	experimental data	.xlsx	1-10 GB
		New	Lab system	EGA archive	sequencing data	.bam	> 1 TB

4. Store and analyze

4.1 Where will the research data be stored during the research phase (e.g. for combining, processing, and/or analyzing data)? Mention both digital and paper data storage.

- Data will be stored on a department server
- Data will be stored on another storage option

4.2 Give a short description of all storage options that will be used for the study data.

The key file will be stored in H:\Infectieziekten\Mycobacteriën\MyCoS\Investigator Site File\E. Informatie Proefpersonen\E7. Identifcation log

Research data - including clinical patient and laboratory data - will be stored in H:\Exp Int Gnk\MyCoS.

4.3 How will data security be ensured during the project?

Data stored in an eCRF from a collaborating party (such as CASTOR ED) will be protected by controlled acces by applying user management.

4.4 How often, where and by whom will backups be made of the data?

Radboudumc network servers are used for data storage.

4.5 How is access to the data arranged for all parties, internal and external (if applicable); which restrictions will be applied to the access to the data during research?

- The data manager and members of the research team that have a care relationship with the participant have acces to the non-pseudonymized patiënt data in the EHR. The research team only has access to the pseudonymized data in Castor EDC.
- The Subject Identification Log wil only be accessible to the data manager and the members of the research team that have a care relationship with the participant.
- Access to the project folder on the H:share will be restricted to members of the research team, only.

4.6 Which software or tools will be needed to process or analyze the data?

- Microsoft Excel
- R / R studio
- SPSS
- Graphpad Prism

4.7 Will standard facilities, (like Zero Clients "werkplek 2.0", Fat clients provided by Radboudumc ICT or DRE), be sufficient to process and analyze the data or will extra computing power and memory be required ?

- The Radboudumc virtual 'werkplek 2.0' will suffice for processing and analyzing the data

4.8 What will be the estimated costs for managing the data during the study, and how will these be covered?

4.9 How will you structure your data? Briefly elaborate on the naming conventions and the structure of the files and directories.

Research data - including clinical patient and laboratory data - will be stored in H:\Exp Int Gnk\MyCoS.
Subfolders will be created for the study documents, and data.
Documents will be dated using the yymmdd_DOCUMENT NAME format. Every new version will be saved with the date and version number and older documents will be transferred to an archive folder within the subfolder.

4.10 How will version control be applied, with clear version numbers, to maintain all changes that will be made to the data?

Documents will be dated using the yymmdd_DOCUMENT NAME format. Every new version will be saved with the date and version number.

- **version number system:** Any major changes to a file can be indicated by whole numbers, for example, v01 would be the first version, v02 the second version. Minor changes can be indicated

by increasing the decimal figure for example, v01_01 indicates a minor change has been made to the first version, and v03_01 a minor change has been made to the third version.
- when draft documents are sent out for amendments, upon return they should carry additional **information to identify the individual who has made the amendments** . Example: a file with the name datav01_20130816_SJ indicates that a colleague (SJ) has made amendments to the first version on the 16th August 2013. The lead author would then add those amendments to version v01 and rename the file following the revision numbering system.

4.11 Which documentation will be added to the data to (further) describe the data collection?

- I will make use of a codebook, that describes all data items in the data collection
- I will document the research process (data cleanings, methodology of data collection, quality controls, statistics)
- I will include syntaxes that will be used to process and/or analyze the data
- I will include software (versions) that will be used to process and/or analyze the data
- I will link the data to one or more scientific publications

5. Archive and share

5.1 Will there be any issues that affect the sharing of (parts of) the data collection after the research? If so, briefly describe these issues.

- Yes, there are issues that affect the sharing of (parts of) the data after the publication of the results. (please specify)

Data were colllected with informed consent from participants. The conditions stated in the informed consent form should be met in case data are reused by others. Moreover, privacy issues prevent the data from being shared freely. We will tackle these issues by including in the consent form that data will be made available online only for reuse for research on the topics of infection and the immune system. We thereby included that the future researchers will only be allowed to access the data after agreeing to the terms of use, which include responsible use of the data.

5.2 If applicable, where will the parts of the data collection that are not suitable for data reuse be stored for the long term?

The patient identification keyfile and other documentation containing directly identifiable data (informed consent files) will be archived within in the department as long as required by regulations.

5.3 Which (part of the) data will be made findable and shared for reuse and/or verification?
(See also the table from question 3.10 that describes the data collection)

Sequencing data

5.4 How will the data be made findable and shared for reuse and/or verification? Select the options that apply, and provide further details the comments field.

- Data will be published in a data repository or other online data archive (e.g. DANS EASY, Radboud Data Repository, disciplinary repository, data archive) (please specifiy)

We foresee to publish the data in the Radboud Data Repository

5.5 Will restrictions be applied to access to (parts of) the data?

- Yes, restrictions will be applied to (parts) of the data (please specify, e.g. how restrictions wi ll be applied, how will access to the data be arranged, who will be made responsible for granting access to the data)

The pseudonimized data will be published under restricted access. Requests for acces will be checked by a data access committee (DAC), against the conditions for sharing the data as described in the signed Informed Consent.

5.6 Will a license be applied to the published data? If yes, what license?

CC-BY

5.7 Will terms of use be set out to describe the conditions for access to the data? If yes, how are these defined, e.g. in a consortium agreement, data use agreement and/or other terms of use?

Yes, in a data use agreement.

5.8 How will metadata be published to describe the data collection, and to enable findability of the data collection? And which metadata will be shared, e.g. will standards be used?

- Dublin Core and DataCite Metadata about the data collection will be registered in RIS
- Metadata about the data collection will be published via the data repository/repositories (see above, please specify which metadata standard(s))

5.9 Will the dataset be made findable by means of linkage to a persistent identifier (PID) like a DOI, a Handle or other PID? Please provide the PID here as soon as it is available!

The DOI (as well as other metadata) will be automatically registered in the Research Information

Services (RIS) and will be shared in the respective publications.

5.10 Which documentation will be added to the published data collection to enable reuse? (see also 4.11 for examples of data documentation)

refer to 4.11

5.11 What will be the estimated costs for data archiving and/or publication after the study, and how will these be covered?

The costs for the Radboud Data Repository are expected to be minimal.

5.12 Which (bio)medical or other discipline specific terminology (vocabularies, classifications, ontologies or other standards) will be used whithin the data collection?

NCBI sequence identifiers and LOINC for lab data from GLIMS.

5.13 Which data formats will the data collection contain? See also 3.10, is there a need to migrate the data to (a) preferred data format(s)?

refer to 3.10

5.14 Which part of the data collection will be stored for the long term and for how long will the data be stored?

Scientific research involving humans, a storage period of at least **15 years** applies in the Radboudumc.

5.15 If applicable, describe your strategy for publishing the analysis software that will be generated in this project.

NA