Grant agreement no: 101097101

TERahertz ReconfigurAble METAsurfaces for ultra-high rate wireless communications

TERRAMETA

Deliverable D1.1
Data Management Plan (initial)

Technical Manager: George Alexandropoulos
Organization: NKUA

Project Coordinator: Luís Pessoa
Organization: INESC TEC

Start date of project: 01-Jan-2023

Date of issue: 28-06-2023

Due date: 30-06-2023

TERRAMETA Ref:
TERRAMETA_INESC_012_F_ WP1
Deliverable 1.1

Leader in charge of deliverable: Luis Pessoa, INESC TEC

Dissemination level		
PU	***Public***	***X***
CO	***Confidential, only for members of the consortium (including the Commission Services)***	

The TERRAMETA project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101097101, including top-up funding by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee.

Co-funded by the European Union

Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union, SNS JU or UKRI. The European Union, SNS JU or UKRI cannot be held responsible for them.

 Named Distribution Only

Table of Contents

Change register

Version	Date	Author	Organization	Changes
A	08-06-2023	Luis Pessoa	INESC TEC	Initial draft
B	15-06-2023	Joana Tavares	INESC TEC	Gathering contributions from partners
C	21-06-2023	Luis Pessoa	INESC TEC	Revision and formatting. Created a table for the Data Summary and added new contents to the FAIR section.
D	28-06-2023	João Castro	INESC TEC	Revision
E	28-06-2023	Joana Tavares	INESC TEC	Minor revisions
F	29-06-2023	Luis Pessoa	INESC TEC	Final revision, added references

1. Statement of independence

The work described in this document is genuinely a result of efforts pertaining to the TERRAMETA project: any external source is properly referenced.

Confirmation by Authors: Luis Pessoa, INESC TEC
Joana Tavares, INESC TEC
João Aguiar Castro, INESC TEC
Kyriakos Stylianopoulos, NKUA
Antonio Clemente, CEA
Lucca Luci, CEA
José Luis Jimenez, CEA
Juan Andres Vasquez-Peralvo, uni.lu
Bo Kum Jung, TUBS
Thomas Kuerner, TUBS
Qi Luo, UH

2. Abbreviations

ADS – Advanced Design System
3D – 3 Dimensional
CST – Computer Simulator Technology
DMP – Data Management Plan
EC – European Commission
EIRP – Effective Isotropic Radiated Power
EOSC – European Open Science Cloud
FAIR – Findable, accessible, interoperable, and re-usable
GB – Giga Bytes
HFSS – High Frequency Structure Simulator
RF – Radio Frequency
RIS – Reconfigurable Intelligent Surface
VNA – Vector Network Analyzer

3. Executive summary

This report consists of the initial version of the project Data Management Plan (DMP) based on the template recommended by the European Commission (EC), describing how data generated in the project is exploited, accessed, curated, and preserved. This initial version outlines the initial data management plan according to the current development of the project. It identifies a first set of the data to be generated, collected, and processed during the project, according to the different data categories, and proposes the data management process that will be followed in the next developments. It also outlines the approach to make data Findable, Accessible, Interoperable, and Re-usable (FAIR) by indicating what data will be generated, collected, and processed, what standards and metadata will be applied, methods for data preservation and data public sharing. This document will also briefly address the ethical and privacy aspects as well as some data security principles.

4. Introduction

The TERRAMETA project aims to investigate revolutionary technologies for 6G and demonstrate the feasibility of Terahertz (THz) Reconfigurable Intelligent Surface (RIS) assisted ultra-high data rate wireless communications networks. Novel high-performance hardware including the THz RIS and THz transmitter/receiver will be developed and advanced network analysis/optimizations

techniques will be developed using these real THz components. The proposed TERRAMETA THz network will be driven by 6G usage scenario requirements.

The present deliverable describes the data management procedures for the data to be generated, collected, and processed during the project. It outlines the approach to make data Findable, Accessible, Interoperable, and Re-usable (FAIR) by indicating what data will be generated, collected, and processed, what standards and metadata will be applied, methods for data preservation and data public sharing. This document will also briefly address the ethical and privacy aspects as well as some data security principles. As a living document, the DMP will be updated with a periodicity of 6 months by the INESC TEC data steward, João Aguiar Castro, hereafter referred as DMP manager. The DMP manager will directly interact with the Project Coordinator, Luis Pessoa (INESC TEC), which is the responsible for the data management in the project. Therefore, the DMP will be updated according to the evolution of the project:

- The initial version D1.1 (M06, i.e., June 2023) outlines the initial data management plan according to the current development of the project. It identifies a first set of the data to be generated, collected, and processed during the project, according to the different data categories, and proposes the data management process that will be followed in the next developments.
- The interim versions (M12, M18, M24, M30) will provide necessary updates to the document, whether to inform about the availability of data or the practices implemented by the project to manage the datasets.
- The updated version D1.2 (M33, i.e., September 2025) will provide the final description of the data generated, collected, and processed in the project, as well as the final process and the tools that have been established for managing the project data.

5. Data Summary

The planned data generation, collection, and re-use, including purpose and details of data types/formats/origin/volume is given in the data summary Table 5.1. Each row of the table should be read from left to right, starting with the data identification. For each identified type of data, including the identification of the method used for data obtention (generated, collected or re-used) and the data category, namely research, context, administrative data:

- Research data includes all aspects of data generated, collected, or re-used in the context of research, from simulation models and results, experimental measurements and results, and data processing code.
- Context data refers to the collection or generation of data related with the context of the research, such as pictures/videos of experimental setups and conducted experiments, experimental notes (e.g., configuration, topology, hardware models and software versions), prototype fabrication details (materials, methods to fabricate, process guidelines, dimensions, and characterization of materials and functionality).
- Administrative includes data generated by the project management activities (such as meeting minutes, recordings, internal management reports, enabling historical tracking and follow-up), project documentation (such as documents, reports, presentations slides and deliverables) and publications (public scientific/technical article publications, scientific/technical presentation slides, news articles, newsletters, social media posts).

After the data type (including category) identification in the left column, the other table columns provide a more detailed analysis of the purpose of data generation, collection and re-use, description of the data types, data formats, data origin and expected data volume (in Giga Bytes - GB). This section will be updated on an ongoing basis to further detail the information about the data collected by the project.

Table 5.1 – Data summary - overview of planned data generation, collection, and re-use, including purpose and details of data types/formats/origin/volume.

Generated, collected or re-used data type identification	**Purpose of data generation, collection, and re-use**	**Data type description**	**Data formats**	**Data origin**	**Data volume**
Generation of RF hardware simulation data, including the digital models of the designs Re-use of impedance model for the fast characterization of the unit–cell frequency behaviour in terms of scattering parameters Category: research data	Estimate in simulation environment the RF performance of the designed test fixtures, RF switch devices, antenna unit cells, and antenna arrays.	Test fixtures and RF devices electromagnetic and electric circuit models (S-parameters versus frequency), Radiation patterns of antenna unit cells and antenna arrays (gain or directivity versus observation angle), Circuit model of the unit cells and its scattering parameters (S-parameters versus frequency), Radar cross-section of the reconfigurable intelligent surface, Transmitted power (EIRP versus frequency) and receiver noise figure, Unit cell frequency response (S-parameters in s2p format), Full array frequency response (ASCII format), Radiation patterns cuts of full arrays (ASCII format), E- and H- fields from unit cells and full arrays (ASCII format), Power flow simulations of full arrays (ASCII format).	image formats (.png, .jpeg, .eps, .pdf), raw data formats (.csv, .mat, .xlsx, .txt) HFSS/CST/ADS proprietary files Digital models of the designs - gerber file (.grb)	Specialized 3D electromagnetic simulation software (HFSS, CST) and circuit level simulation software (ADS, SPICE), Ad-hoc numerical tools developed in MATLAB.	Simulation data: 100 GB Digital models: 1 GB

		Digital models of the designs: test fixtures, antenna unit cells and antenna arrays.			
Generation of link and system level simulation data Data re-use: Simulation scenarios for the system level simulation of backhaul networks have been created within the ThoR project. Part of that data has been generated by TUBS and can be re-used in other projects. The scenarios, that can be re-used are those form Hanover and Berlin. Category: research data	To assess impact, feasibility, and scalability of RIS in THz backhaul network (e.g., for coverage extension)	Ray-tracing data for the base station coverage analysis. Scenario data for software simulation, set up in the Horizon2020 EU-project ThoR. This is available at TUBS in the Simulator for Mobile networks (SiMoNe).	image formats (.png, .jpeg, .eps, .pdf), raw data formats (.csv, .mat, .xlsx, .txt)	Simone link-level and system-level simulator	10 GB
Collection of experimental characterization data from RF devices Category: research data	Measure the obtained RF performance of the fabricated test fixtures, RF switch devices, antenna unit cells and antenna arrays. Extract spice compact models for the RF	Experimental characterization results of different RF switch devices (S-parameters versus frequency), Experimental characterization results of radiation patterns of antenna unit cells and antenna arrays (gain or directivity versus observation angle).	image formats (.png, .jpeg, .eps, .pdf) raw data formats (.csv, .mat, .xlsx, .txt) S-parameters (.s1p, .s2p)	RF switch device characterization data collected from 300 mm on-wafer probing station equipped with a state-of-the art VNA and a semiconductor parameter analyzer.	10 GB

	switches to be used for system design.			Antenna measurements collected from anechoic chamber and specific characterization facilities equipped with state-of-the-art equipment (e.g., VNA, frequency up and down converters up to 330 GHz).	
Collection of experimental data collected from the wireless channel Category: research data	To enable modelling the characteristic of RIS and the communication channel including RIS	Channel sounding data	raw data formats (.csv, .mat, .xlsx, .txt) S-parameters (.s1p, .s2p)	Collected from channel sounder equipment	10,000 GB – 100,000 GB
Generation (and re-use of previously developed) MATLAB, Java, C, Python codes Category: research data	MATLAB, Java, C, and Python codes can be used to process simulation and measurement data in order to extract new insights from raw data. Re-used code will be used for array synthesis and optimizations (synthesis of phase gradient on the electromagnetic surface aperture and radiation pattern calculation)	MATLAB codes for data/signal processing algorithms.	.mat, .java., .c, .py, .npy.	Generated by the project researchers	0.1 GB

Collection of pictures/videos of experimental setups Collection of experimental notes, Collection of prototype fabrication details <u>Category</u>: context data	Reporting purposes (deliverables and publications).	Screenshots of simulation models, setups, and results. Pictures/videos of experimental setups and conducted experiments. Experimental notes (e.g., configuration, topology, hardware models and software versions) Prototype fabrication details (materials, methods to fabricate, process guidelines, dimensions, and characterization of materials and functionality).	Pictures/videos (.jpeg, .mp4), text notes (.txt, .docx, .xlsx)	Collected using smartphones or other image capture devices.	50 GB
Collection of pictures /videos of project dissemination activities <u>Category</u>: administrative data	Raise awareness to the project and ensure effective dissemination of the project aims and outputs (e.g., in social networks).	Pictures/videos of project dissemination activities.	Pictures/videos (.jpeg, .mp4)	Collected using smartphones or other image / video recording devices.	50 GB
Generation of publications <u>Category</u>: administrative data	To publicly disseminate the project scientific and technical results and create impact.	Scientific/technical article publications, scientific/technical presentation slides, news articles, newsletters, social media posts	.pdf, .docx, .tex, pptx, .html	Generated by the project researchers and project management team including the project manager, the WP leaders and task leaders	10 GB

Generation of project documentation Category: administrative data	Reporting, dissemination, communication, and standardisation (internal and external)	Documents, reports, presentations slides, deliverables	.docx, .tex, pptx, .xlsx, .pdf	Generated by the project management team including the project manager, the WP leaders and task leaders, and possibly researchers and administrative staff	100 GB

In order to gather information about the data produced, or collected by the project, a Data Inventory Register (DIR) will be developed and made available to project members in the project´s drive.

The DIR can be defined as a catalogue of metadata, where incremental information about the data can be updated. The DMP manager is responsible for maintaining the DIR and update the DMP accordingly. The information in the DIR needs to be completed by the people responsible for the data, and should be as rich as possible.
Table 5.2. outlines the metadata elements to collect with the DIR.

Table 5.2 – Data Inventory Register metadata elements

Metadata information	**Descriptive information**
Responsible Party	The person(s) responsible for the creation/maintenance of the dataset.
Title	A descriptive name for the dataset, with contextual information – *what, where, when.*
Data Type	The type of data, for example, experimental, observational or simulation data.
Date Created	Date of creation of the dataset. YYYY-MM-DD.
Number of Data Files	The total number of files that make up the dataset.
Source	Where the data originates. Is the data reused? Recommended best practice is to identify the related resource.
Format	The files format - .txt .xml .csv
Size	The total weight of the dataset.
Mode of Data Collection	The method(s) used to collect data.
Data Collection Instrument	The instruments used to generate and or process the data.
Support Documentation	Whether documents have been created to provide context for the data – experimental protocols and *readme files.*
Location	Place where the data is stored – institutional drive, data repository.
Availability	How is the data available? – private, project partners or open.
Access conditions	Who have access to the data? What are the conditions for accessing the data.
Rights and Restrictions	Data licence. CC BY-SA 4.0 is recommended for data sharing.

6. FAIR data

TERRAMETA will adopt an *open as possible but close as necessary* approach to data sharing. Taking into the different requirements that the project data may have, the most suitable approach to make each dataset available must be evaluated on case-by-case basis, although a general approach is described in this document. Whenever possible, project data will be made openly available, with respect to possible embargo periods, through data repositories; Persistent identifiers are assigned to the published data and are included in its metadata, while also be listed in the DMP; Metadata should be accessible in the circumstances that access to the data itself is restricted; Metadata is based on standard vocabularies, whenever possible, but customizable metadata may also be required.

6.1. Repository selection

By default, processed data and data supporting publication results, will be made publicly available. TERRAMETA will consider a set of different data repository solutions, depending on the datasets and on specific journal data policies – in this case the data can be published on repositories recommended by the publishers, with the support from the DMP manager, if needed

6.1.1. INESC TEC data repository

The INESC TEC data repository [IDR], based on the CKAN open-source data management system, is the institutional service to enable TERRAMETA partners to publish their data. The INESC TEC data repository can be accessed by any user at any time and from any location. Therefore, selected datasets will be deposited, preserved, and made available via the INESC TEC data repository. The workflow for data publishing is the responsibility of the DMP manager, which ensures that all the datasets are granted a DOI via the DataCite Fabrica service. By the time a dataset reaches the deposit stage, the DMP manager will provide a template for the representation of a minimal set of metadata elements, including the necessary metadata for citation information.
The published datasets are indexed in Dataset Search and other services, while the repository is registered in the Re3data.org.

6.1.2. Zenodo

As a complementary data publication approach, for the sake of data dissemination, Zenodo, the catch-all repository for EU funded research, will be part of the TERRAMETA data publication plan. Zenodo enables the definition of communities responsible for the curation of their resources. TERRAMETA will consider the creation of a project community in order to publish not only the data, but also to disseminate different project outputs. Zenodo provides a DOI as part of the dataset upload workflow will also be integrated to reporting lines for research funded by the EU.

6.1.3. European Open Science Cloud

The European Open Science Cloud (EOSC) is a trusted digital platform for the scientific community, providing seamless access to data and interoperable services that address the whole research data cycle, from discovery and mining to storage, management, analysis and reuse across borders and scientific disciplines. The EOSC Portal available on the web is the universal access channel to EOSC services and resources. TERRAMETA will align with the EOSC recommendations and aim to provide Open Data at different technical levels:

- Open datasets of experimental and simulated data. The following datasets are planned:
  - Simulation results of test fixtures, RF devices electromagnetic and electric circuit models (S-parameters versus frequency), radiation patterns of selected antenna unit cells and antenna arrays (gain or directivity versus observation angle),
  - Simulation results using SiMoNe (system/link level simulation)
  - Processed channel sounder measurements
  - Experimental RF characterization data of antenna unit cells, antenna arrays and RF switch devices which should complement publication of the results in

conferences or journals since original data availability is important to allow for other researchers to reproduce the research results. This includes raw data, processed measurement data, and figures.

- Open-source code (e.g., real-time RIS phase profile array synthesis)
- Open access to scientific publications
- Open access to research outputs beyond datasets, such as software, models, algorithms, and workflows through deposition in the above-mentioned trusted repositories (while retaining partners IPR).

6.2. Data documentation and metadata standards

Data sharing, especially with external parties, must include metadata and the necessary documentation to make the data findable and reusable beyond the project timeframe. Research and context data is expected to be useful other researchers working in the field of THz wireless communications, enabling the reproduction of project outputs and results as well as helping to validate/cross check simulation and experimental results using other platforms beyond the scope of this project. Administrative data, including public project documentation and publications are expected to be useful to the wider research community, to the media, policy makers and to the general public. as well as to stakeholders interested in the application of the project.
Data may need to be documented at different levels, considering task, file and variable or item level. Hence, the documentation of TERRAMETA data needs to inform on how the data was collected or processed, citation information, study design, instruments and units of measures used, how the different files in a dataset relate to each other, labels explaining the meaning of variables, among others.
In the absence of other types of documents, such as research protocols, a Readme file, as part of the dataset should be adopted as a common practice.
Research data also needs to be made interoperable. This means that data needs to be prepared and handled so that it can be exchanged and combined with other data, moved between services and systems without information content loss or disordering. The integration of metadata in a structured fashion in ReadMe files, as well as an integral part of data repositories data deposit forms, together with the adherence to established standards are an important good practice which will be followed to ensure data interoperability. Both the INESC TEC data repository and Zenodo already implement an appropriate metadata standard. The data will be made available, namely in the metadata provided by the INESC TEC data repository. The datasets deposit at the INESC TEC data repository can be completed with the necessary custom metadata fields, preferably based on available metadata standards.
Concrete examples of additional custom metadata fields that can be considered are:

- Simulation options and conditions, RF/electric parameters of the considered devices, considered design dimensions and material properties.
- In RF characterization best practices will be followed. All data files will contain in easily recognizable and plain readable format at least: timestamp, RF probes, cables and instrument serial numbers, calibration specs, wafer and die IDs, unequivocal device, bias parameters and laboratory identification (institution, location).

6.3. Other practices to promote FAIR

Licence

Provided that the datasets do not have conditions to access and reuse, as a standard rule, they will be made available under Creative Commons Attribution-ShareAlike 4.0 International [CCAS], the recommended licence for research data, which specifies that credit must be given to the creator, but adaptations must be shared in the same terms. Other licences may be considered depending on the dataset requirements.

DOI and citation information

When making data available in the INESC TEC data repository the DOIs will be registered and managed via the DataCite Fabrica service [DFS], while the data citation information will include the minimal

elements according to the DataCite (APA reference style) – *Creator (PublicationYear): Title, Publisher. DOI.*
The DMP manager is the responsible for minting the DOI. The published dataset titles will provide information about the subject matter, and of the most important parameters to contextualize their content, in order to be as descriptive as possible.

File naming

File naming conventions will also be considered to make sure that the file names remain meaningful and useful beyond their original creation and storage location – outlasting the person who created the file. Some conditions must be guaranteed, such as, file names have included descriptive information, must be consistent and avoid the use of special characters, full-stops or spaces. When adopting a file name strategy, a **quick access guide** enabling decoding the elements that make up the filename.

File date formatting

To maintain the chronological order and simplify the process of sorting and browsing data files, when applicable, dates need to use the Year-Month-Day format, ensuring compliance with the standard for the representation of date and time, ISO 8601.

Versioning

The recommended practice is to use consecutive numbering for major version changes, with the use of decimals for minor changes. If the number of versions is high, the priority is to retain a copy of the original “raw” data and most recent version. Together with a well-documented record of changes.

7. Other research outputs

In addition to the management of data, the management of other research outputs will be considered following the same principles. Other research outputs that may be generated or re-used throughout the project can include software, workflows, protocols, and models, beyond those already identified in the data summary in this document.

8. Allocation of resources

The Project Coordinator (Luis Pessoa, INESC TEC) is responsible for the data management, in person, with the support from INESC TEC’s data management specialist João Aguiar Castro. Specific costs related with the management and maintenance of INESC TEC’s institutional data repository, to ensure that the generated, collected and re-used data is FAIR, may be covered by the project.

9. Data security

TERRAMETA will define a set of preliminary recommendations regarding data storage and backup, to be refined in the following DMP updates.
Data must be stored in the institutional network drive, who must be routinely backed up with account authentication systems to prevent unauthorised access. Personal storage devices are suitable solutions for daily work; however, project members must ensure that the backup of data is regularly made through suitable network drive. Data backups will be made regularly, and data should be stored, preferably, in at least two different media.
Managing access to specific files also must consider more granular control of who is able to handle specific datasets, being recommended that an (external) user account is created for individuals requiring access to data. Moreover, strong passwords are also recommended for security purposes.

The datasets on the INESC TEC’s institutional data repository is backed-up every night, and a backup is created every week to be stored externally from the main premises.

10. Ethics

We do not expect any ethical issues within the project. e.g., we are not collecting any personal data.

11. Other issues

The transfer of private research data between the researchers from different partners will be done using the repository platform of the project, which is established at INESC TEC using a drive service provided by Nextcloud, which runs on premises. This service has a proper access control to block unauthorized access.

12. Conclusion

This Data Management Plan presents the initial overview (at Month 6) of the plan that was prepared for managing data and other research outputs within the project. It describes the data summary, including the planned data obtention method, the associated purpose, and the details of data types / formats / origin / volume. The adopted data identification includes the method used for data obtention (generated, collected or re-used) and the data category, namely research, context, administrative data. The document also describes how FAIR principles will be addressed as part of the data management. It is important to note that the project commits to align with the EOSC recommendations and aim to provide Open Data at different technical levels: (1) Open datasets of experimental and simulated data, (2) Open-source code, (3) Open access to scientific publications, (4) Open access to research outputs beyond datasets. By ensuring FAIR principles on the provided data, we expect to increase of project impact since the data should be useful to other researchers working in the field of THz wireless communications, enabling the reproduction of project outputs and results as well as helping to validate/cross check simulation and experimental results using other platforms beyond the scope of this project.
This document represents the initial release of the Data Management Plan. However, the Data Management Plan is a living document that will be regularly updated until the end of the Project. In the final release (expected at month 33), more details will be provided comparatively, specifically regarding the description of the datasets (data summary), as well as regarding how FAIR principles are addressed.

13. References

[IDR] https://rdm.inesctec.pt/
[CCAS] https://creativecommons.org/licenses/by-sa/4.0/
[DFS] https://doi.datacite.org