Project title:	Medical Integrated Photonic Ultrasound Transducer
Project acronym:	Med-IPUT
Grant Agreement:	101100633
Call identifier:	HORIZON-CL4-2022-DIGITAL-EMERGING-01

D1.3: Data Management Plan

Lead partner:	EIBIR
Author(s):	Peter Gordebeke, Valentina Belmar
Delivery date:	30/06/2023
Dissemination level:	Public
Version:	1

Funded by
the European Union

Table of Contents

Abbreviations

CC	Creative Commons
FAIR	Findable, Accessible, Interoperable, and Reusable
IPUT	Integrated Photonic Ultrasound Transducer
MZI	Mach Zehnder Interferometer
OEM	Original Equipment Manufacturers
PA	Photoacoustic Imaging
PDK	Process Design Kit
PIC	Photonic Integrated Circuit
SiN	Silicon Nitride
SOI	Silicon on Insulator
SOP	Standard Procedure
SOTA	State of the Art

Disclaimer

The opinions stated in this report reflect the opinions of the author(s) and not the opinion of the European Commission.

1. Introduction

This Data Management Plan (DMP) outlines the strategies for managing the datasets generated within the context of the Med-IPUT project. This collaborative initiative focuses on developing innovative medical diagnostic technologies based on Integrated Photonics for Ultrasound Transducer (IPUT). Throughout the project, various datasets will be generated, comprising optical simulation results, mask designs, process instructions, in-line measurements during device processing, and optical measurements of fabricated components.

The purpose of this DMP is to ensure effective and secure handling of these datasets, handling accessibility, integrity, and long-term preservation. By adhering to robust data management practices, the Med-IPUT project aims to facilitate data sharing, replication of results, and adherence to legal and ethical requirements.

This document presents an initial overview of the data management strategies at the start of the project, including data collection, documentation, storage, sharing, and preservation. Throughout the project's lifecycle, this will be updated to reflect the latest status. The DMP serves as a reference point for all project participants, ensuring consistency and clarity in data management practices.

By implementing this DMP, the Med-IPUT project aims to maximize the value and impact of the generated datasets, enabling future research, validation, and exploitation of the innovative technologies developed. The data management plan will be updated on a six-monthly basis and the compliance will be continuously monitored.

2. Data summary

2.1 Purpose of the data collection/generation

The Med-IPUT project encompasses a range of activities aimed at developing groundbreaking medical Integrated Photonics for Ultrasound Transducer. As part of this endeavor, the project will generate datasets during the realization of its objectives. These datasets will play a fundamental role in advancing knowledge, validating research outcomes, and enabling future innovation in the field of medical diagnostics.

The purpose of data collection and generation in the Med-IPUT project is multi-faceted. It serves several key objectives:

1. Evaluation and Optimization: The generation of optical simulation results allows for the evaluation and optimization of the proposed IPUT technologies. Through detailed simulations, the project team can assess the performance, efficiency, and feasibility of different design concepts, enabling informed decision-making in the development process.

2. Design and Fabrication: The creation of mask designs serves as a critical component in the manufacturing of IPUT devices. These designs provide the blueprint for the fabrication process, guiding the production of intricate structures necessary for ultrasound transducer. The datasets generated through mask designs aid in accurate and reproducible manufacturing, ensuring the quality and consistency of the final devices.

3. Process Instructions: Process instructions encompass detailed guidelines and protocols for carrying out specific manufacturing steps during the device fabrication process. These instructions capture the essential parameters, procedures, and methodologies required to achieve desired outcomes. The datasets associated with process instructions enable consistency, standardization, and reproducibility in device manufacturing, promoting reliable results and facilitating scalability.

4. Quality Control and Assurance: In-line measurements performed during the device processing phase provide critical feedback on the manufacturing process. These measurements capture real-time data on various parameters, allowing for continuous monitoring, adjustment, and optimization of the fabrication steps. The datasets generated through in-line measurements support quality control efforts, ensuring that the manufactured devices meet predetermined specifications and performance criteria.

5. Performance Evaluation: The results of optical measurements conducted on the fabricated components serve as a comprehensive assessment of the device's performance characteristics. These measurements provide quantitative data on key parameters such as sensitivity, resolution, and signal-to-noise ratio, which are essential for evaluating the device's diagnostic capabilities. The datasets derived from optical measurements enable the validation of the developed IPUT technologies, providing evidence of their effectiveness and potential clinical application.

By systematically collecting and generating these datasets, the Med-IPUT project aims to build a robust foundation for advancing medical diagnostic technologies. These datasets will contribute to scientific knowledge, support evidence-based decision-making, and facilitate the future commercialization and adoption of IPUT devices in healthcare settings. The management and utilization of these datasets are crucial for maximizing the impact and societal benefits derived from the project's outcomes.

Med-IPUT will not incorporate data generated outside of the consortium.

2.2 Types and formats of the collected/generated data

The Med-IPUT project generates various types of data, which can be categorized primarily into two main types: measurement data and plans/designs. The formats and storage methods for each type of data are carefully selected to ensure accessibility, compatibility, and compliance with legal requirements.

1. Measurement Data:

Measurement data encompasses both numerical values and images obtained through various experimental procedures and simulations. This data plays a critical role in assessing the performance, characteristics, and reliability of the developed IPUT technologies. The majority of measurement data will be stored in MATLAB format, providing a widely used and versatile platform for data analysis and processing. Additionally, optical and electro-optical measurement data will be stored as .csv files, ensuring compatibility with common spreadsheet applications and facilitating data sharing and collaboration.

2. Plans/Designs:

Plans and designs form an essential component of the data generated in the Med-IPUT project. These include mask designs, process instructions, and other related documentation that guide the fabrication and manufacturing of IPUT devices. The formats for plans and designs are typically in electronic file formats such as PDF, CAD files (e.g., DWG, DXF), or specific software formats used by the respective design tools employed by the partners. These formats ensure the integrity and accuracy of the design information and facilitate easy access, revision control, and sharing among project partners.

The detailed information on the different types and formats of collected/generated data, including specific data types used by each participating institution, can be found in the appendix. The appendix provides a comprehensive overview of data types, their characteristics, and the intended use for each institution, ensuring transparency and clarity in data management.

The Med-IPUT project is committed to ensuring full compliance with all applicable legal requirements, including data protection regulations such as the General Data Protection Regulation (GDPR). As the measurement data primarily originates from non-personal sources such as phantoms, there are no foreseen issues with GDPR standards, as no personal data is collected during the research and development activities. Data storage costs are accounted for by funds allocated to each partner specifically for storage resources, and personnel are designated by each partner for data curation tasks to ensure the proper organization, documentation, and preservation of the datasets.

3. FAIR Data

Within Med-IPUT, quality-controlled data and associated data may be shared in a way that they are Findable, Accessible, Interoperable, and Reusable (FAIR), provided there is IP protection in place, and that there is value in sharing the data. E.g., sharing validation data may be useful for external reviewers or auditors, but sharing design plans may not be, as it would interfere with IPR and confidentiality.

3.1 Making data findable

2.1.1 Discoverability of data

Discipline compliant metadata elements will be used describing the data to aid data discovery and potential re-use. Data will be deposited in a trustworthy repository.

2.1.2 Identifiability of data – identification mechanism

A digital object identifier is generated when the datasets are uploaded to the repository, which in Med-IPUT's case is Zenodo for publicly shared data.

2.1.3 Naming conventions

Standard naming conventions will be used wherever possible. For optical measurement data, the following naming conventions will be used: the file name contains enough information to identify the origin of the data. The naming format used is Batch ID+Wafer ID +Reticle row number+Reticle column number+Device name+.file extension, for example: MED-IPUT_Wafer01_R1_C1_MZI1.csv

2.1.4 Approach for clear versioning

Keywords will be extracted from the file name and metadata generated from the file header containing information from the measurements.

2.1.5 Standards for metadata creation

Med-IPUT will use JSON files with a standardized metadata scheme. As data becomes available, this will be further developed.

3.2 Making data openly accessible

The Med-IPUT project recognizes the importance of disseminating research findings and knowledge generated within the consortium. The project aims to make some deliverables openly accessible to ensure widespread access to the latest advancements in IPUT technologies. However, it is crucial to note that data sharing will only occur if it aligns with the protection of intellectual property and confidentiality considerations.

The project website will serve as a central platform for sharing project deliverables and updates, ensuring easy access for stakeholders and interested parties. Additionally, the project will leverage social media channels to promote and disseminate key findings, research outputs, and advancements to a broader audience, fostering engagement and knowledge exchange.

Research outputs from the Med-IPUT project will be presented at international and national conferences, allowing experts in the field to gain insights into the project's progress and outcomes. Furthermore, the consortium aims to publish its research findings in reputable scientific journals, ensuring that the scientific community can benefit from and build upon the project's results. Whenever possible, these publications will be made openly accessible, promoting collaboration and further advancement in the field.

Regarding data accessibility within the consortium, measurement results obtained during the Med-IPUT project can be openly shared among consortium members. This enables effective collaboration and facilitates the execution of individual project work. However, simulation results, mask files, and in-line measurement results may be shared exclusively within the consortium, safeguarding the intellectual property and confidential aspects of the project.

Process instructions, which contain sensitive information critical to the development of IPUT technologies, will remain strictly confidential and will not be shared, ensuring the protection of proprietary knowledge and expertise.

For internal data sharing within the consortium, a dedicated platform Med-IPUT TNO Sharepoint will be utilized to securely exchange relevant information and maintain effective communication among project partners.

In cases where particularly valuable or interesting data emerges from the project, efforts will be made to make it available via open access platforms. Repositories such as https://repository.tno.nl/ or Zenodo provide suitable channels for preserving and sharing data of value to the broader research community. However, careful consideration will be given to the intellectual property rights and the significance of the data before deciding on open access sharing, ensuring that confidential and proprietary information is adequately protected.

In summary, the Med-IPUT project is committed to making its research outputs openly accessible and disseminating knowledge through various channels, including the project website, social media, conferences, and scientific publications. While measurement results can be openly shared, data that requires confidentiality, such as process instructions, will be strictly protected. Internal data sharing will be facilitated through secure platforms, and if appropriate, valuable data may be made available via open access repositories. The emphasis is placed on striking a balance between openness and the protection of intellectual property, ensuring the project's success and promoting advancements in the field of IPUT technologies.

2.3 Making data interoperable

A strong emphasis is placed on ensuring that the data generated throughout the research process is interoperable. This means that the data will be structured and formatted in a way that promotes seamless integration and compatibility with existing open applications and tools. By adopting widely used standard formats, the consortium aims to facilitate data exchange, collaboration, and the utilization of available resources effectively.

To achieve interoperability, the project will prioritize the use of standard and open formats whenever possible. These formats are widely recognized and accepted within the scientific community, enabling efficient data sharing and integration across different platforms and systems. By adhering to these established formats, the Med-IPUT project ensures that its data can be easily accessed, processed, and analyzed by researchers and stakeholders.

In cases where standard or open formats are not readily applicable, the project will define and provide format specifications. These specifications will outline the structure, organization, and data representation requirements necessary for effective data interoperability. By providing clear guidelines, the project ensures that the data can be accurately interpreted and utilized by external parties or future research endeavors.

Furthermore, the project acknowledges the importance of metadata, which provides essential information about the data's content, context, and characteristics. Metadata standards will be followed to ensure consistency and comprehensiveness in describing the datasets. By including detailed metadata, the project enhances the discoverability and interpretability of the data, allowing other researchers to understand and leverage its value more effectively.

The adoption of interoperable data formats and metadata standards within the Med-IPUT project not only facilitates data sharing and collaboration among project partners but also promotes the wider dissemination and utilization of the research outputs. By aligning with established practices, the project contributes to the development of a more interconnected research ecosystem, where data can be seamlessly integrated and leveraged to advance scientific knowledge and innovation.

In summary, the Med-IPUT project recognizes the importance of data interoperability and commits to using widely adopted standard formats and metadata standards whenever possible. By doing so, the project ensures that its data can be easily integrated into existing applications and shared with the scientific community. The adoption of interoperable data practices enhances collaboration, facilitates the exchange of knowledge, and contributes to the broader research ecosystem.

4. Data Security

Med-IPUT’s partners will not produce sensitive data.

The appendix includes detailed information about data security for each participating institution separately.

5. Ethical Aspects

For Med-IPUT, no ethical issues are foreseen. All participants will be in compliance with the national legal frameworks.

Appendix includes detailed information about ethical aspects for each participating institution separately.

APPENDIX

1. Data Summary. Institution: TNO Provide a summary of the data addressing the following issues
1.1 State the purpose of the data you generate yourself, or the data you collect from other sources
The main purpose is to generate ultrasound data from a phantom with a probe based on a state of the art transducer elements and a probe based on IPUT sensors. The data will be used for benchmarking.
1.2 Explain the relation of the data to the objectives of the project
The main objective of Med-IPUT is to develop and demonstrate IPUT sensing technology with the ambition to be 100x more sensitive.
1.3 Specify the types and formats of data you generate or collect (e.g., what standards are you using, such as DICOM, what is the volume of the data, as in number of datapoints not the actual file size)
Most data contains pulse echo or PA measurements. The data consist of 32 to 256 traces each of 10k - 25k sample points. The data will be stored in Matlab format (mathworks.com).
1.4 Specify if existing data is being re-used (if any)
Not applicable.
1.5 Specify the origin of the data
The data will be originated from a standard phantom for medical ultrasonic imaging and a standard photo acoustic phantom.
1.6 State the expected sample size of the data or storage size (if known)
Each record will be between 2.5 and 50 MB.
1.7 Outline the data utility: to whom will it be useful
In the near future for other ultrasound research lab that are using phantoms to develop ultrasonic hardware and algorithms. When IPUT is matured, the benchmarking data is useful for the OEMs and MDs.

2. FAIR Data
2.1 Making data findable, including provisions for metadata
2.1.1 Outline the discoverability of data (metadata provision)
Good description of the used phantom. **Good description of the set-up used.** Good description of the KPIs and their values and how they are extracted from the data.
2.1.2 Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

This will be described in a separate document, and preferable in an article.
2.1.3 Outline naming conventions used
Standard keywords will be used.
2.1.4 Outline the approach towards search keywords
Standard keywords will be used.
2.1.5 Outline the approach for clear versioning
Data and documenting descriptions will be reviewed first before made available.
2.1.6 Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
It will be developed when data is available.
2.2 Making data openly accessible
2.2.1 Specify which data will be made openly available? If some data is kept closed/private, provide a rationale for doing so
The data mentioned in section 1.1 will be made available.
2.2.2 Specify how the data will be made available
Internal data via Med-IPUT Sharepoint. If interesting data is available, it will be made available by open access via https://repository.tno.nl/ or Zenodo.
2.2.3 Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Documentation will be in Office365. Data and scripting in matlab format.
2.2.4 Specify where the data and associated metadata, documentation and code are deposited
Via https://repository.tno.nl/ or Zenodo.
2.2.5 Specify how access will be provided in case there are any restrictions
If data is made available internally, the Med-IPUT TNO Sharepoint will be used.
2.3 Making data interoperable
2.3.1 Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability
Matlab is well known and matlab data / code is easy accessible. The description of the data will be in accomplished self-explanatory documentation.
2.3.2 Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Documentation will be fully understandable by experts in the field.
2.4 Increase data re-use (through clarifying licenses)
2.4.1 Specify how the data will be licensed to permit the widest reuse possible
No licenses will be needed.
2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
If interesting phantom data is available, it will be made available for open access.
2.4.3 Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
If the data obtained from the IPUT sensor technology is significantly better, other users can verify the data that is made available. Algorithms to further improve the data quality or to extract additional data can be validated on the sota and IPUT data sets.
2.4.4 Describe data quality assurance processes
See 2.1.5.
2.4.5 Specify the length of time for which the data will remain re-usable
5 years after the end of the project.

3. Allocation of Resources Explain the allocation of resources, addressing the following issues:
3.1 Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
10,000 EUR per data set, will be part of the Med-IPUT WP6 and 7.
3.2 Clearly identify responsibilities for data management in your project
This is part of the validation and dissemination activities in Med-IPUT. Vermon and EIBIR are WP-leaders.
3.3 Describe costs and potential value of long-term preservation
Not applicable in the repository of TNO.

4. Data Security
4.1 Address data recovery as well as secure storage and transfer of sensitive data
No sensitive data will be produced.

5. Ethical Aspects
5.1 To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Briefly describe: • the project tasks (number), that require ethics approval • procedures and criteria that will be used to identify/recruit research participants • in case children and/or adults unable to give informed consent are involved, provide details on how the consent of the legal representatives will be acquired • confirm that templates of the informed consent/assent forms and information sheets will be kept on file • Describe details on incidental findings policy • Provide copies of approvals by ethics committees, if available. If not available, indicate the date of availability • For clinical studies provide: (i) Final version of study protocol as submitted to regulators/ethics committee(s), (ii) Registration number of clinical study in a WHO-or ICMJE-approved registry (with the possibility to post results), (iii) Approvals (ethics committees and national competent authority if applicable) required for invitation/enrolment of first subject in at least one clinical centre. For each clinical study, a report on the status of posting results in the study registry(s), including timelines if/when final posting of results is scheduled after end of funding period. If not yet available, indicate the date of availability
No ethical issues are foreseen. The data will only be generated from standard phantoms.
5.2 Provide contact details of the person responsible for the ethics approval
Not applicable.
5.3 Processing of Personal Data • Submit a declaration of compliance with respective national legal framework(s) • confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research • Confirm that the data used in the project is publicly available and can be freely used for the purposes of the project, • In case of further processing of previously collected personal data, confirm that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.
Not applicable.
5.4 Please provide the name of the person(s) that can be approached for further information regarding the above information.
Not applicable.

6. Other **6.1 Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)**
Not applicable.

1. Data Summary. Institution: Ligentec Provide a summary of the data addressing the following issues
1.1 State the purpose of the data you generate yourself, or the data you collect from other sources
Generate data for standard process procedure for Membrane fabrication in the LGT SiN platform. Simulation and measurement data for SiN-based membrane ring resonator sensors and sensor arrays for the MED-IPUT project, which can further be used in the PDK and process portfolio.
1.2 Explain the relation of the data to the objectives of the project
The main objective for the project is to provide Ring Resonator based IPUT devices including heater for frequency tuning. The SiN PICs will then be integrated in photoacoustic IPUT demonstrator devices by the project partners.
1.3 Specify the types and formats of data you generate or collect (e.g., what standards are you using, such as DICOM, what is the volume of the data, as in number of datapoints not the actual file size)
Definition of a standard procedure (SOP) for the fabrication process (pdf). **PDK component development.** Optical and electro-optical measurement data stored as (.csv).
1.4 Specify if existing data is being re-used (if any)
Not applicable.
1.5 Specify the origin of the data
Not applicable.
1.6 State the expected sample size of the data or storage size (if known)
Not known.
1.7 Outline the data utility: to whom will it be useful
Data will be useful for project partners TNO and Vermon. Generated data will also be used in the future LGT component PDK.

2. FAIR Data
2.1 Making data findable, including provisions for metadata
2.1.1 Outline the discoverability of data (metadata provision)
Not applicable.
2.1.2 Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
Not applicable.
2.1.3 Outline naming conventions used

Standard keywords will be used.
2.1.4 Outline the approach towards search keywords
Standard keywords will be used.
2.1.5 Outline the approach for clear versioning
Data and procedure will be reviewed before releasing.
2.1.6 Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
Not applicable.
2.2 Making data openly accessible
2.2.1 Specify which data will be made openly available? If some data is kept closed/private, provide a rationale for doing so
No answer.
2.2.2 Specify how the data will be made available
Project related will be made available to all project partners via the TNO-Sharepoint drive.
2.2.3 Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Documentation will be in Office365 format. Data will be in .csv format.
2.2.4 Specify where the data and associated metadata, documentation and code are deposited
No answer.
2.2.5 Specify how access will be provided in case there are any restrictions
If data is made available internally, the Med-IPUT sharepoint will be used.
2.3 Making data interoperable
2.3.1 Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability
Not applicable.
2.3.2 Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
Not applicable.
2.4 Increase data re-use (through clarifying licenses)
2.4.1 Specify how the data will be licensed to permit the widest reuse possible
No answer.

2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
No answer.
2.4.3 Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
No answer.
2.4.4 Describe data quality assurance processes
No answer.
2.4.5 Specify the length of time for which the data will remain re-usable
No answer.

3. Allocation of Resources Explain the allocation of resources, addressing the following issues:
3.1 Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
No answer.
3.2 Clearly identify responsibilities for data management in your project
This is part of the validation and dissemination activities in Med-IPUT. Vermon and EIBIR are WP-leaders.
3.3 Describe costs and potential value of long-term preservation
No answer.

4. Data Security
4.1 Address data recovery as well as secure storage and transfer of sensitive data
No sensitive data will be produced.

5. Ethical Aspects
5.1 To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Briefly describe: • the project tasks (number), that require ethics approval • procedures and criteria that will be used to identify/recruit research participants

- in case children and/or adults unable to give informed consent are involved, provide details on how the consent of the legal representatives will be acquired
- confirm that templates of the informed consent/assent forms and information sheets will be kept on file
- Describe details on incidental findings policy
- Provide copies of approvals by ethics committees, if available. If not available, indicate the date of availability
- For clinical studies provide: (i) Final version of study protocol as submitted to regulators/ethics committee(s), (ii) Registration number of clinical study in a WHO-or ICMJE-approved registry (with the possibility to post results), (iii) Approvals (ethics committees and national competent authority if applicable) required for invitation/enrolment of first subject in at least one clinical centre. For each clinical study, a report on the status of posting results in the study registry(s), including timelines if/when final posting of results is scheduled after end of funding period. If not yet available, indicate the date of availability

No ethical issues are foreseen.

5.2 Provide contact details of the person responsible for the ethics approval

Not applicable.

5.3 Processing of Personal Data

- Submit a declaration of compliance with respective national legal framework(s)
- confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research
- Confirm that the data used in the project is publicly available and can be freely used for the purposes of the project,
- In case of further processing of previously collected personal data, confirm that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.

Not applicable.

5.4 Please provide the name of the person(s) that can be approached for further information regarding the above information.

Not applicable.

6. Other

6.1 Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

No answer.

1. Data Summary. Institution: Vermon Provide a summary of the data addressing the following issues
1.1 State the purpose of the data you generate yourself, or the data you collect from other sources
Data will be mostly parameters measured during acoustic and electrical characterization of transducers.
1.2 Explain the relation of the data to the objectives of the project
Transducers will be made, and Vermon is in charge of a providing piezo-based probes. The piezoelectrical transducer will be measured to check the performances. Vermon may perform measurements on IPUTs as well.
1.3 Specify the types and formats of data you generate or collect (e.g., what standards are you using, such as DICOM, what is the volume of the data, as in number of datapoints not the actual file size)
Data will be in the MATLAB format (.mat). The precise content is unknown for now.
1.4 Specify if existing data is being re-used (if any)
No.
1.5 Specify the origin of the data
Internal measurements on characterization setups.
1.6 State the expected sample size of the data or storage size (if known)
Unknown.
1.7 Outline the data utility: to whom will it be useful
It will be useful for TNO and for publication of the project's results.

2. FAIR Data
2.1 Making data findable, including provisions for metadata
2.1.1 Outline the discoverability of data (metadata provision)
No answer.
2.1.2 Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
Unknown.
2.1.3 Outline naming conventions used
There is no naming conventions. Name of data is understandable for people skilled in the art.
2.1.4 Outline the approach towards search keywords
No answer.

2.1.5 Outline the approach for clear versioning
No answer.
2.1.6 Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
No answer.
2.2 Making data openly accessible
2.2.1 Specify which data will be made openly available? If some data is kept closed/private, provide a rationale for doing so
General data about the acoustic performances may be available. To be confirmed. Data which could give hints about fabrication secrets will be kept private.
2.2.2 Specify how the data will be made available
Unknown.
2.2.3 Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
Any software which can open .mat files.
2.2.4 Specify where the data and associated metadata, documentation and code are deposited
Unknown.
2.2.5 Specify how access will be provided in case there are any restrictions
No answer.
2.3 Making data interoperable
2.3.1 Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability
No answer.
2.3.2 Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
No answer.
2.4 Increase data re-use (through clarifying licenses)
2.4.1 Specify how the data will be licensed to permit the widest reuse possible
It will not be.
2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
No answer.

2.4.3 Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
No answer.
2.4.4 Describe data quality assurance processes
No answer.
2.4.5 Specify the length of time for which the data will remain re-usable
No answer.

3. Allocation of Resources Explain the allocation of resources, addressing the following issues:
3.1 Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
No answer.
3.2 Clearly identify responsibilities for data management in your project
No answer.
3.3 Describe costs and potential value of long-term preservation
No answer.

4. Data Security
4.1 Address data recovery as well as secure storage and transfer of sensitive data
No answer.

5. Ethical Aspects
5.1 To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Briefly describe: • the project tasks (number), that require ethics approval • procedures and criteria that will be used to identify/recruit research participants • in case children and/or adults unable to give informed consent are involved, provide details on how the consent of the legal representatives will be acquired • confirm that templates of the informed consent/assent forms and information sheets will be kept on file • Describe details on incidental findings policy

- Provide copies of approvals by ethics committees, if available. If not available, indicate the date of availability
- For clinical studies provide: (i) Final version of study protocol as submitted to regulators/ethics committee(s), (ii) Registration number of clinical study in a WHO-or ICMJE-approved registry (with the possibility to post results), (iii) Approvals (ethics committees and national competent authority if applicable) required for invitation/enrolment of first subject in at least one clinical centre. For each clinical study, a report on the status of posting results in the study registry(s), including timelines if/when final posting of results is scheduled after end of funding period. If not yet available, indicate the date of availability

No answer.

5.2 Provide contact details of the person responsible for the ethics approval

No answer.

5.3 Processing of Personal Data

- Submit a declaration of compliance with respective national legal framework(s)
- confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research
- Confirm that the data used in the project is publicly available and can be freely used for the purposes of the project,
- In case of further processing of previously collected personal data, confirm that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.

No answer.

5.4 Please provide the name of the person(s) that can be approached for further information regarding the above information.

No answer.

6. Other

6.1 Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

No answer.

1. Data Summary. Institution: VTT **Provide a summary of the data addressing the following issues**
1.1 State the purpose of the data you generate yourself, or the data you collect from other sources
All data will be generated by VTT. The generated data includes optical simulation results, mask designs, process instructions, results of the in-line measurements during the device processing, and results of the optical measurements of the fabricated components.
1.2 Explain the relation of the data to the objectives of the project
Data is related to the design and fabrication of sensitive Mach-Zehnder based IPUT devices and their manufacturability (Objectives 2-4) and developing the read-out chip for mass parallelization of the sensors (Objective 5).
1.3 Specify the types and formats of data you generate or collect (e.g., what standards are you using, such as DICOM, what is the volume of the data, as in number of datapoints not the actual file size)
Optical simulation results: plain text files (.txt), ASCII Mode File (.amf), Binary Mode File (.bmf), Excel Open XML Spreadsheet (.xlsx). Number of data points: Hundreds to thousands. Mask layout designs: Open Artwork System Interchange Standard (OASIS) files (.oas) and/or Graphic Data System files (.gds) Process instructions: text files (.txt) In-line measurements: SEM and optical micrographs (.tif; .jpg) Optical measurements: Comma separated values files (.csv) and text files (.txt). Number of datapoints: Hundreds to thousands.
1.4 Specify if existing data is being re-used (if any)
No existing data will be re-used.
1.5 Specify the origin of the data
Optical simulation results are obtained by commercial optical simulation tools from Ansys Lumerical and Photon Design. Data can also result from numerical models developed by the designers. Mask files are generated by in-house python-based mask design software. Process instructions are generated by engineers in charge of the process based on the internal process guidelines. In-line measurement data is generated by SEM or optical microscopy of the fabricated wafers and devices. Optical measurement data is generated by measuring the fabricated components and test structures. Measurements are made to evaluate the spectral characteristics of the devices on different polarizations. In-house lab software for instrument control based on LabVIEW is used to acquire and save the data in files containing a header with information about the measurement, such as variables and units.
1.6 State the expected sample size of the data or storage size (if known)
Simulation results: <100 MB Mask files: <100 MB Process instructions: <1 MB SEM & optical micrographs: < 1MB Optical measurement data: <1 MB
1.7 Outline the data utility: to whom will it be useful
Simulation, design and processing data will be useful for VTT team members for developing and optimizing the fabrication process of the MZI based IPUTs and read-out circuit and measurement data for Med-IPUT partners to verify the properties of the fabricated devices.

Data set description	Type	Format	Standards & meta-data	Size
Optical measurement data	Plain text	Comma-separated	JSON Schema, Metadata:	<10 MB

		values files (.csv)	variable labels, value labels	
SEM images	Images	.jpeg, .tif	none	<1 MB

2. FAIR Data
2.1 Making data findable, including provisions for metadata
No answer.
2.1.1 Outline the discoverability of data (metadata provision)
Discipline compliant metadata elements will be used describing the data to aid data discovery and potential re-use. Data will be deposited in a trustworthy repository.
2.1.2 Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
A digital object identifier is generated when the datasets are uploaded to the repository, which in our case is Zenodo for publicly shared data.
2.1.3 Outline naming conventions used
Optical measurement data: The file name contains enough information to identify the origin of the data. The naming format used is Batch ID+Wafer ID +Reticle row number+Reticle column number+Device name+.file extension, for example: MED-IPUT_Wafer01_R1_C1_MZI1.csv
2.1.4 Outline the approach towards search keywords
Keywords will be extracted from the file name and metadata generated from the file header containing information from the measurements.
2.1.5 Outline the approach for clear versioning
For data made available publicly, we will exploit Zenodo's DOI versioning in which a DOI per version is created in addition to a Concept DOI representing all versions of a record and semantically linking all the per-version DOIs.
2.1.6 Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how
JSON metadata standard
2.2 Making data openly accessible
2.2.1 Specify which data will be made openly available? If some data is kept closed/private, provide a rationale for doing so
Only measurement results can be made openly accessible. Simulation results, mask files, and in-line measurement results may be shared between the consortium members if that is needed for the execution of their own project work. Process instructions are strictly confidential and will not be shared outside VTT.
2.2.2 Specify how the data will be made available
Measurement results can be made available when publishing the results of the project. Measurement and other data (except the process instructions) can be made available for the consortium members by sharing through project's Teams site, in consortium meetings or by email to specific partners.
2.2.3 Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
The measurement result data (.csv) is directly readable e.g., by Excel, Notepad++.
2.2.4 Specify where the data and associated metadata, documentation and code are deposited
All data will be deposited at VTT's file repositories. Part of the data will also be made available for the consortium at the consortium's Teams site. Openly accessible data will be shared through Zenodo or CSC's data repository IDA.
2.2.5 Specify how access will be provided in case there are any restrictions
For the openly accessible data there are no restrictions.
2.3 Making data interoperable

2.3.1 Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability
Standard file formats will be used.
2.3.2 Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
Standard vocabulary in the photonics research field will be used
2.4 Increase data re-use (through clarifying licenses)
2.4.1 Specify how the data will be licensed to permit the widest reuse possible
Creative Commons (CC) licenses for opened data will be used.
2.4.2 Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed
Openly accessible measurement data does not need a data embargo. It will be available upon publishing.
2.4.3 Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why
Openly accessible measurement data is accessible and useable by third parties. All the other data will not be made useable for the third parties due to its confidential nature.
2.4.4 Describe data quality assurance processes
During the data generation stage, we will perform repeated measurements. The data will be captured following pre-stablished internal procedures and we will follow a file naming convention to identify the origin of the results. Finally, we will implement data review to evaluate the scientific quality of the datasets, including methods and reusability.
2.4.5 Specify the length of time for which the data will remain re-usable
No definite period or time limit is planned for access or re-use of the data.

3. Allocation of Resources Explain the allocation of resources, addressing the following issues:
3.1 Estimate the costs for making your data FAIR. Describe how you intend to cover these costs
Costs related to research data management and opening are eligible as part of the project grant.
3.2 Clearly identify responsibilities for data management in your project
VTT is responsible for managing the data generated within work package 3 related to SOI platform development.
3.3 Describe costs and potential value of long-term preservation
VTT will use Zenodo or IDA as repositories, which are free to use. Zenodo is operated by CERN and IDA is offered by the Finnish Ministry of Education and Culture and produced by CSC – IT Center for Science Ltd, both of which comply with FAIR principles. Data will be stored there for as long as the repositories exist. Backups will be kept internally.

4. Data Security
4.1 Address data recovery as well as secure storage and transfer of sensitive data
No sensitive data will be transferred or stored outside VTT.

5. Ethical Aspects
5.1 To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Briefly describe: • the project tasks (number), that require ethics approval • procedures and criteria that will be used to identify/recruit research participants

• in case children and/or adults unable to give informed consent are involved, provide details on how the consent of the legal representatives will be acquired • confirm that templates of the informed consent/assent forms and information sheets will be kept on file • Describe details on incidental findings policy • Provide copies of approvals by ethics committees, if available. If not available, indicate the date of availability • For clinical studies provide: (i) Final version of study protocol as submitted to regulators/ethics committee(s), (ii) Registration number of clinical study in a WHO-or ICMJE-approved registry (with the possibility to post results), (iii) Approvals (ethics committees and national competent authority if applicable) required for invitation/enrolment of first subject in at least one clinical centre. For each clinical study, a report on the status of posting results in the study registry(s), including timelines if/when final posting of results is scheduled after end of funding period. If not yet available, indicate the date of availability
Not relevant
5.2 Provide contact details of the person responsible for the ethics approval
Not relevant
5.3 Processing of Personal Data • Submit a declaration of compliance with respective national legal framework(s) • confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research • Confirm that the data used in the project is publicly available and can be freely used for the purposes of the project, • In case of further processing of previously collected personal data, confirm that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects.
No personal data will be processed
5.4 Please provide the name of the person(s) that can be approached for further information regarding the above information.
Not relevant

6. Other 6.1 Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)
No answer.

EIBIR will neither collect nor produce any data.