DMP title

Project Name CURE - DMP title
Project Identifier IDN/21/013
Grant Title IDN/21/013
Principal Investigator / Researcher Emmanuel Vander Poorten
Project Data Contact emmanuel.vanderpoorten@kuleuven.be
Description robotiC gUided canceR nanomedicinE (CURE)
Institution KU Leuven

1. General Information

Name of the project lead (PI)
Emmanuel Vander Poorten

Internal Funds Project number & title
IDN-21-00195 - robotiC gUided canceR nanomedicinE (CURE)

2. Data description

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

- Generate new data

2.2. What data will you collect, generate or reuse? Describe the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a numbered list or table and per objective of the project.

Type of Data	Format	Volume	Method of Creation
Electron Microscopy Images	.tiff	10GB	1. Transmission Electron Microscopy (STEM) 2. High-Resolution Tunnel Electron Microscopy (HTEM) to analyze the size and shape of PLGA nanomedicine in **WP1, T1.1**
X-Ray Diffraction Images	.tiff	1 GB	1. X-RAY Diffraction (XRD) 2. Small Angle X-RAY Diffraction 3. Dynamic Light Scattering 4. Nano Material Tracking Analysis (Nano Sight) for characterization of nanomedicine in **WP1, T1.1**
Optical Imaging (IVIS Spectrum)	jpeg	50 GB	- To analyze parameters related to NM injection via the robotic platform (Dose, Number of injections, the speed of injection, the location, Distribution rates) in **WP1, T1.3** - Measurement of therapeutic efficacy of NM during in vivo studies in **WP1, T 1.4**
High-Pressure Liquid Chromatography (HPLC) Chromatograms	jpeg, gif, CSV	15GB	Estimation of the concentration of encapsulated drugs inside nanomedicine in **WP1, T1.1**

Image-Based Flow Cytometry (ImageStream Mark II)	.rif or FCS file	100 GB	- Study of hydrodynamic size, stability, charge, and number of NMs in **WP1, T1.1** - Measurement of therapeutic efficacy of NM during in vivo studies in **WP1, T 1.4**
Immunoblotting Data	jpeg, gif, CSV	5 GB	Estimation of emission of danger signals from the dying cells in **WP1, T1.1 via conditioned media analysis**
Antibody Conjugated magnetic Beads, Enzyme-linked immunosorbent assay (Elisa)	Charts, Raw data , CSVs	5 GB	Estimation of emission of danger signals from the dying cells in **WP1, T1.1**
User Studies (Meta data)	CSV, jpeg	15 GB	User studies involving the use of robotic EUS and ERCP by endoscopists and surgeons **WP2 , T2.3** ,
Multiplex Immunofluorescence	jpeg	50 GB	Measurement of therapeutic efficacy of NM during in vivo studies in **WP1, T 1.4**
MRI / CT Scans	DICOM	1 TB	- Scanning for getting 3D models of anatomical structures to help manufacturer create mockup in-cilico models **WP3, T3.2** - Measurement of therapeutic efficacy of NM in **WP3 , T3.3** **-** Verification of location of delivered NM relative to Tumor Cells **WP3, T3.4**
Data for Deep Learning	jpegs, mp4	500 GB	Data for deep learning and trained models to be used for **WP2, T2.3**

3. Ethical and legal issues

3.1. Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to the file in KU Leuven's Record of Processing Activities. Be aware that registering the fact that you process personal data is a legal obligation.

No

3.2. 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,

In WP3, clinical experiments will be prepared, performed, and supported by the clinicians from CDO and TCTR. Prior to the experiment, the related ethical approval will be obtained from the local Ethics Committee. Project partners at CDO and TCTR have ample experience in running and supporting clinical experiments. In themeantime,in-silico studies are planned from Month 1-18. By then, ethical approval will be prepared and obtained. Relevant ethical approval will be added upon receiving.

3.3. Does your research 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?

The specific design of the robotics catheter and NM formulations have the potential for valorization (WP1 and WP2), and this will be subject to discussion during the project. Underlying design methodologies, modeling techniques, building blocks, and process steps rely on the background. Valorization would likely translate into a development-on-demand project with an industrial partner, further re-designing and optimizing the hardware and medical techniques for a large scale use upon successful validation. Restrictions on design data may apply but normally not on the measurement and test data obtained with these designs and fabricated samples.

3.4. 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 regarding reuse and sharing are in place?

No.

4. Documentation and metadata

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

Software developed in this project will be documented using Doxygen (http://www.doxygen.nl/ and accompanied by build and launch scripts to facilitate (re)use. We will also use Docker (https://www.docker.com/) to capture a fully functional software environment that runs our software, including its dependencies. This allows anyone with a Docker installation to start up and use our software. We also describe the full procedure of how to install and run our software in the Docker domain-specific language (Dockerfile and/or Docker Compose). Each experiment will contain metadata answering in as much detail as possible what was measured, when, why, and how. Metadata will be stored in various ways, such as headers in the CSV files, annotations in a TXT file, or records in a database. Measurement protocols, that describe data collection during a surgical procedure, will be stored as PDF files and/or provided to stakeholders via a web portal. Representative measurement data on fabricated samples will be documented and made available to the other project partners

4.2. Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.

No, Most measurement data will be stored in CSV and DICOM and will be accompanied by metadata headers. We will ensure all metadata is described consistently during the project and well documented. This will facilitate the reuse of our datasets, both internally and externally. For any datasets we would wish to publish, we will investigate if the use of domain agnostic metadata standards, such as Dublin Core or DDI, is feasible within the project

5. Data storage and backup during the project

5.1. Where will the data be stored?

All developed software is version controlled and stored at our GitLab server. Data generated during the project will be stored at the data center of KU Leuven, on which the research unit.

5.2. How will the data be backed up?

All storage locations (KU Leuven GitLab, Data Center, one drive) have adequate backup strategies in place. In case of data loss or damage, we will be able to revert our software or data fairly easily. GitLab also allows tagging to annotate a particular version of our software. The research data manager of this project, Johan Philips (KU Leuven) will inform the researchers on how to safely upload the datasets to make sure that all of them are correctly backed up

5.3. 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, All storage locations (KU Leuven GitLab, Data Center, one drive) have sufficient capacity to store all our data and software. .

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

Data storage costs on GitLab and one drive are covered by the internal ICT contributions, while additional storage at KU Leuven's data center will be covered by the consumables foreseen in the FWO proposal. We estimate the total storage cost at 3500 EUR, which covers the estimated data volume of about 1.8 TB during the 4-year project and the following 5 year preservation period

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

The project will not use any personal data but, nonetheless, will ensure that all generated datasets and developed software are stored securely on KU Leuven's servers (either GitLab, one drive, or the Data Center) and behind proper authentication. During the project, we will evaluate how and when to release or publish software and datasets, making them available to the community. This step will also involve the curation of documentation and data.

6. Data preservation after the end of the project

6.1. Which data will be retained for the expected 10 year period after the end of the project? If only a selection of the data can/will be preserved, clearly state why this is the case (legal or contractual restrictions, physical preservation issues, ...).

The software will be preserved on our GitLab server, including build and launch scripts to ensure reuse. Public releases of our software might also be stored on GitHub to maximize dissemination. Measurement data that is discussed or used in publications will be preserved on either KU Leuven storage or community data repositories such as Zenodo. Raw footage of movies to support publications and presentations will be annotated and released on KU Leuven's video channel or our research group's YouTube channel. Project information and data are at least preserved as long as grant and collaboration agreements with the other partners demand.

6.2. Where will these data be archived (= stored for the long term)?

The software will remain on our Gitlab server both during the project as well as afterwards. Data and software will be stored on the university's central servers with automatic daily backup procedures.

6.3. What are the expected costs for data preservation during these 10 years? How will the costs be covered?

Since we estimate to preserve a maximum 1.8 TB, this amounts to about 3500 EUR for the whole 4 year period.

7. Data sharing and re-use

7.1. 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 or because of IP potential)?

No, 3rd party agreements or legal restrictions will prevent the sharing of software or data. If we, however, see potential exploitation of our developed software, we will consider the use of a commercial license. Specifications and interface specifications related to the catheter needed in the project can be shared. Underlying information on design methodologies, modeling techniques, building blocks, and process steps may contain proprietary information and shall not be disclosed; note that such information is not needed by the other project partners.

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

Curated datasets will be uploaded to a community data repository such as Zenodo under a CC-BY or MIT license. The software will be released with a permissive license unless exploitation potential prohibits this

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

- In an Open Access repository
- Other (specify):

Curated datasets will be published in an open access repository such as Zenodo and/or on the research group's website. The software will be released via GitLab and/or GitHub.

7.4. When will the data be made available?

- Upon publication of the research results

Datasets and software will be published and released as soon as we see fit and definitely upon

publication of the research results. If particular software might have the potential for valorization, we might consider a closed (commercial) license. Specifications, interfacing specifications, and measurement data shall be published.

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

Datasets will be published under a CC-BY or MIT license, allowing maximum reuse. The software will be released under a permissive license unless we see the potential for exploitation.

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

The costs for data sharing will be covered completely by the project's consumables budget. We estimate this cost at 3500 EUR. (same as the data storage cost, which will provide the feature for data sharing without any additional cost)

8. Responsibilities

8.1. Who will be responsible for the data documentation & metadata?

The researchers are responsible for documenting their code and datasets with proper metadata. This will be supervised and guided by Johan Philips, the departmental data manager.

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

The researchers are responsible for safely storing all software and data on the university's servers, such that backup strategies are automatically taking care of it. Johan Philips will guide and supervise the researchers in this task.

8.3. Who will be responsible for ensuring data preservation and sharing?

The researchers are responsible for ensuring data preservation and reuse and will be guided and supervised by Johan Philips.

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

The PI bears the end responsibility of updating & implementing this DMP and will be guided by the project data manager.