**DMP TITLE**

RESEARCH PROJECT TO IDENTIFY EARLY MECHANISMS IN PULMONARY FIBROSIS WITH FOCUS ON AIRWAYS AND VASCULAR STRUCTURES

ACRONYM: REPAIR

**ADMIN DETAILS**

**Project Name:** Research project to identify Early mechanisms in Pulmonary fibrosis with focus on AIrways and vasculaR structures
Acronym: REPAIR
DMP title: DMP FWO 1832522N FKM Wim A Wuyts
**Project Identifier:** W A Wuyts FKM
**Grant Title:** 1832522N
**Principal Investigator / Researcher:** Prof. Wim A Wuyts

**Description**: The DMP lists all research activities of the involved PI's or researchers in respect to the tasks specified within the project. DMP activities are described in respect to task owners and work packages in the workplan. The overall responsibility for the DMP lies in the hand of the grant holder (Wim Wuyts), while the involved PI's and researchers take the responsibility to ensure the correct execution of the DMP for their assigned tasks and research activities.

**Institution:** KU Leuven

**1. GENERAL INFORMATION**

**Name applicant**
Wim A Wuyts

**FWO Project Number & Title**
1832522N
Research project to identify Early mechanisms in Pulmonary fibrosis with focus on AIrways and vasculaR structures
Acronym: REPAIR

**Affiliation**

- KU Leuven

## 2. DATA DESCRIPTION

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

I will generate new data and make use of existing data obtained from databases generated by myself as part of my former postdoctoral projects.

**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).**

The project is divided in 4 work packages (WP1-4). As the nature of the data might differ from one WP to the other the information is provided per work package.

*WP1: Identification of patients with subclinical stages of pulmonary fibrosis, identification of risk factors for the development of progressive pulmonary fibrosis.*

This WP will focus on two specific groups of patients that are at high risk on developing pulmonary fibrosis, namely family members of patients with familial fibrosis and patients with newly diagnosed rheumatoid arthritis.

This WP consists of 2 parts:

WP1.1 screening of relatives of patients with familial fibrosis and WP1.2 screening of rheumatoid arthritis patients

*WP1.1: screening of relatives of patients with familial fibrosis*

**Type of study**

This WP consists of 2 parts:

Part A: the identification and characterization of familial IPF patients and Part B multimodal screening program in their relatives consisting of a questionnaire, HRCT, pulmonary function test, genetic testing and telomere length assessment.

**Types and format of scale of data**

In part A Clinical data (e.g. clinical characteristics) will be extracted from the electronic medical records in KWS (Klinisch Werk Station). Extracted analyzed data will be converted to an electronic case report form (eCRF) in REdCAP (Research Electronic Data Capture). Three blood samples will be collected:

1) one blood sample for telomere length assessment will be collected and will be sent to UCL (Université catholique de Louvain)

2) one blood sample for genetic testing (DNA analysis) will be collected and will be sent to Centre of Human Genetics (Leuven)

3) one blood sample will be collected and will be stored in our biobank. Results of genetic testing will be extracted from KWS and will be converted to an eCRF in REdCAP.

Results of telomere length assessment will be sent to us by UCL through LiquidFiles. These files (.pdf +/- 50MB) will be uploaded in REdCAP. Data will be extracted from these files and will be converted to a eCRF in REdCAP.

In part B clinical data (e.g. HRCT, pulmonary function tests variables) will be extracted from the electronic medical records in KWS. Extracted analyzed data will be converted to an eCRF in REdCAP. Questionnaires (.pdf (+/-500MB)) will be collected and will be uploaded in REdCAP. Data will be extracted from these questionnaires and will be converted to an eCRF in REdCAP. Three blood samples will be collected:

1) one blood sample for telomere length assessment will be collected and will be send to UCL (Université catholique de Louvain)
2) one blood sample for genetic testing (DNA analysis) will be collected and will be send to Centre of Human Genetics (Leuven)
3) one blood sample will be collected en will be stored in our biobank.

Results of genetic testing will be extracted from KWS and will be converted to an eCRF in REdCAP. Results of telomere length assessment will be sent to us by UCL through LiquidFiles. These files (.pdf +/- 50MB) will be uploaded in REdCAP. Data will be extracted from these files and will be converted to a eCRF in REdCAP. R-studio software will be used for statistical analysis. Data will be presented in Word and PowerPoint files (+/-100MB).

| **Type of data** | **Format** | **Volume** | **How created** |
|---|---|---|---|
| Clinical data (e.g. clinical characteristics, HRCT report, pulmonary function tests) | REDCAP | REDCAP server | Clinical data (e.g. clinical characteristics, HRCT and pulmonary function test) will be extracted from the electronic medical records in KWS (Klinisch Werk Station). Extracted analyzed data will be converted to an eCRF in REdCAP (Research Electronic Data Capture). |
| Questionnaires | pdf | 500MB | Questionnaires (.pdf (+/-500MB)) will be collected and will be uploaded in REdCAP. Data will be extracted from these questionnaires and will be converted to a eCRF in REdCAP. |
| Results of telomere length assessment | Pdf | 50Mb | Blood samples for telomere length assessment will be collected and will be sent to UCL (Université Cc~~c~~atholique de Louvain). Results of telomere length assessment will be sent to us by UCL through LiquidFiles. These files (.pdf +/- 50MB) will be uploaded in REdCAP. Data will be extracted from these files and will be converted to a eCRF in REdCAP. |
| genetic sequencing | REDCAP | REDCAP server | Blood samples for genetic testing (DNA analysis) will be collected and will be sent to |

| | | | |
|---|---|---|---|
| | | | Centre of Human Genetics (Leuven). Results will be extracted from the electronic medical records in KWS. |
| Presentations/papers | Word/ppt | 100MB | Data will be presented in Word and PowerPoint files (+/-100MB). |

*WP1.2: Screening of RA patients*

The project consists of the identification and characterization of patients with interstitial lung disease associated with rheumatoid arthritis (RA-ILD), as well as the development of a multimodal screening program for patients with rheumatoid arthritis, consisting of clinical characteristics, HRCT, lung ultrasound (LUS), pulmonary function testing, proteomics and genetic testing.

**Types and format of scale of data**

Clinical data (e.g. clinical characteristics, HRCT, LUS, pulmonary function tests variables) will be extracted from the electronic medical records in KWS. Extracted analyzed data will be converted to an eCRF in REdCAP. Two blood samples will be collected: One blood sample for genetic testing (DNA analysis) and one blood sample for proteomics analysis. Both will be stored in the Breathe biobank (S63978). Genetic testing will take place through the Genomics Core. Results of genetic testing will be extracted from KWS and will be converted to an eCRF in REdCAP. R-studio software will be used for statistical analysis. Data will be presented in Word and PowerPoint files (+/-100MB).

Existing data from the PROOF, prospective RA registry (ReCoRD) and CareRA registries will also be extracted and converted to a eCRF.

| Type of data | Format | Volume | How created |
|---|---|---|---|
| Clinical characteristics | REDCAP | REDCAP server | Clinical data, lung ultrasound protocols, results from genetic and biomarker testing, high resolution CT protocols and pulmonary function test results will be extracted from the electronic medical records in KWS (Klinisch Werk Station). Extracted analyzed data will be converted to an eCRF in REdCAP (Research Electronic Data Capture). |
| HRCT reports | | | |
| Lung Ultrasound | | | |
| Pulmonary function testing | | | |
| Proteomics | | | |
| Genetic sequencing | | | |
| Data presentation | .docx, .pptx | 100 MB | Data will be presented in Word and Powerpoint files. |

*WP2: structural changes over time with specific focus on early signs of fibrosis (ILA).* The main aim is to map the structural changes in lung parenchyma, airways and vascular structures over time from the early (subclinical) phase towards end stage fibrosis using state of the art technology.
In WP2.1 we will rely on the existing biobank (S63978) consisting of data and tissue from patients with different forms of ILD diagnosed and followed-up in UZ Leuven that underwent lung transplantation. This will be based on the data collected in WP1. In WP2.2 serial HRCT scans from patients with pulmonary fibrosis will be further analysed using CALIPER software, an artificial intelligence software developed at the Mayo clinic in US and UCL in London. In WP2.3 we will apply micro CT scanning on the explant material from WP2.1. The project will focus on parenchymal changes, airway changes and lung vasculature. For this project the vast majority of tissue has already been collected and stored in the Breathe biobank.

| Type of data | Format | Volume | How created | Analysis software |
|---|---|---|---|---|
| CT scans of lungs and explants | .DICOM | 1GB/scan 150GB | HRCT scanning via clinical CT scanners | RadiAnt, ITK-SNAP, NeuronStudio, ImageJ |
| MicroCT Scans of whole lungs | Serial .tiff or .jpeg | 8GB/scan 630GB | Custom microCT set-up Gent University Centre for X-ray tomography | |
| MicroCT scans of small cores of lung tissue | Serial .tiff or .mha | 5GB/core 1.6TB | Bruker Skyscan | |
| Clinical data | .xlsx | 10MB | Data from already existing database | |

*WP3 The guided search for disease-relevant cell clusters and subsequently disease related pathways with a focus on reversibility of the disease.*

In this WP we will use single nuclear (sn)RNA seq and spatial transcriptomics on the same lung tissue specimens. We will use all samples characterized in WP2. Once we identified the morpho-molecular pathophysiological traits prevailing in early fibrotic regions, we will seek to identify treatment targets by determining transcriptomic regulatory networks and analysing the connectome landscape of these molecular determinants.

| **Type of data** | **Format** | **Total Volume** | **How created/analyzed** |
|---|---|---|---|
| microscopy | .nd2 files (Nikon software); .tiff | 200GB | light microscopy of explant tissue sections |
| snRNA seq | .bam (Sequencing) .fastq, .csv, (CellRanger, R-Studio);.xlsx, .txt (databases) | 40GB | Visium 10x Genomics pipeline |
| spatial transcriptomics | .fastq (Raw sequencing files) .txt | 5GB | Nanostring GeoMx pipeline |

*WP4: Lung on chip dynamic model*

The treatment of Progressive Fibrosing Interstitial Lung Disease (PF-ILD) is hampered by the lack of knowledge of their mechanistic driving factors. Current experimental models to study this disease rely on in vitro (e.g. 'simple' cell cultures) and in vivo (e.g. mice) experimental models, which are poorly representative of the human pathophysiology. Therefore, more relevant 2D/3D patient-derived cell cultures are gaining momentum as more reliable disease models compatible with personalized medicine. In collaboration with Prof. M. Carlon and Prof. X. Casadevall, we will develop a microphysiological organ on chip model of chronic lung disease. Besides recapitulating the complex cellular and environmental conditions of the pulmonary regions of interest for this disease, this system will ultimately be fully personalizable, forming the basis for the development of novel therapeutic strategies to prevent its early disease onset.

| **Type of data** | **Format** | **Volume** | **How created** |
|---|---|---|---|
| Microscopy | .czi .tiff | 3TB | Confocal microscopy (Zeiss LSM880) of cell composition and dynamic biological processes |
| Sanger sequencing | .abi | 250GB | Genetically engineered cells by CRISPR or vector technology |
| RT-PCR | .csv, .xlsx | 2GB | CFX96 RT-PCR with CFX Maestro software for gene expression quantification |
| FACS | .fcs, .txt | 500MB | Guava/BD FACS for analysis/ sorting of lung cells |
| ELISA | .xlsx | 500MB | Envision optical density (O.D.) plate reader for inflammatory/ pro-antifibrotic cytokines/ growth factors |
| Sc/snRNAseq | .bam (Sequencing) .fastq, .csv, | 20GB | Visium 10x Genomics pipeline |

| | | | |
|---|---|---|---|
| | (CellRanger, R-Studio);.xlsx, .txt (databases) | | |

### 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.**

- Yes

Privacy Registry Reference:
WP1.1: Ethical committee approval: S63694

Existing PROOF registry Ethical committee approval: S60645

WP1.2: protocol in preparation and will soon be submitted for ethical committee approval

Approval of the existing prospective RA registry (ReCoRD): S63021

Approval of the existing RA registry Carrera: CareRA S51411 and CareRA2020 S59474

WP2:

All ethical and biosafety issues related to the human lung collection and analysis have been approved by the university hospital for many years already via our biobank S51577

There is an ethical approval for our imaging program S52174

An ethical approval for collecting lobectomy specimen from patients with mild disease has also been obtained through S63093.

Additionally, the snRNAseq and spatial transcriptomics research is covered by ethical approval S64458.

There is an EC approval for the collaboration with UCL and J Jacob: S63737

WP3

Running ethical approval for snRNAseq and spatial transcriptomics: S64458

WP4 EC approval is in place to biobank lung explant derived primary cells (epithelium, endothelium, fibroblasts) for use in cell culture experiments, including organ on chip (old biobank numbers S51577, S55877, now included in the recent biobank ethical approval S63978).

Short description of the kind of personal data that will be used:
*WP1.1 screening of relatives of patients with familial fibrosis*

*Part 1. (patients)*
Personal data collected are demographic information (age, gender, birth date), results of genetic testing and telomere length assessment, family history, specific information on the lung condition (pulmonary function, radiological evaluation, pathology of the lungs, differential cell counts of bronchoalveolar lavage)
and smoking status, occupational/environmental exposure, medication use, medical history, time to death/transplantation.
*Part 2. (relatives)*

Personal data collected are demographic information (age, gender, birth date), results of genetic testing and telomere length assessment, results of HRCT and pulmonary function tests, occupational/ environmental exposure, medication use, medical history and family history

WP 1.2 Screening of RA patients
Following data will be collected:

- *Retrospective data (sex, age, medication, date of diagnosis of rheumatoid arthritis, duration of disease, comorbidities, medical history, familial history, characteristics upon clinical examination, …)*
- *Imaging: HRCT images and protocols, LUS images and protocols*
- *Pulmonary function testing data*
- *Physical data:*
    - *Blood samples*
    - *BAL samples if available*

WP2 and 3 structural changes over time and the search for disease-relevant cell clusters and pathways.

Personal data collected are general demographic information (age, gender, length, body weight, home address...) but also specific medical information on the lung condition (lung function, radiological evaluation, pathology of the lung) and general health status and therapy (CRP, smoking behavior, steroid use, immunosuppression… ).
We will be using intact human lungs collected within our biobank S51577 for studying airway morphometry, cellular composition, aerodynamics, and specific in vitro experiments on the airway epithelial cells.
Gene expression data will be collected as well as spatially resolved gene expression data.

WP4: Lung on chip dynamic model
See above.

**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

WP1.1: Ethical committee approval: S63694
WP1.2: In preparation for ethical committee approval
WP2-4: deal with explant human lung material collected at the time of transplantation:
Running ethical approval to collect the human lungs:  biobank S51577.
Running ethical approval for microCT imaging: S52174.
Running ethical approval for sncRNAseq and spatial transcriptomics: S64458

Running MTA approved via UZLeuven for WP 2 and 3 for the collaboration with Jim Hogg (Vancouver, Canada) and Joe Jacob (London, UK)
MTAs currently being installed to collaborate with Naftali Kaminski (Yale, USA).
MTA is currently installed with UZA (Stijn Verleden)

**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?**

WP 1.
A code book/data dictionary is integrated in REdCAP. The study design/protocol, informed consent forms, questionnaire and information leaflet that have been approved by the ethical committee will be stored in a .pdf file.

WP2,3.
List of biobanked lung explants with their respective code is available from the Breathe lab manager (K. Maes) and involved Breathe lab PI's which enables the coupling of the patient data to the lung explant of interest.

| Data type | Metadata description |
|---|---|
| microscopy | Images will be clearly labeled to identify experiment, sample, and date. Metadata concerning the technical specifications of the images (e.g. laser/gain/dimensions/etc) are collected automatically with the .czi files |
| CT scanning | Images will be clearly labeled to identify pseudonym and date. Metadata concerning the technical specifications of the images are collected with the .dicom files |
| microCT scanning | Images will be clearly labeled to identify pseudonym, core ID and date. Metadata concerning the technical specifications of the images will be noted down by each experiment. |
| Sequencing | All files will be clearly labeled, to include experiment number and conditions. |

WP4

| Type of data | Metadata description |
|---|---|
| Microscopy | Images will be clearly labeled to identify experiment, sample, and date. Metadata concerning the technical specifications of the images (e.g. laser/gain/dimensions/etc) are collected with the .czi files, or else will be noted down by each experiment. |
| Sanger sequencing | All files will be clearly labeled, to include experiment number and conditions. |
| FACS | Experiments will be clearly labeled, to include cell model used, date, staining and conditions. .fcs files prior and after analysis are stored, containing metadata as laser settings, cell count, etc. |
| ELISA | All files will be clearly labeled, to include experiments, date, etc. Final measurements are conducted via a standardized protocol which will be stored with the data. All deviations from this protocol will be noted clearly. |
| RT-PCR | |
| Sc/snRNAseq | |

WP2-4:

Standard experimental procedures (SOPs) and practices will be fully documented as word (and PDF) and saved on the two-factor authentication protected KU Leuven GBW-0076_LTx server.

SOPs include:

Raw experimental data from the analytical SOPs are collected per experimental test, and will include a MS Word file with a clear description of what the data represent and how they were generated. This description will be documented in notebooks (with page numbers), as well as in electronic format (OneNote).

The name of the folder always contains the date, name of the experiment, and the name of the person who performed the experiment.

Each individual file with experimental data contains information on the study design, the origin of the samples, and all necessary information for an independent analyst to use or reuse the data accurately and efficiently.

**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.**

Currently the use of metadata standards has not been implemented in the daily routine of the research group. In case this will change in the course of the project this change and its timing will be reported at the end of the project. For all data common metadata are collected: (1) title, (2) author, (3) data type, (4) date created and date modified, (5) file size, (6) equipment reference (such as manufacturer and model identification). Depending on the nature of data additional metadata are collected. For instance, for *Microscopy:* (7) lens type, pin holes, gain, laser strength and magnification; or for *FACS:* (7) channels used. *Fluorescence:* (8) wavelength. *RNA sequencing*: protocol, sequencing depth, analysis cut-offs.

## 5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

**Where will the data be stored?**

WP 1

Part 1/ Data will be mainly stored in REDCAP, a web-based application provided by the University Hospitals Leuven and the KULeuven. REDCAP data are stored on the KULeuven REDCAP servers.

Part 2/ Other datafiles will be stored at the KU Leuven's secure environment, of which daily backup are made by the ICT.

Our dataservers are:

GBW-0076_LTx is a smaller (0.5Tb) server but faster server.
GBW-0017_LTx is a larger (15Tb) server but slower on which we store all raw data long time.
Access to these files is controlled by the lab manager (K. Maes) and PIs from the Breathe lab, linked to this project.

WP2-4

The physical data will be stored in appropriate storage rooms in the Breathe laboratory (ON1b) such as histology rooms, fridges, freezers (-20, -80, -150°C) and cryotanks, including the centrally managed KU/UZ Leuven Biobank (https://www.uzleuven.be/en/uz-kuleuven-biobank; contact: Lieve Bruggeman).
Hard copies of paper lab notebooks are kept in locked cabinets in the lab of the PIs concerned.

**How is backup of the data provided?**

The data that will be stored in REdCAP and on KU Leuven central servers with automatic daily back-up procedures. REdCAP data are saved on the KUL/UZ Leuven REDCAP 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 and this is still expandable:

REDCAP: extra storage space is purchased per project as needed (80€/year/project)

Our KU Leuven dataservers are:
GBW-0076_LTx is a smaller (0.5TB)
GBW-0017_LTx is a larger (15TB)
Each researcher has a private OneDrive account with storage capacity of 2TB per person and provides periodic backups.

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

WP 1:
The storage capacity needed for this project will be minimal (max. 10GB) and therefore the additional costs to store the data on the servers will be minimal (5 euro/year). For REDCAP, the annual costs are 80 euro for each project.
WP2-4:

For GBW-0076_LTx the annual cost is 519 euro/TB. We currently use for other projects +/-400 GB of the 500 GB reserved, resulting in an annual cost of +/- 200 euro/year currently. For GBW-0017_LTx the annual costs are 156,6 euro/TB.
Both storage servers are shared by the different PI's associated to the Breathe lab. These costs are financed through grant applications.

The OneDrive (2TB) comes without charge.

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

**WP1**
The data will be stored in REDCAP which has a two-factor authentication process, ensuring maximal security.

WP2-4
GBW-0076_LTx and GBW-0017_LTx KU Leuven servers likewise are protected by a two-factor authentication process.
Access to these servers is restricted to the Breathe lab researchers. The lab manager (Karen Maes) and the PI's of the projects control who has access to the servers.

## 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 will be stored for a minimum of 25 years (requirement Ethical committee).
Big datasets that impact on the need of server space are datasets of microCT scan, digital slides scans, RNA profiling datasets. Datasets of clinical data, analyzed data in excel or Graphpad prism files do not impact on server space.
Raw digital data of scans, slide picture and molecular profiles are stored on the large volume server. Analyzed subdatasets and clinical and end-analytical excel or Graphpad prism data stored at the smaller server by the individual researchers are always evaluated at the end of the PhD. The excess of data is removed and valuable data remain at the large volume server under the name the fellow in a folder 'old fellows". Currently this concept is permanent storage but this can be reconsidered in the future.

**Where will the data be archived (= stored for the longer term)?**
WP1: REDCAP

WP2-4: the large volume server (GBW-0076_LTx)

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

WP1 The storage capacity needed for this project will be minimal and therefore the additional costs to store our data on these servers will be minimal (+/- 50 euro/25years). The costs for REDCAP will be 2000 euro/25 years.

WP2-4

Considering the current yearly cost, we expect costs for data preservation to be about 12500 euro/year. The department CHROMETA reserves for each separate group per year a budget which is enough to cover these annual (and total) costs of basic storage.

## 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)?**

For data sharing, MTA's are in place (J Hogg, J Jacob) and in preparation with N. Kaminski from Yale New Haven. At the collaborating centers a DMP is also installed compliant with this DMP.

No 3rd parties are involved

Data sharing per group includes:

| | |
|---|---|
| N Kaminski | Physical data are lung cells and RNA |
| | Digital data are morphological data and RNA profiles |
| J Jacob | Digital data: HRCT scans |

In general, with respect to the data produced/collected no restriction for data sharing apply.

Considering the use of patient materials, data will be either shared in a coded manner through publications, or, if more specific information is required, following ethical approval.

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

For the vast majority of the project, raw data will not be made open access available.

The only expected open access will be for RNA profiling (anonymized) at the time of publishing in international peer reviewed as this is an international requirement.

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

- In an Open Access repository

The full (anonymized) dataset will be uploaded in a cvs format in Zenodo

Data will be made publicly available post publication depending on the journals policy (post-publication data repository). Non-published data will remain confidential until a final decision on publication of the data has been taken. RNA-seq data will be deposited in the Zenodo or Gene expression omnibus (GEO) database after publication in a peer-reviewed journal. The anonymized data will be made available upon publication of the results. Other data will be available upon request.

**When will the data be made available?**

- Upon publication of the research results

RNA sequencing data will be deposited in the Zenodo database and will be available after publication in a peer-reviewed journal. Other data will be available after signing a data sharing agreement which will be established with the support of KUL R&D after a request. Data will be available after signing a data sharing agreement which will be established with the support of KUL R&D after a request. Once KU Leuven has established a university managed and owned data repository sharing of data (or a subset of data) on this repository will be evaluated depending on the policy and conditions of this repository. Changes to the data sharing policy will be reported in the final DMP of this grant.

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

Only RNA datasets will be made accessible upon acceptance of the manuscript and this for everybody without restrains.

Due to the potential commercial value of patient data no general and full open access to data will be provided by default. Data which will be shared with third parties will exclude commercial use and will require appropriate credit to the data owners.

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

No costs are expected.

The KU Leuven repository will not request any cost contribution for KU Leuven researchers. Data shared through journal repositories will be covered by publication costs. Bilateral agreements for data sharing will be established through the services of KU Leuven R&D. The costs expected for data sharing are thus low and will be reported in the final DMP at the end of the project. They will be covered through funds of the project. For primary patient cell or tissue sharing if legally approved, the KUL biobank where the samples will be registered will calculate the cost for sample handling and shipment of the cell/tissue cultures and also for administrative work.

The upload on the Zenodo database by the Genomics Core includes a charge of 500 €. This charge will be covered by the project. Deposition to the Gene Expression Omnibus database is free of charge.

## 8. RESPONSIBILITIES

**Who will be responsible for data documentation & metadata?**

Data documentation and metadata will be organized by the PIs and fellows organizing the laboratory and project namely Karen Maes (lab manager) and Celine Aelbrecht (lab technician).

The individual researcher producing data will have the final responsibility for data documentation and metadata. In case of PhD students and technical personnel, the collection will be supervised by the scientific project responsible (involved PI, post-doc,..). However, the final responsibility of data integrity lies with the researcher performing the experiments.

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

Both servers are dedicated to the PI of the project and access is managed by the PI and the lab manager. ICT (Gert Goos as contact person and PI) is handling back-up and if needed expansion of storage capacity. Data storage and back-up underlies the responsibility of the individual researcher, who is supervised by the scientific coordinator. The responsibility for maintaining the infrastructure access for data storage lies in the hands of the IT responsible of the research team. Finally, the maintenance of servers and integrity of data stored on these servers underlies the ITC services of the university. The maintenance of the integrity of the external drives and the data stored on them will be responsibility of the lab manager.

**Who will be responsible for ensuring data preservation and reuse ?**
All persons mentioned before being the PI and Lab manager.

**Who bears the end responsibility for updating & implementing this DMP?**
The PI bears the end responsibility of updating & implementing this DMP.