HORIZON 2020

Terahertz based Ultra High Bandwidth Wireless Access Networks

Deliverable ID:	Preparation date:
D4.6	08 April 2019

Milestone: Final Proposed

Title:

Interim Data Management Plan and Report

Editor/Lead beneficiary (name/partner):
Cameron Barlow / NPL

Internally reviewed by (name/partner):
Dr. Zeta Dooly / TSSG
Dr. Bruce Napier / Vivid

Approved by:
PSC

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

Revisions				
Version	**Date**	**Author**	**Organisation**	**Details**
0.1	01.02.19	Cameron Barlow	NPL	First draft
0.2	03.04.19	Cameron Barlow	NPL	Second draft
0.3	04.04.19	Alan Davy	WIT	Amendments to current draft
0.4	04.04.19	Bruce Napier	Vivid	Minor mods based on discussions
0.5	05.04.19	Cameron Barlow	NPL	Proposed final version for review

The research leading to these results has received funding from the EU H2020 under grant agreement n° 761579.

Table of contents

List of tables

Executive summary

The Interim Data Management Plan is a framework for data management. It details issues and questions to be addressed in developing a data management policy within the project, and outlines the methodologies and solutions that will be adopted. Types of data expected in the project are described, together with planned provisions for making data findable and available to project partners and other interested parties. This deliverable acts as a guideline to ensure consistent use of data (formats) and data management for the duration of the research project. A final data management plan is planned for M36.

1 Introduction

1.1 Summary

This deliverable presents the Interim Data Management Plan. It lists issues and questions to be addressed in developing a data management policy within the project, and outlines the methodologies and solutions that will be adopted. Types of data expected in the project are described, together with planned provisions for making data findable and available to project partners and other interested parties.

1.2 Structure of this document

This document is structured as proposed in the H2020 Guidelines on data management that makes data findable, accessible, interoperable and reusable (FAIR).

1.3 Relationships with other deliverables

This document is the second deliverable of a set of deliverables describing TERAPOD's data management plan. Therefore, it is related to deliverables D4.5 and D4.7 as its precursor and its successor. Deliverables which have data as integral to their progression include D5.2, D5.4 and D5.6 for systems modelling and protocol development; D6.1, D6.2, D6.3, D6.4, D6.5 and D6.6 for demonstration purposes; D4.1, D4.2, D4.3 and D4.4 for characterisation of devices & systems and finally; D3.6, D3.7 and D3.8 for revised device design, integration and packaging.

1.4 Contributors

The following partners have contributed to this deliverable:

- Cameron Barlow (NPL)
- Mira Naftaly (NPL)
- Johannes Eckhardt (TU Braunschweig)
- Alan Davy (TSSG)
- Zeta Dooly (TSSG)
- Bruce Napier (Vivid)

2 Data Summary

The following purposes of data collection have been identified (this list may be extended as the project progresses):

- Device characterisation data required for device specifications and for link design.
- Free-space propagation data required for link design.
- Measurements, modelling and simulation data are necessary inputs to model and verify the wireless THz data transmission system.

The data will be used for device engineering, for modelling and simulations, and as a basis for designing wireless links. It will also be a development resource for THz communications technologies.

Types of data produced:

- Measurement data: device characterisation, free-space propagation, channel measurement.
- Scenario data: measured and simulated scenarios based on use cases defined in WP2, typical/generic geometrical environments.
- Modelling data: channel modelling produced in task 4.2.2.
- Simulation data: simulation results from WP5, physical and system layer simulations, open-source simulation software.

No data-re-use is foreseen at this point.

All data will be generated by the project.

Size of the data.

- Measurement data originating from device characterisation and free-space propagation will be low-volume; expected total size below 1 Gb.
- Channel measurement data will be high-volume, of the order of 1 Tb.
- Modelling data will be medium-volume, of the order of 100 Gb.

The data will be utilised by project partners, and will be made available to third parties as open-source data (see Section 3.2).

3 FAIR data

3.1 Making data findable

- Metadata provision:
  - Data will be sorted by category
  - A file will be provided and regularly updated, listing the type of data, its filename, and relevant information as to its nature.
  - A file will be provided listing all abbreviations in use.
  - For data originating from measurements, simulation, and device characterisation a table of contents will be provided, showing data structure.
  - Each category of data will have the same folder structure.
  - Each scenario will have its own identifier.
- Data files will have standard identifiers listed in Table 1.
- The naming convention will be:

  TERAPOD, ID, name, version, date (e.g. TERAPOD_D4.6_InitialDMP_v0.1_20171128)

Table 1: standard identifiers

Standard identifier	Document type
PSC	documents relating to the Project Steering Committee
PM	documents relating to Plenary Meetings
R	project review documents
DMP	documents relating to the Data Management Plan
WP#	work package related documents (e.g. WP4)
D#	deliverable reports and related documents (e.g. D4.5)
CM	channel measurement data
DC	device characterisation data
P	propagation measurement data

3.2 Making data openly accessible

As far as possible the processed data will be openly available.

Types of data not available:

- Raw data.
- Data violating personal rights.
- Data violating company interests.

Most of the processes data will be available as clear text files, importable into software packages such as Matlab, Excel, and Origin. Where necessary, software conversion script will be provided.

Data will be stored in the Project OwnCloud hosted at the TSSG, Waterford Institute of Technology. On a case by case basis, links to data will be registered with appropriate Open Access Data Repositories such as OpenAIRE, with agreed access restrictions and procedures in place.

3.3 Making data interoperable

Processed measurement data from device characterisation and free-space propagation measurements will be stored as tabulated values in tab-separated columns with named column headings in ASCII plain-text files that can be read or imported into suitable software tools (e.g. MATLAB, EXCEL, Origin).

Processed channel measurement data will be provided as MATLAB data files that can be directly loaded into MATLAB.

Scenario data will be in the form of a generic data type containing the scenario data in an XML file, where every point in space is characterised by several properties.

Modelling data will be recorded as statistical parameters listed in an appropriate table format (e.g. EXCEL). The format of the stochastic channel models will be defined in the course of the project.

Simulation data will be produced by MATLAB or C#, with the exact format determined by the simulation software. If open-source simulation software is provided, the data can be directly imported and used.

3.4 Increase data re-use

A licensing policy will be introduced within the project. Processed measurement data will be made available as open source. Raw unprocessed data will not be made available.

Open source data will be made available at the conclusion of the project. The project will have innovation cycles that will depend on confidentiality. Final analysed data will be published.

Re-using the data:

- Measurement data pertaining to devices and propagation will be re-usable.
- Channel measurements will take place in a specific environment in a data centre, and therefore are not generally usable.
- Simulation data will be re-usable or reproducible, because simulation software will be open-source.

Data will be maintained for 10 years, for use by other projects, researchers or development engineers, as recommended in the Guidelines on the Handling of Research Data by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG).

4 Allocation of resources

Costs for cloud service: TSSG Data Centre will host any services required by the project. This will cost approximately €4,000 over the duration of the project to provide maintenance and management of the resources being accessed by the project.

Responsibilities for data management: Each partner will contribute to the Data Management Plan for their own generated data sources. Each partner generating data will be responsible for operating this DMP.

OwnCloud will be hosted at the TSSG, Waterford Institute of Technology, this is a data and project documentation repository. Costs will be included in the overall maintenance and management costs of TSSG Data Centre resources to the project.

5 Data security

Data will be hosted on TSSG OwnCloud or on partner's own repository. Data security will be provided by the host.

Raw data will be held in local repositories; with data security provided locally.

Data will be secured by TSSG OwnCloud backup policies, local as well as remote (off site).

6 Ethical aspects

The project addresses aspects of data transmission in the THz domain. There will be no personal data acquired or revealed. All stored data will be anonymised.

7 Other

Country-dependent issues:

- Germany: minimum of 10-year storage must be guaranteed for research data as recommended in the Guidelines on the Handling of Research Data by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG).

8 Conclusion/Further work

A framework for data management has been described. The issues relating to data management have been identified and the methodologies for addressing them have been outlined. Types of data expected in the project have been described, together with planned provisions for making data findable and available to project partners and other interested parties. A detailed data management plan will be developed for each identified data source as the project progresses. These Data Management Plans will be kept up to date and will be reported as amendments to the DMP deliverables.

References

Deutsche Forschungsgemeinschaft: Guidelines on the Handling of Research Data, online available at http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/antragstellung/nachnutzung_forschungsdaten/ , accessed 18.01.2018

European Commission: H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, European Commission Directorate-General for Research & Innovation, Version 3.0, July 2016

Digital Curation Centre: DMPonline, online available at https://dmponline.dcc.ac.uk/, Version v0.3.2, 2017

OpenAIRE: How to create a DMP Plan, online available at https://www.openaire.eu/opendatapilot-dmp , November 2016, accessed 30.11.2017