Horizon 2020 Grant Agreement no: 814523

Tera**h**ertz end-to-end wireless systems supp**o**rting ultra-high data **R**ate applications

ThoR

Deliverable D1.5

Mid-term Data Management Plan

Coordinator (EU): Thomas Kürner
Organisation: Technische Universität Braunschweig

Coordinator (Japan): Tetsuya Kawansihi
Organisation: Waseda University

Start date of project: 01-Jul-2018

Date of issue: 18-Dec-2019
Due date: 31-Dec-2019
ThoR Ref: ThoR_VIV_191218_A_WP1

Leader in charge of deliverable: Bruce Napier
Vivid Components

Project co-funded by the European Commission within the Horizon 2020 programme and the National Institute of Information and Communications Technology in Japan (NICT)		
Dissemination level		
PU	**Public**	**X**
PP	**Restricted to other programme participants (including the Commission Services)**	
RE	**Restricted to a group specified by the consortium (including the Commission Services)**	
CO	**Confidential, only for members of the consortium (including the Commission Services)**	

Contents

Change register

Version	Date	Author	Organisation	Changes
A_DRAFT	18-Dec-2019	Bruce Napier	Vivid	Initial
A	18-Dec-2019	Bruce Napier	Vivid	No reviewer changes: typos and formatting only.

Reviewed by Thomas Kürner TUBS A_DRAFT 18-Dec-2019

Statement of independence

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

Confirmation by Authors: Bruce Napier Vivid Components
[ThoR consortium also provided input]

Executive summary

This document summarises the data management plan being utilised in the ThoR project, which is based mainly on existing procedures at each partner, and the project management manual (D1.1).

The automated tool DMP Online was used to produce this document.

https://dmponline.dcc.ac.uk

1. Data summary

State the purpose of the data collection/generation
In general, the data collection is in order to facilitate research and development both within the project and for future work.

Individual partner comments include:

CIT
Data is collected to enable future THz antenna and propagation research.

DTAG
Data is collected and generated for future use in DTAG.

Gifu
Data is generated and collected to enable current and future research.

IAF
Data is generated and collected to enable current and future research and development.

Siklu
Data is collected to enable future use by Siklu employees to deliverables and research results

USTUTT
Data is generated and collected to enable current and future research.

Vivid
Contact information for the ThoR Community and Advisory Group

Waseda
Data is collected to enable future THz system research.

TUBS
Data is generated and collected to enable current and future research.

Explain the relation to the objectives of the project

In general, the data collection is necessary to perform the project work, including dissemination.
Individual partner comments include:

CIT
The data will be used for THz antenna and propagation research.

DTAG
The data is/will be used for the analysis and evaluation of 5G and beyond 5G network deployment scenarios.

Gifu
The data will be used for THz and millimeter-wave research.

IAF
The data is required for the development and analysis of solid-state THz front ends, components and technologies, which are enabling high-data-rate THz communications.

Siklu

The data is likely to be used in the design of future wireless products

USTUTT
The data is required for the design of future THz front ends, which are enabling high-data-rate THz communications and research in the field of ultra-high data rate transmissions.

Vivid
Necessary for effective dissemination

Waseda
The data will be used for THz and millimetre-wave radio system research.

TUBS
The data used and collected to enable current and future research and development of THz communication systems including planning algorithms for backhaul/fronthaul links

Specify the types and formats of data generated/collected

There are various formats, including:

CIT
Data includes experimental results and simulation results of THz antenna and propagation.

DTAG
Data includes the measured and simulated results, network design.

Gifu
Data includes experimental results and simulation results (CST) related to THz research. Data includes near-field patterns, radiation patterns, transmission results, etc.

IAF
Simulation data is generated using EM and RF circuit simulation tools like ADS and CST.
Simulation/modelling as well as measurement data of devices is processed using appropriate software tools like Python, Origin, EXCEL etc.
Processed measurement and simulation results of devices and integrated circuits is stored using ASCII text files as well as Touchstone files (usable in conventional RF circuit simulators).
Layouts of devices and circuits are mainly stored in GDSII format. 3D models of devices/modules are stored in STEP format.

Siklu
Data includes system design related information (design documents, schematics, PCB layout, etc.), system measurement results

USTUTT
Data includes system design-related information generated using EM and RF circuit simulation tools like ADS and CST.
Data includes also the results of transmission experiments.

Vivid
Contact information in Microsoft Outlook and Excel

Waseda
Data includes experimental results and simulation results related to THz system research.

TUBS
Data includes the measured and simulated results, network design.

Specify if existing data is being re-used (if any)

No reuse of existing data is being done.

Specify the origin of the data

There are various origins of the data in the project, including:
CIT
Data are generated in the course of the experiment and simulation about THz antenna and propagation research in this project.

DTAG
Data is generated/created in the simulation tools and network planning tools.

Gifu
Data is generated/created in the simulation and experiment.

IAF
The data is generated in device modelling, simulation and characterization (see above).

Siklu
Data is typically being generated in the course of the project for any advancements done within its scope

USTUTT
Data is generated/created in the simulation tools and by the measurement equipment.

Vivid
All WP6 data is freely provided by the ThoR Community and Advisory Group.

TUBS
Data is generated/created in the simulation tools and network planning tools.

State the expected size of the data (if known)

The total amount of data generated in the project is expected to be <100 Gb.

Outline the data utility: to whom will it be useful

Partner comments on data utility from their perspectives:

CIT
The data is useful for antenna and propagation researchers.

DTAG
The data is useful for the network planning engineers.

Gifu

The data will be useful for researchers of the Gifu University for the future development of Terahertz communication systems, new types of antennas, propagation research, etc.

IAF
The data will be useful for employees and partners of Fraunhofer IAF for the development of THz devices and systems.

Siklu
The data is useful to Siklu employees if deciding to develop products based on the technology investigated in the project

USTUTT
The data will be useful for researchers of the University of Stuttgart for the future development of Terahertz communication systems.

Vivid
This contact information will not be used beyond ThoR.

Waseda
The data is useful to researchers and collaborators of the Waseda University for future THz research and standardization.

TUBS
The data will be useful for students and researchers at TUBS for studies and the future development of Terahertz communication systems.

2. FAIR data

2.1. Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Some of the data (such as design documents) is searchable as text. Other types of data, such as schematics, layouts and measurements are not directly searchable but are described by attributes such as filename, or are linked from searchable documents

Re. TUBS SiMoNe data:
The data TUBS is mainly using and archiving is background and simulation data implemented in the in-house-developed tool Simulator for Mobile Networks (SiMoNe). This tool stores its data in an SQL data base which is searchable both via SQL commands and through the Graphical User Interface of SiMoNe. For the exchange of data with partners import and export routines (with converters) exist enabling the exchange of ASCII data, e.g. which is searchable through filenames.

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?

The data will have standard identifiers.

Outline naming conventions used

All partners will use the agreed ThoR naming conventions are used as defined in the project management manual (D1.1, section 7.3):

"In a project of this size, many hundreds of documents will be generated. For ease of use a register of all project documents based on a unique reference based on the document creation date will be included on the website. The initial version will be A; updates B, C etc. will retain the original creation date so that this acts as an identifier.
E.g. ThoR _TUBS_180715_B_WP3 Document title XYZ referring to:
ThoR_<partner>_<doc creation date (YYMMDD)>_<version>_<WP> Document title XYZ
Please keep the document title brief and descriptive. If it is a deliverable then simply state the deliverable number and (possibly abbreviated) name.
Examples:
ThoR_TUBS_180823_A_WP1 Regular call minutes 2018-08-23
ThoR_DTAG_181022_B_WP2 KO meeting presentation
ThoR_VIV_180831_A_WP1 D1.1 Project management manual"

Outline the approach towards search keyword

N/A for most of the project data.
Keywords are important for not searchable data. When such data is linked or named, describing keywords should be included.

Outline the approach for clear versioning

Draft versions are indicated with "DRAFT" in the filename.
Issued versions are named A, B, C, etc. as per the naming convention above (D1.1).
Change registers are included on all project documents.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

N/A

2.2. Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Data in public deliverables is available through the deliverables on the project website.

Most data from the project is confidential, for several reasons, including:
- geographic data has licensing conditions
- network data provided by partners is confidential.
(-Selected parts of (artificial but still realistic) reference scenario data will be made publicly available through the web page of TUBS and the ThoR website.)
-Other data is kept confidential for academic contributions (e.g. contents for journal or conference papers should be unpublished.)
-Company confidential material

Specify how the data will be made available

Data will be available through public deliverables, conferences, and journal papers.

The public data is published at ThoR web site and on some members' webpages.

Selected parts of (artificial but still realistic) reference scenario data will be made publicly available through the web page of TUBS and the ThoR website.

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)?

Standard tools such as PDF reader and web browser (HTML) are needed to access the public data.
Reference scenario data will be made available in ASCII format along with a format description.

Specify where the data and associated metadata, documentation and code are deposited

The data is stored in Powerfolder hosted at the TUBS, and at ThoR public web site.

Specify how access will be provided in case there are any restrictions

No restrictions to access public deliverables. Internal data, which is not publicly available is available through the password-protected Powerfolder at TUBS.

2.3. Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Data is mostly in text format so should pose no limit to interoperability. If non-standard data formats are used, the data format will be described and the document with the description will be made available along the data.
Two-byte characters will be avoided.

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?

ThoR will use standard vocabulary.
If necessary a mapping will be made available with an accompanying document.

2.4. Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

No restriction for public data.

Non-public data will be kept confidential.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

The public data is available via ThoR web site.
Internal data is kept as confidential.

TUBS will also make the public simulation scenarios available via TUBS and ThoR website.
No restrictions in terms of life-time of the data. Good scientific practice rules enforces to keep data available at least for a period of 10 years.

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

Public data may be re-used if properly cited.

Describe data quality assurance processes

Data quality is assured by the review process as described in D1.1 project management manual.

Specify the length of time for which the data will remain re-usable

No restriction on reusability of public data.
Good scientific practice suggests data be available at least for a period of 10 years.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Costs TBD
These costs will be borne by the individual partners as necessary.

Clearly identify responsibilities for data management in your project

Each partner to be responsible for its own data management as per this document and D1.1.

TUBS administrators for SiMoNe and Powerfolder will be responsible for maintaing the data.
Bruce Napier (Vivid) will maintain the mailing lists and contact info.

Describe costs and potential value of long term preservation

N/A

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Partners each have internal back-up procedures.

For project documents on Powerfolder, the data is backed-up and mirrored to back-up servers.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

N/A

6. Other

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

N/A