FWO DMP Template - Flemish Standard Data Management Plan

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	Sophie Heymans (0000-0002-8686-2247)
Contributor name(s) (+ ORCID) & roles	Supervisor: Koen Van Den Abeele (0000-0003-0327-6478) Co-supervisor: Jan D'hooge (0000-0002-2346-142X) Co-supervisor: Edmond Sterpin (0000-0001-9764-546X)
Project number[1] & title	12E0923N: Proton range verification enabled by acoustic modulation of superheated nanodroplets
Funder(s) GrantID[2]	12E0923N
Affiliation(s)	KU Leuven
Please provide a short project description	Proton therapy has grown exponentially over the past 20 years. The success of proton therapy relies on its ability to inflict maximal tumor damage while minimizing the exposure of organs at risk. However, range uncertainties currently prevent proton therapy to reach its full potential, and hamper its adoption for moving targets in the abdomen, which hosts particularly deadly cancers. Although several techniques for in vivo proton range verification have been developed to overcome those barriers, none has reached widespread clinical adoption. Recently, we showed that injectable nanodroplets can vaporize into microbubbles when exposed to proton radiation. The generated ultrasound contrast could be related to the proton range with sub-millimeter reproducibility, thereby opening the door for ultrasound-based range verification. In this proposal, we address important challenges for clinical translation. As the indirect nanodroplet radiation response at 37°C currently limits the performances, we propose to reach direct sensitivity to protons by lowering the vaporization threshold using a novel technique called acoustic modulation. Moreover, we will build advanced volumetric ultrasound imaging sequences to actively or passively localize vaporization events and infer the proton range, in parallel with acoustic modulation, using a single ultrasound array. Finally, we will conduct an experimental performance assessment in clinical settings and nearly physiological conditions.

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

2. 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.

				ONLY FOR DIGITAL DATA	*ONLY FOR DIGITAL DATA*	*ONLY FOR DIGITAL DATA*	*ONLY FOR PHYSICAL DATA*
Dataset Name	Description	New or Reused	Digital or Physical	Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
RF ultrasound recordings of experiments	Raw recordings (channel data) of acoustic modulation experiments, irradiation experiments. Active (pulse-echo) or passive (receive-only) ultrasound recordings.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☐ other: ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	/
Ultrasound beamformed images	Beamformed ultrasound volumetric images/movies of nanodroplet vaporization	New data	Digital	Experimental	.mat	< 1 TB	/
Nanodroplet vaporization maps	Vaporization maps extracted from US	New data	Digital	Compiled	.mat	< 1 GB	/

	beamformed images/movies (see above) or from passively acquired RF data						
Raw hydrophone recordings	Characterization of the acoustic pressure fields of acoustic modulation transducers and matrix arrays	New data	Digital	Experimental	.mat or .tdms	< 100 GB	/
Acoustic pressure fields	Processed resulting pressure fields	New data	Digital	Compiled	.mat	< 1 GB	/
Ultrasound synthetic data	Computer simulations of high frame rate imaging sequences	New data	Digital	Simulation data	.mat	< 1 TB	/
Codes for high frame rate imaging	Codes to generate delay files and corresponding delay files	New data	Digital	Software	.m & .law	< 1 GB	/
Dynamic Light Scattering recordings	Nanodroplet size distribution results	New data	Digital	Experimental	.pdf, .csv, .m, .mat	< 1 GB	/

NMR recordings	Nanodroplet concentration results	New data	Digital	Experimental	.mat	< 1 GB	/
3D models of experimental equipment	SolidEdge CAD designs of experimental parts for flow phantoms	New data	Digital	Models	.par, .asm	< 1 GB	/
Processed data from acoustic modulation experiments	Peak negative pressure thresholds for acoustic modulation in different conditions	New data	Digital	Compiled	.mat	< 1 GB	/
Processed data from irradiation experiments	Dose response curves, limit of detection, range verification performances, sensitivity range : full dosimeter calibration and performance assessment	New data	Digital	Compiled	.mat	< 1 GB	/
Ultrasound image acquisition, analysis and	Software used to process and analyse passive or active	New data	Digital	Software	.m	< 1 GB	/

processing software	ultrasound recordings						
Monte Carlo simulations	Monte Carlo simulations to generate vaporization response functions	New data	Digital	Software & simulation results	TBD	< 100 GB	/
Modelled distribution of protons and charged particles	MCSquare models of proton distribution in the target	New data	Digital	Software & simulation results	TBD	< 100 GB	/
Predicted vaporization response	Code and files containing the predicted vaporization response based on the beam model and MC simulation results	New data	Digital	Software & compiled data	.m & .mat	< 1 GB	/
Scanned laboratory notebooks, procedures, protocols	Electronic lab notebook, scans of paper notes, experimental protocols	New data	Digital	Observational	PDF, docx	< 1 GB	/
Paper manuscripts	Manuscripts of research papers	New data	Digital	Compiled/ aggregated data	docx	< 1 GB	/

	and technical reports						
Powerpoint presentations and posters	Presentations and posters for seminars, conferences, dissemination activities	New data	Digital	Compiled/ aggregated data	pptx, pdf	< 1 GB	/
Metadata	Brief description accompanying each dataset	New data	Digital	Compiled/ aggregated data	txt	< 100 MB	/

GUIDANCE:

DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.

EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[3] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.

EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.

DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.

PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).

[3] These data are generated by combining multiple existing datasets.

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, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:
Will you process personal data[4]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:

[4] See Glossary Flemish Standard Data Management Plan

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 ☐ No If yes, please comment: Imaging technologies to be developed in WP1 and WP2 are of potential interest for vendors of ultrasound systems, as they can improve contrast enhanced ultrasound imaging for a variety of applications. The predictive model for proton range verification (WP4) also has potential value should the range verification application be translated to the clinic. For this reason, for both situations, we aim to restrict access until its valorization potential has either been realized (e.g. patenting) or abandoned. Advice from KU Leuven LRD Tech Transfer office will be sought whenever valorization potential is identified.
Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)? If so, please explain to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. 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).	Every experimental dataset (nanodroplet characterization, ultrasound acquisitions) will be accompanied by a filled experimental protocol (.doc, txt or pdf) describing the content of the data and the experimental conditions corresponding to its acquisition. All experimental files will be stored in folders clearly labelled with the date of the experiment and name of the phantom/repeat, which are referenced in the experimental protocol. Whenever necessary, additional files containing supplementary information (axis, image dimensions, acquisition parameters) will be generated and stored in the same folder. Ultrasound simulation data will be accompanied by a .txt file with a clear description of what the data represents and how it has been generated. The input files related to the simulations will be stored in the same folder. The folder names will follow a clear structure defined in advance and maintained in a specific .txt file. Ultrasound imaging, acquisition and processing softwares will be managed in Gitlab for versioning and carefully documented (structured comments within the source code). Codes will be hierarchically organized, with a main code for execution accompanied by clearly labelled subfunctions. Each code entity (main + subfunctions) will be accompanied by a read-me file containing detailed information about the software dependencies and providing detailed installation and compilation instructions. Software related to the predictive vaporization model (e.g. MCSquare simulations, Geant4 simulations) will also be clearly documented with read-me files containing detailed explanation about the code structure. A tutorial document will be provided, describing the high-level design and organization of the software packages. Processed data from experiments and simulations will be classified according to the experiment date, accompanied by the experimental protocol and a read-me file containing information about the date of creation, code version used to generate the processed data, and experimental dataset to which it corresponds (if applicable). Manuscripts, protocols, presentations, reports, electronic lab books : will be stored in specific folders with clear names.

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. *REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.*	☐ Yes ☒ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: If no, please specify (where appropriate per dataset or data type) which metadata will be created: There is no standard metadata format available for most of our data. Wherever applicable, the Dublin Core Metadata will be used. The original DCMES (Dublin Core Metadata Element Set) Version 1.1 consists of 15 metadata elements: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, and Rights. Each file will thus contain information on study design, sampling methodology, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	Experimental data, simulation results, processed data, lab notes, manuscripts and presentations will be stored online using KU Leuven OneDrive for Business. In addition, the experimental data will be backed up on external solid state drives. Generated software (ultrasound image processing codes, predictive model for range verification, simulations) will be stored in Gitlab Enterprise repositories at KU Leuven, which are managed by the ICTS services of KU Leuven.

How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.*[5] *REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.*	Gitlab repositories at KU Leuven and OneDrive for Business are managed by the university ICT staffs and automatically backed-up. In addition, a back-up of the experimental data is foreseen on external hard drives.
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 ☐ No If yes, please specify concisely: The total required storage capacity was estimated to be < 5 TB. Upgrade of Data Storage to 5 TB was foreseen, as indicated in the data management plan The Gitlab repositories have a space of 1 GB per user (extendable), which is sufficient for the software that is expected to be generated within the project. If no, please specify:

[5] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE.* [5]	Data security is provided by the ICT departments of KU Leuven. They will maintain the servers on which the data is stored, backup the data regularly, and fight against any malware attacks. OneDrive for Business has a restricted access (user-specific, with multifactor authentication with the KU Leuven Authentificator app) and users who wish access to the data need to be approved by the owner of the specific folder(s). Any additional user who wish access to the restricted data will receive it only if approved by all project investigators (Dr Heymans, Prof. Van Den Abeele, Prof. D'hooge and Prof. Sterpin). Communication (data transfer) with the OneDrive for Business cloud storage is via SSL / TLS. All SSL connections with OneDrive for Business via the internet are done with 2048-bit keys. Data movements between data centers happen over a private network and are further encrypted.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	The cost for storage of data on the Kulak servers is free during the project (including back-up). Storage on OneDrive and Gitlab is expected to be free of cost due to the sufficient available space provided by KU Leuven (up to 5 TB).

5. Data Preservation after the end of the Research Project	
Which data will be retained for at least five 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...).	After the research, data will be stored on and shared (internally) via the KU Leuven RDR Repository, with automatic back-up procedures for at least 10 years after the end of the project, following the KU Leuven research data management policy.

Where will these data be archived (stored and curated for the long-term)?	The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Storage on KU Leuven Archiving disks (K drive): the software data, processed data, and all the data stored on Git Lab and OneDrive will be transferred to an archive (K drive) for long-term storage. In addition, a copy will be made on an external hard drive, which will be stored in the office of Prof. Van Den Abeele. The storage cost is 99.55 euros/TB/year. These costs will be covered by available research funds and general consumer funds of the team of Prof. K. Van Den Abeele.

6. 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. *NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS*	☐ Yes, in an Open Access repository ☒ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: A decision on valorizing the generated data will be made by the end of the project. If the valorization potential is abandoned, the data will be made available.
If access is restricted, please specify who will be able to access the data and under what conditions.	The data uploaded on the KU Leuven repository RDR will be available to third parties under restrictions as defined by members of the project (e.g. email request).
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 ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	Annotated data are uploaded to KULeuven repository RDR and/or as supplemental material to publications, if requested by the journal.

When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	● Immediately after the end of the project ● Upon publication of the research results Data shared via supplemental material is shared upon publication. Other data are shared after the valorization potential has been fully explored.
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [6]	Data from the project that can be shared (i.e., after the valorization potential has been abandoned) will be made available under a Creative Commons Attribution License (CC-BY-4.0). Software source code will be made available with an open source software licence (copyleft license, which ensures that the code always remains open, e.g. GNU GPL v3).
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☒ Yes ☐ No If yes: Data uploaded to KU Leuven RDR will get a DOI that can be used to cite the data.
What are the expected costs for data sharing? How will these costs be covered?	The costs for data sharing with journal platforms is covered by the publication fee. For data stored at RDR, the costs are covered by KU Leuven.

[6] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	During the project, Dr. Heymans will be responsible for managing the data documentation and metadata
Who will manage data storage and backup during the research project?	Dr Heymans
Who will manage data preservation and sharing?	Prof. Van Den Abeele will be responsible for data preservation and sharing
Who will update and implement this DMP?	Dr Heymans bears the overall responsibility for updating the DMP, while Prof. Van Den Abeele bears the responsibility to ensure that the DMP is correctly implemented.