TERA6G

TERAHERTZ FOR 6G

TERAhertz integrated systems enabling 6G Terabit-per-second ultra-massive MIMO wireless networks

Deliverable D6.3a: Data Management Plan (FIRST RELEASE)

Date of issue: 31-07-2023
Due date: 31-07-2023

Leader in charge of deliverable: UC3M

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)**	

TERA6G 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 101096949

Document updates

Version	Date	Author	Organization	Changes
00	20/06/2023	JMD Mendinueta	UC3M	Deliverable template
01	20/06/2023	JMD Mendinueta	UC3M	Deliverable index and expected contributions from partners
02	27/07/2023	JMD Mendinueta	UC3M	First complete version
03	28/07/2023	Several authors	Several organizations	Incorporated dataset and storage facility information from partners
04	31/07/2023	JMD Mendinueta	UC3M	Incorporated contributions by LIX. Final version of the deliverable

Abbreviations

APC	Author Processing Charges	**OA**	Open Access
DMP	Data Management Plan	**ORCID**	Open Researcher and Contributor IDentifier
DOI	Digital Object Identifier	**VCS**	Version Control System
MIMO	Multiple-input multiple-output		

Statement of independence

The work described in this document is genuinely a result of efforts pertaining to the TERA6G project. Any external source is properly referenced.

Confirmation by authors:

José Manuel Delgado Mendinueta, Carlos III University of Madrid (UC3M)
Guillermo Carpintero del Barrio, Carlos III University of Madrid (UC3M)

1 Index

Page intentionally left blank.

1 Executive Summary

This document provides the FIRST RELEASE of the Data Management Plan (DMP) for the TERA6G project. This project participates in the Open Science Practices, introduced in the Annotated Grant Agreement (Article 17), which aims at the publication of the project results in open access form, the re-use and reproducibility of the data generated by the project, and the use of open data repositories.

To align with this directive, in its initial phase the TERA6G project has identified several data sets which will be generated within the project and that will be shared with the research community. The main objectives of the project are to improve the exchange and dissemination of the research results, hence enabling and promoting a wider validation of the project results, and to encourage a fair comparison and evaluation of different solutions in the technical areas of the TERA6G project.

The DMP of the TERA6G project, described in this deliverable, tackles with all the concerns about the treatment of the data that will be collected, processed, or generated during the whole project lifecycle and has been structured in compliance with the guidelines and the templates conveyed by the European Commission (EC).

The DMP, for each of the datasets identified in the TERA6G project, will consider the following aspects:

- Types of data generated, collected, or processed
- Standards used to manage data
- Data exploitation methodology
- Accessibility to data produced by the project
- Data dissemination level
- Data preservation and re-use

The DMP is a document with a **rolling-update policy**, in which the information will be made available on a finer level of granularity through updates of the DMP document as the implementation of the project progresses and when significant changes occur. This document is the **first version of the DMP**, providing just an initial description of the expected datasets and their management plan, while more details and further datasets might be added as the project progresses. The document will be actively maintained and regularly updated with additional datasets or details about the existing ones for the whole duration of the project.

Keywords: Data management plan, DMP, open source, open science practices, project dissemination, project exploitation.

2 Definitions of Terminology

arXiv

arXiv is an online, open access repository of electronic scientific paper preprints and postprints. A **paper preprint** is a drafted scientific paper that has not been peer reviewed, or it's in the process of being peer reviewed for its acceptance into a journal. This process is known as green open access. A **paper postprint** is a peer reviewed, accepted-for-publication paper. Typically, a postprint has not the official template or formatting of the final published version, but its contents are identical to the final version of the published paper.

Dataset

Digital information created because of a research process which is not published in the form of research output such as a journal paper. Research data excludes purely administrative records or any other king of bureaucratic documents. The highest priority research data is the one which underpins a research output such as a publication in an archived journal. Research data do not include publications, articles, lectures, or presentations.

Data Management Plan (DMP)

A formal working document which outlines how datasets will be handled both during the active research phase and after the project is completed. A DMP in some form or another is now a requirement for research grant proposals, and therefore must be addressed at the earliest phase of the research project lifecycle.

Digital Object Identifier (DOI)

A persistent alphanumeric identifier used to uniquely identify objects and to redirect communications to the correct Internet server where the object is stored, developed by the International DOI Foundation. DOIs are in widespread use mainly to identify academic, professional, and government information, such as journal articles, documents in conference proceedings, research technical reports, research datasets, and official publications.

External data

This type of data is defined as existing data that has already been generated and reported by other research initiatives or companies, which might be used as inputs in the TERA6G project for experimental validation activities or simulation studies. This type of data has a strong potential since it enables the comparison of the TERA6G solution with other existing architectures and technologies. TERA6G will act only as recipient of this type of data and this data will be just re-used "as-is", hence the project will not define any specific strategy for their management because external data is already handled by other organizations or entities.

Github

Github is an online platform to store repositories that make use of the **git version control system** (VCS). Git is primarily use by software developers for source version control, access control, bug tracking, and task management. However, the git VCS can be used to control any kind of data, whether it is computer source code or not. Github can be used for free, and it is currently the biggest repository for open-source projects and the largest source code host. Github does not generate digital object identifiers (DOIs), but the data can be protected and verified by the own cryptographic algorithms of the git VCS.

IEEE DataPort

IEEE DataPort is a dataset repository created and managed by the Institute of Electrical and Electronics Engineers (IEEE). Due to the nature of the research fields covered by the TERA6G project, naturally most of the publications will be in IEEE-managed conferences, transactions, or journals. DataPort allows IEEE authors to link the datasets into the published research journals and to automatically update their ORCID profile. IEEE DataPort creates a DOI for each uploaded dataset, which enables the citation of the dataset in other research papers.

Internal data

This type of data will be produced within the project as reference inputs for TERA6G validation activities in simulation, emulation, or experimental test-bed environments as well as outcomes of the TERA6G technological blocks evaluation and demonstration. For each internal data dataset, the project will define a suitable data management strategy, identifying proper procedures for their documentation, sharing, and maintenance.

Metadata

Information template about datasets stored in a repository or database. For example, an image may include metadata that describes how large the picture is, the color depth, the image resolution, when the image was created, the image's author, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, when the document was last modified, and a short summary of the document.

Postprint

A postprint is a full draft of a research paper that is uploaded to an open access repository after the peer review process and has been accepted for publication. However, a postprint does not have the final typesetting or template of the journal. The term e-print is used as a synonym for postprint. Also, postprint is considered a form of green open access but no gold open access.

Preprint

A preprint is a full draft of a research paper that is uploaded to an open access repository before the peer review process and the journal acceptance. The process of uploading a preprint is also known as self-archiving or green open access. Most preprint repositories, such as arXiv, create a unique DOI for each uploaded preprint paper.

Repository

A digital repository is a mechanism for managing and storing digital content such as scientific documents, datasets, electronic books, or pictures, etc. For example, the Internet Archive is a general data repository. A digital repository can have a specific purpose. For example, arXiv is a repository for paper preprints and postprints and Zenodo is a repository for research datasets.

The term repository is also used as a synonym to database and may refer to the elements stored in a data repository. For example, a dataset may be stored as a git repository whereas Github is a collection of git repositories or a repository that stores repositories.

Secondary data

Sources that contain commentary on or a discussion about a primary source.

Zenodo

Zenodo is a general-purpose, open-access repository developed under the European OpenAIRE program and operated by CERN. It allows researchers to deposit datasets, research software, reports, and any other research related digital artifacts. For each submission, a persistent and unique digital object identifier (DOI) is minted, which makes the stored items easily citable.

3 Introduction

The European Commission (EC) has introduced the "COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (— ARTICLE 17)" guidelines in Annex 5 of the Annotated Grant Agreement[1], introduced in the Horizon Europe (HORIZON) Framework Programme, which aims to improve and maximize access to scientific information and re-use of research data generated by projects.

The European Community (EC) has adopted the **Open Science Policy** for publications and datasets as a strategy to improve the quality of the scientific results, the efficiency, and the transparency of the scientific processes. These **Open Science guidelines** are expected to boost the benefits of public investment in research funded under Horizon Europe Framework Programme, since knowledge and innovation have a central role in generating economic growth.

In this context, research data refers to information, facts, digital information, or numbers, collected to be processed and elaborated as a basis for reasoning and discussion[2],[3]. These Open Science requirements are based on both **green open access** and **gold open access**. Green open access, also known as self-archiving, consists of the submission of a draft of the research manuscript to an open repository such as arXiv. Gold open access means that the final, peer reviewed version of the paper, is publicly available and can be read by anybody for free. Apart from make the publications open access, projects must also aim to deposit the research data needed to validate the results presented in public, open access data repositories. To this end, projects need to consider at an early stage how they are going to manage and share the data they create or generate.

The TERA6G project[4] is governed by the Open Science guidelines, with the objective of improving the exchange and dissemination of the project research results. The generation, collection, and publication of research data within the project, as well as their sharing with the wider research community constitute an excellent opportunity to promote an effective validation of the project results and to encourage a fair comparison and evaluation of different solutions in the technical areas of the project.

This document introduces the **first version** of the Data Management Plan (DMP) that has been elaborated for the TERA6G project, in accordance with the guidelines of the EC. The DMP describes how to select, structure, store, and make public the information used or generated during the project, both considering scientific publications as well as generated research data. In particular, the DMP include following issues:

- What types of data will be generated or collected
- What data will be exploited, and which will be the data dissemination level
- What standards will be used to manage data and its associated metadata

[1] EU Grants, AGA – Annotated Grant Agreement, EU Funding Programmes 2021-2027, version 1.0 DRAFT (01 April 2023)
[2] EU Funding & Tender opportunities, Online Manual, accessed on 27 July 2023.
[3] Horizon Europe (HORIZON) Programme Guide, version 3.0 (01 April 2023), section 16. Open Science
[4] TERA6G website: https://www.uc3m.es/research/TERA6G

- How data will be preserved, including after project completion

It should be noted that the DMP reported in this document is an initial version, which provides an overview of the data sets initially identified by the TERA6G partners. For this reason, this document provides just an initial description of the expected data sets and their management plan, while more details and potentially further data sets might be added during the project implementation. The document will be actively maintained and regularly updated with additional data sets or details about the existing ones during the total duration of the project. As a bare minimum, according to the proposal updated versions of the DMP document are expected in time with the end of second report period of the project, i.e., in M24 (deliverable D6.3b in December 2024).

4 TERA6G project overview

The TERA6G project main purpose is to push forward the state-of-the-art of Integrated Radiofrequency Photonics (IRFP) and its associated network management and network integration. TERA6G is a crossroad of previous H2020 projects including TERAWAY, ARIADNE, TERRANOVA and FUDGE-5G, and aims to provide massive multiple-input multiple-output (MMIMO) multi-antenna techniques with carrier frequencies well above 100 GHz. The key enabler technology of the TERA6G project is the hybrid photonic integration.

The disruptive characteristics of the TERA6G project include:

- **Agility**. TERA6G aims to produce an agile fronthaul based on photonic RF signal generation. This enables the generation of ultra-wide bandwidth signals, with a 30 GHz for each channel, and the ability to use any modulation format or scheme. In addition, the carrier frequency can be continuously tuned over the range from 30 GHz to 450 GHz.
- **Scalability**. The TERA6G project uses the hybrid photonic integration technology as an enabling technology for scalable MIMO devices, capable of handling a large number of independent spatial beam channels with 2-dimension antenna arrays. The optical beamforming network (OBFN) is based on a Blass matrix.
- **Reconfigurability**. The frequency agility of the TERA6G modules and the ability to generate multiple spatial beam channels unravels the implementation of several functionalities, including wireless data transmission, radar ranging and channel sounding. In addition, the TERA6G projects aims to demonstrate the integration of the TERA6G modules with a standard network orchestrator to provide functionalities such as switching spatial beams ON or OFF, combine several spatial beam channels to increase the power and hence the reach, or to search for alternate routes in the case of a line-of-sight temporary blockage, showing network restoration and self-healing functionality.

Finally, the main objectives of the TERA6G project can be summarized as follows:

1. **Objective 1**. The development of a scalable MMIMO transmitter module based on a Blass matrix OBFN with a wide tunability in the range from 30 to 450 GHz.
2. **Objective 2**. The development of a scalable MMIMO incoherent multi-band receiver module.
3. **Objective 3**. The assembly and functional demonstration of the transmitter and receiver modules and the use of the spatial beam channels for several functionalities, including but no limited to terabit-per-second wireless transmission, channels sound and radar ranging.
4. **Objective 4**. “Fiber over the air” THz smart management, which includes the integration of THz wireless subsystems and technologies with current 3GPP and B5G mobile networks.
5. **Objective 5**. The development of algorithms for an adaptive, energy-efficient mobile fronthaul based on the agile spatial beam channels.

5 TERA6G – Data Management and Open Science Planning

This section describes the plans of the TERA6G consortium for the management of the different datasets of research data that will be produced, collected, and used for internal processing and validation in the project.

It should be noted that the present deliverable is released at the early stage of the project (M06), when the design of the modules and specifications of their components, the design of the network use cases, the TERA6G system and module specifications, as well as the specific validation plan are still under definition. For this reason, this report is an initial description of the expected datasets and their management plan, while more details and potentially further data sets will be added during the execution steps of the project that will follow on at later stages of its deployment.

5.1 Open access considerations

The EC Horizon Europe (HORIZON) - Programme Guide states that,

> *Some open science practices are mandatory for all beneficiaries per the grant agreement. They concern:*
>
> - *open access to scientific publications under the conditions required by the grant agreement;*
> - *responsible management of research data in line with the FAIR principles of 'Findability', 'Accessibility', 'Interoperability' and 'Reusability', notably through the generalised use of data management plans, and open access to research data under the principle 'as open as possible, as closed as necessary', under the conditions required by the grant agreement;*
> - *information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data;*
> - *digital or physical access to the results needed to validate the conclusions of scientific publications, unless exceptions apply;*
> - *in cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency18.*
>
> *These obligations are described in the Model Grant Agreement (Article 17) and detailed guidelines on complying with them are provided in the Annotated Grant Agreement (Article 17).*

The scientific information and datasets generated within the TERA6G project will be made public accordingly as long as there are no potential conflicts against commercialization, Intellectual Property Rights (IPR) protection of the knowledge generated such as patents or other forms of protection, and accordingly to the provisions of the Consortium Agreement (CA) and only if they are not associated to ethical issues as they are defined in the CA of the TERA6G project.

The datasets that will be published within TERA6G DMP will belong solely to the partners that generated the associated IPR, according to the provisions of the CA. Before a dataset is deposited in an open access repository, the partners responsible for uploading and preserving the data will seek consent for data preservation and sharing from all partners involved, according to the provisions of the CA and with the guidance of **Task 6.2 Protection and management of IPR, contribution to standardization activities and Data Management** leader (CMC partner). The respective partners, in collaboration with the Project Coordinator (PC) and the Task 6.2 Protection and management of IPR, contribution to standardization activities and Data Management leader, CMC, will be responsible for postponing or restricting data sharing to allow enough time for publishing the results in peer-reviewed

journals or for seeking patents. To facilitate handling of datasets especially where multiple partners are involved, to expedite data dissemination and ensure that no ethical issues are associated to the respective datasets, the process will be overseen by Task 6.2 Protection and management of IPR, contribution to standardization activities and Data Management leader.

5.2 Open access to publications

The first aspect to be considered in the DMP is related to the **open access** (OA) nature of the publications generated within the context of the project, meaning that the project publications will be available online to any user free of charge. In line with the Horizon Europe (HORIZON) Framework Programme Open Science guidelines about open access for scientific publications, the publications that will arise from the TERA6G project will be made public through the following three channels:

- **Open access publishing,** also referred to as **gold open access,** means that an article is immediately placed in open access mode on the publisher or scientific journal website. In this option, beneficiaries publish their results in open access journals, or in hybrid journals that sell subscriptions and offer the possibility of making individual articles openly accessible via the payment of author processing charges (APCs).

  TERA6G will preferably target gold open access journals which apply article processing charges for the article to be published and then it is freely available online via the Open Access journal site.

- **Self-archiving,** also referred to as **green open access** or **publication as a preprint**. In this option, the beneficiaries deposit a drafted version of the manuscript in an online reprint repository of their choice, such as arXiv. If this route is chosen, the beneficiaries will ensure open access to the publication within a maximum of six months, in-line with the open access obligations established by the EC.

- **Upload of scientific datasets to a repository** such as Zenodo, IEEE DataPort, or Github. Zenodo is free for the uploader and submitted datasets receive a DOI. IEEE DataPort requires that the uploader pays a fee, but the datasets are freely available for download. IEEE DataPort also generates a unique DOI for each uploaded dataset. Finally, Github is free to use for everyone. It does not generate a DOI, but the integrity and timestamp of the dataset are guaranteed by the cryptographic algorithms of the git version control system. Furthermore, Zenodo enables easy importation of Github repositories.

Additionally, according to the EC recommendation, whenever possible the TERA6G consortium will retain the ownership of the copyright for their work using a *License to Publish*, which is a publishing agreement between the author and the publisher. With this agreement, authors can retain copyright and the right to deposit the article in an Open Access repository, while providing the publisher with the necessary rights to publish the article.

All the publications will acknowledge the project funding. This acknowledgment will be included also in the metadata of the generated information since it allows to maximize the discoverability and visibility of publications and to ensure the acknowledgment of EU funding.

To ensure open access—via the data repository—to the bibliographic metadata that identify the deposited publication, the metadata must be in a standard format and will include all the following:

- the terms “European Union (EU)” and “Horizon Europe (HORIZON) Framework Programme”
- the Name of the action, the Acronym and Grant Number
- the publication date and length of embargo period (if applicable)

- a persistent identifier such as a DOI

5.3 Research data

Apart from the open access to publication explained in the previous section, the Open Science guidelines also applies to research data that will be generated within the TERA6G project. Specifically, two types of data sets will be considered in TERA6G project:

- **Internal data:** These data will be produced within the project as reference inputs for TERA6G validation activities in simulation, emulation, or experimental test-bed environments as well as outcomes of TERA6G technological blocks evaluation and demonstration. For each of these datasets, the project will define a suitable data management strategy, identifying proper procedures for their documentation, sharing and maintenance.
- **External data:** These data will be existing data that have been already generated and reported by other research initiatives, which might be used as inputs in TERA6G for experimental validation activities or simulation studies. This type of data has a strong potential since they allow for the comparison of TERA6G solution with other existing technologies. TERA6G will act only as recipient of this type of data and, if just re-used "as-is", the project will not define any specific strategy for their management which is already handled by other organizations.

6 Data and Datasets Summary

TERA6G will produce several datasets during the lifetime of the project. The data will be both quantitative and qualitative in nature and will be analyzed from a range of methodological perspectives for project development and scientific purposes. The data within the datasets will be available in a variety of easily accessible formats, including Postscript (PS, PDF, PDF/A), Excel (XLSX), Word (DOCX, RTF), Power Point (PPTX), image (PNG, SVG, JPEG, TIFF, WEBP), compressed formats (TAR.GZ, MTZ), MATLAB binary format (MAT, FIG), and text formats (CSV, XML, JSON, YAML).

Additionally, when possible, the datasets will be stored using the Open Document Format for Office Applications (ODF) open standard for data storage. The ODF standard, also known as OpenDocument, guarantees interoperability, accessibility of the data and future proof data readability owing to the open nature of the file formats. The OpenDocument standard formats include ODT for word processing documents, ODS for spreadsheets, ODP for presentations, ODG for graphics and ODF for formulas and mathematical equations.

Table 1 includes a summary of the expected datasets produced by the TERA6G project.

Table 1. Prospective data that will be generated in the TERA6G project. The datasets are listed for each Working Package, taken from the project proposal.

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
WP1. Project management								
			Choose an item.			Choose an item.		Choose an item.
WP2. System specifications and concept design. Evaluation of demonstration results								
3D model of TERA6 transceiver demonstrator module	The data set contains the 3D model of the complete packages including all components /chips and interconnections of TERA6G transceiver demonstrator modules	D3.2, D3.3	Internal	STEP, DXF	<200MB	Restricted / Data accessible by the consortium members only	The data is generated for the partners involved in the assembly to design the thermal, mechanical, and electrical parts of the module as well as to define the assembly steps and address the integration limitations. It will not be disclosed as involves restricted technical know-how	5. PHIX
Final integration and assembly results	The data set contains the detailed pictures of the modules, chip-to-chip optical insertion loss	D3.2, D3.3	Internal	PDF, PNG, DOCX		Restricted / Data accessible by the consortium members only	The data is generated for the partners and reporting. It will not be disclosed as involves restricted technical know-how	5. PHIX
Data generated from analytical studies and simulations of wireless environment and algorithms	Codebooks and beam forming algorithms	D2.4	External			Open		6. UPRC

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
Channel and propagation models for sub-THz frequencies	Physical and statistical models explaining the propagation of sub-THz in various scenarios	D2.4	Internal	PDF	N/A	Open	N/A	7. OULU
WP3. Fabrication and Assembly of wireless Terahertz transceiver modules								
Simulation of the 3D waveguide routing network	This data set includes the files generated by photonic simulation software.	D3.2, D3.3	Internal	TXT, XLSX, PPT	< 100 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the system specifications and design in WP2. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Simulation of InP modulators	This data set includes the files generated by photonic simulation software.	D3.2, D3.3	Internal	TXT, XLSX, PPT	< 100 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the system specifications and design in WP2. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Simulation of InP THz emitter arrays (pin photodiodes)	This data set includes the files generated by photonic simulation software.	D3.2, D3.3	Internal	TXT, XLSX, PPT	< 100 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the system specifications and design in WP2. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Simulation of InP THz receiver arrays (photoconductive antennas)	This data set includes the files generated by photonic simulation software.	D3.2, D3.3	Internal	TXT, XLSX, PPT	< 100 MB	Restricted / Data accessible by the consortium	This data is generated for the partners involved in the system specifications and design in WP2. It will not be disclosed as involves	3. FhG-HHI

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
						members only	restricted technical know-how	
Layout of the 3D waveguide routing network	This data set includes the CAD files that will be used for the fabrication of the mask-sets.	D3.2, D3.3	Internal	GDS, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Layout of the InP modulators	This data set includes the CAD files that will be used for the fabrication of the mask-sets.	D3.2, D3.3	Internal	GDS, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Layout of InP THz emitter arrays (pin photodiodes)	This data set includes the CAD files that will be used for the fabrication of the mask-sets.	D3.2, D3.3	Internal	GDS, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Layout of InP THz receiver arrays (photoconductive antennas)	This data set includes the CAD files that will be used for the fabrication of the mask-sets.	D3.2, D3.3	Internal	GDS, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
Characterization data of the 3D waveguide routing network	This data set includes the pictures of the fabricated devices and measurement data acquired with the optoelectronic characterization setups.	D3.2, D3.3	Internal	JPEG, TIFF, PNG, TXT, XLSX, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules as well as in their testing in WP5. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Characterization data of the InP modulators	This data set includes the pictures of the fabricated devices and measurement data acquired with the optoelectronic characterization setups.	D3.2, D3.3	Internal	JPEG, TIFF, PNG, TXT, XLSX, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules as well as in their testing in WP5. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Characterization data of InP THz emitter arrays (pin photodiodes)	This data set includes the pictures of the fabricated devices and measurement data acquired with the optoelectronic characterization setups.	D3.2, D3.3	Internal	JPEG, TIFF, PNG, TXT, XLSX, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules as well as in their testing in WP5. It will not be disclosed as involves restricted technical know-how	3. FhG-HHI
Characterization data of InP THz receiver arrays (photoconductive antennas)	This data set includes the pictures of the fabricated devices and measurement data acquired with the optoelectronic characterization setups.	D3.2, D3.3	Internal	JPEG, TIFF, PNG, TXT, XLSX, PPT	< 200 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules as well as in their testing in WP5. It will not be disclosed	3. FhG-HHI

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
							as involves restricted technical know-how	
Layout of the lasers and the optical beamforming networks	This data includes waveguides designs in TriPleX platform	D3.2, D3.3	Internal	GDS	<100 MB	Restricted / Data accessible by the consortium members only	This data is generated for the partners involved in the assembly and packaging of the transceiver modules. It will not be disclosed as involves restricted technical know-how	4. LXI
WP4. Development of network protocols and algorithms, including extension of network slicing tools								
Data generated from analytical studies and simulations of wireless environment and algorithms	Codebooks and beam forming algorithms	D4.2	External			Open		6. UPRC
mMIMO algorithms for beamforming, tracking and energy efficient communications	Physical and mathematical models for mMIMO signal processing methods for various beamforming scenarios	D4.2	Internal	PDF	N/A	Open	N/A	7. OULU
Slice metadata	Data structure including slice ID, QoS profile associated to each slice and Network Functions associated to the slice		Internal	JSON	1KB per slice. System can have 256 slices maximum	Restricted	Accessible to network administrator	8. CMC
WP5. TERA6G network demonstration. Integration and Validation and of wireless Terahertz modules in demonstrator test bed								
Characterization of fully integrated OBFNs based	The data set will include parts of the experimental results of	D5.3	Internal	PPTX, DOCX	<100 MB	Restricted / Data accessible by	This data is generated for the partners involved in the assembly, to design the	2. ICCS

Data Set	Data Description	Related Deliverable	Type	Format	Estimated Volume	Access	Restriction Level Rationale	IPR Owner
on Blass matrix architecture	the TERA6G modules (2nd gen) at ICCS's lab.					the consortium members only	probe head. It will not be disclosed as involves restricted technical know-how.	
Test data for the validation	Test data produced with UOULU channel sounders to benchmark and test TERA6G prototypes	D5.2, D5.3	Internal	CSV	<10 GB	Restricted / Data accessible by the consortium members only	Data made open access to partners of TERA6G via the repository	7. OULU
WP6. Dissemination, exploitation, standardization and roadmapping								

7 FAIR (Findable, Accessible, Interoperable, and Reusable) data

7.1 Making data findable, including provisions for metadata

For the open datasets, a Digital Object Identifier (DOI) will be created for effective and persistent citation when the dataset is uploaded to the data repositories Zenodo or IEEE DataPort. This DOI can be used in any relevant publications to direct readers to the underlying dataset.

Each dataset generated during the project will be recorded in an Excel spreadsheet with a standard format and allocated a dataset identifier according to Table 2. The owner of each dataset will be responsible for the creation of the spreadsheet associated with the generated dataset, while UC3M (the Project Coordinator) will be responsible for gathering and preserving the information of all the datasets in its local repository.

TERA6G naming convention for project datasets will comprise of the following:

- A unique chronological number of the datasets in the project will be added.
- The title of the dataset.
- A prefix “TERA6G” indicating an TERA6G dataset.
- A unique identification number linking with the dataset work package, e.g., "WP1".
- For each new version of a dataset, a version number will be allocated. The starting version numbers will be, for example, v1.0 or rev_01.

Search keywords will be provided when the dataset is uploaded to Zenodo, which will optimize possibilities for data re-use. Zenodo follows the minimum Data Cite metadata standards. The information and metadata stored in the record-spreadsheet for each generated dataset is summarized in Table 2.

Table 2. Fields of the record-spreadsheet for each dataset of TERA6G.

1. Data set Reference Name	*The reference name will be based on the naming convention outlined above in Section 7.1.*
2. Description	*Description of the Dataset*
3. Dataset Title	*The title of the dataset which should be easily searchable & findable*
4. Description of Data	*Brief description of the open data to be included in the metadata*
5. Name(s) of dataset creator(s)	*Lead partners responsible for the creation of the dataset*
6. Data source	*How/why was the dataset generated*
7. Creation date	*Date of generation of the dataset*
8. Format	*Possible formats of the datasets, e.g., DOCX, XLSX, PDF, JPEG, TIFF, etc.*
9. Expected Size	*Approximate size of the dataset*
10. Data Location	*Institutional repository where the data are stored*
11. Digital Object Identifier (DOI)	*The DOI can be entered once the dataset has been deposited in the repository*
12. Access Status	*Type of Dataset "Open" or "Restricted"*
13. Embargo period	*Embargo period of the dataset (if applicable)*
14. Funding statement	*A disclaimer indicating that the specific dataset was generated within the* TERA6G *project that received funding from the* Horizon Europe (HORIZON) Framework Programme *under G.A. No 101096949.*
15. Work Package	TERA6G W*ork Package associated with this dataset*
16. Related Publications	*Bibliographical details of publications based on the dataset will be listed, with links to abstracts and, where possible, full text.*
17. Dataset citation	*A 'ready-to-use' citation reference for the dataset, incorporating the core descriptive elements*
18. Keywords	*Dataset related Keywords*
19. Version Number	*Dataset version number to keep track of changes to the dataset*
20. Repository	*Expected repository to be submitted*
21. Date of Repository Submission	*The date of submission to the repository will be added once it has been submitted*

7.2 Making data openly accessible

Research data that are created during the project are owned by the beneficiary that generates them (G.A. Article 16). Each beneficiary must disseminate its results as soon as possible unless there is legitimate interest to protect the results. A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries—unless agreed otherwise—at least 45 days in advance, together with enough information on the results it will disseminate (G.A. Article 16).

Generated research data will be deposited as follows:

- ***Restricted data*** will be deposited in the repository of the partner that owns it.
- ***Data accessible by the consortium members only*** will be deposited in the ***DropBox Document Repository*** of TERA6G, in a private area reserved for consortium members only. More specifically, a directory or directories will be created in the project's private area, for easy upload of project datasets visible to the consortium members.
- ***Open data*** will be deposited in Zenodo repository, IEEE DataPort or similar open access repository. Zenodo.org is open, free, searchable, and structured with flexible licensing allowing for storing all types of data: datasets, images, presentations, publications, and software. In addition, Zenodo repository allows researchers to deposit both publications and data, while providing tools to link them.

The TERA6G project has chosen to use Zenodo.org as the reference repository for storing the open project data for the following reasons:

- It enables *Shared Research* through the reposition of all research outputs from across all fields of research and science.
- It is *Citeable and Discoverable:* uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
- It *fosters the establishment and curation of research Communities* through the creation of digital repositories, fully controlled by the owner.
- Allows the *identification of grants*, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
- Allows for *Flexible licensing* and,
- *Ensures security of the stored research results* for the future in the same cloud infrastructure as CERN's own LHC research data.

Data objects will be deposited in Zenodo under:

- Open access to data files and metadata and data files provided over standard protocols such as HTTP and OAI-PMH
- Use and reuse of data permitted
- Privacy of its users protected

For the data being deposited in an external repository, e.g., Zenodo, a dataset registry record will also be created in the TERA6G Document Repository in DropxBox. The registry record will be updated by the partner that is responsible for the specific generated data and will include relevant metadata explaining what data exists and a DOI linking to where the data is available in the external repository.

Research data needed to validate the results in the scientific publications should be deposited in the data repository as soon as possible (GA Article 17 and GA Annex 5).

During embargo periods, information about the restricted data will be published in the data repository, and details of when the data will become available will be included in the metadata. Where a restriction on open access to data is necessary, attempts will be made to make data available under controlled conditions to other individual researchers. In accordance with GA Article 25, data must be made available to partners upon request, including in the context of checks, reviews, audits, or investigations. Data will be made accessible and available for re-use and secondary analysis.

In parallel to the available Open Research Datasets, several restricted datasets will be generated within the project. These datasets that contain critical details on TERA6G findings will be protected in a restricted area. Each TERA6G partner is responsible to identify these datasets and to communicate to the consortium members any updates on these datasets. Of course, these datasets can be moved from the Restricted Area to the Open Data zone of TERA6G if the data owners select to disclose them partially or fully. This change will be resembled in the Data Management Plan of the project in a rolling-update of this document.

The datasets that will be restricted are summarized in the Table 3 below.

Table 3. Restricted datasets in the TERA6G project.

Dataset Description	Related WP	Related deliverables	Dataset Owner
Characterization of fully integrated OBFNs based on Blass-matrix architecture	**WP5**	D5.3	2. ICCS
Simulation of the 3D waveguide routing network	**WP3**	D3.2, D3.3	3. FhG-HHI
Simulation of InP modulators	**WP3**	D3.2, D3.3	3. FhG-HHI
Simulation of InP THz emitter arrays (pin photodiodes)	**WP3**	D3.2, D3.3	3. FhG-HHI
Simulation of InP THz receiver arrays (photoconductive antennas)	**WP3**	D3.2, D3.3	3. FhG-HHI
Layout of the 3D waveguide routing network	**WP3**	D3.2, D3.3	3. FhG-HHI
Layout of the InP modulators	**WP3**	D3.2, D3.3	3. FhG-HHI
Layout of InP THz emitter arrays (pin photodiodes)	**WP3**	D3.2, D3.3	3. FhG-HHI
Layout of InP THz receiver arrays (photoconductive antennas)	**WP3**	D3.2, D3.3	3. FhG-HHI
Characterization data of the 3D waveguide routing network	**WP3**	D3.2, D3.3	3. FhG-HHI
Characterization data of the InP modulators	**WP3**	D3.2, D3.3	3. FhG-HHI
Characterization data of InP THz emitter arrays (pin photodiodes)	**WP3**	D3.2, D3.3	3. FhG-HHI
Characterization data of InP THz receiver arrays (photoconductive antennas)	**WP3**	D3.2, D3.3	3. FhG-HHI
Layout of the lasers and the optical beamforming networks	**WP3**	D3.2, D3.3	4. LXI
Test data for the validation	**WP5**	D5.2, D5.3	7. OULU
Slice metadata including slice parameters such as ID, QoS information, Network functions ID	**WP4**		8. CMC

7.3 Making data interoperable

The TERA6G project aims to collect and document the data in a standardized way to ensure that, the datasets can be understood, interpreted, and shared in isolation alongside accompanying metadata and documentation. The generated data will be preserved either in the TERA6G DropBox Document Repository or in the institutional intranet platforms until the end of the Project (see Section 9 of this document).

a) General Information

- Title of the dataset
- Dataset Identifier
- Responsible Partner
- Author Information
- Date of data collection
- Geographic location of data collection
- Title of project and Funding sources that supported the collection of the data

b) Sharing/Access Information

- Licenses/access restrictions placed on the data
- Link to Data Repository
- Links to other publicly accessible locations of the data
- Links to publications that cite or use the data
- Is the data derived from another source?

c) Dataset/File Overview

- This dataset contains X sub-dataset as listed below: ……
- What is the status of the documented data? – “complete”, “in progress”, or “planned”
- Are there plans to update the data?

d) Methodological Information

- Description of methods used for experimental design and data collection: <*Include links or references to publications or other documentation containing experimental design or protocols used in data collection*>
- Methods for processing the data: <*Describe how the submitted data were generated from the raw or collected data*>
- Instruments and software used in data collection and processing-specific information needed to interpret the data
- Standards and calibration information, if appropriate

- Environmental/experimental conditions
- Describe any quality-assurance procedures performed on the data
- Dataset benefits

7.4 Increase data re-use (through clarifying licenses)

The open datasets will be made available for re-use through uploads to the Zenodo community page for the project. In principle, the data will be stored in Zenodo after the conclusion of the project without additional cost. All the research data will be of the highest quality, have long-term validity and will be well documented in order other researchers to be able to get access and understand them for at least after 5 years after the project concludes.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data. Quality control of the data is the responsibility of the relevant responsible partner that generates the dataset.

7.5 Foreseen open research datasets

Table 1 shows the list of the data sets identified as relevant for TERA6G results validation and evaluation which are foreseen to be Open Access, for which the corresponding description is provided in the tables within this section.

Table 4. List of Open data sets relevant to TERA6G research activities.

Dataset number	Data set name	WP related	Type	Partner
Dataset 01	**Slice metadata**	WP4	Internal	CMC

7.5.1 Dataset 01: Slice metadata

[TODO] complete the table that describes this dataset according to section 7.1

8 Allocation of resources

There are no immediate costs anticipated to make the datasets produced FAIR. The open datasets will be deposited in the Zenodo repository and will be preserved for at least 5 years after the conclusion of the project and in-line with the Deposit Data Policy of the European Commission. Any unforeseen costs related to open access to research data in Horizon Europe (HORIZON) Framework Programme are eligible for reimbursement during the duration of the project under the conditions defined in the GA, Article 6.2.C.3.

UC3M (the Project Coordinator) will be the responsible for the implementation of Data Management Plan within TERA6G project. All the consortium partners will also undertake to support the Data Management protocol by providing the types of generated data that have been described in the methodology of the DMP report. Specifically, UC3M will:

a) prepare the Deliverable D6.3a (Data Management Plan),

b) ensure its update during the project course of action,

c) ensure the recording of the produced datasets,

d) have the overall responsibility for the implementation of the Data Management Plan.

Each TERA6G partner must respect the policies set out in this DMP. Datasets must be created, managed, and stored appropriately and in line with European Commission and local legislations. Dataset validation and registration of metadata and backing-up data for sharing through repositories is the responsibility of the partner that generates the data in the respective Work Package.

9 Data security

The data storage facilities and its security policies of the consortium members are summarized in the Table 5 below.

Table 5. Description of data storage facilities of the TERA6G project consortium members.

Partner name	Description of Data Storage Facility
1. UC3M	UC3M storage services for the researchers are based on an institutional Google Drive account for each researcher, with unlimited capacity, enabling to grant access between members of UC3M. Furthermore, UC3M library gives support to UC3M group on issues related to Open Access/Open Science. This service cooperates intensively with the research service of UC3M, that has an institutional repository that belongs to the Spanish repository RECOLECTA and is compatible with the European portal OpenAIRE. All these tools will be used, and the library's consultant office will support the PC.
2. ICCS	University-owned server at ICCS (PCRL). The server offers real-time data mirroring through RAID (redundant array of independent disks) and weekly backups to external disk drives. The Data will be preserved for at least 5 years after the project end and the associated costs will be covered by ICCS through own funds.
3. FhG-HHI	Internal servers at FhG-HHI premises with daily backups. The data will be preserved for at least 5 years after the project ends.
4. LXI	LXI owned server with daily backups, where common set of folders related to specific tasks/projects are accessible by LXI employees only.
5. PHIX	PHIX owned server with daily backups where common set of folders related to specific tasks/projects are accessible only by the responsible staff.
6. UPRC	UPRC storage services for the researchers are based on personal Dropbox accounts for each researcher. Data generated by algorithms are also stored locally in institutional PCs, where the algorithms are tested.
7. OULU	UOULU uses Microsoft services for internal data storage. These are controlled by the UOULU IT services that ensure security and access to the UOULU employees. These include, e.g., Teams and SharePoint, among other services. These data services rely on cloud storage that is backed up automatically to ensure data security and allow recovery of data. UOULU also hosts an open access repository Jultika where all the publications produced by researchers of UOULU (journals, conferences, white papers, etc.) are uploaded and are accessible to everyone in the world.
8. CMC	CMC uses Mongo DB for storing all 5GC related information. The information is stored and accessible through module called Cumucore Network Configuration (CNC) which provides a generic interface for storing and retrieving information from internal DB.

Partner name	Description of Data Storage Facility
9. ICOM	
10. TID	Microsoft SharePoint is used as the company-level general tool for project data storage and document management. Both on premises and cloud-based online versions are available and supported by specific internal security and IT teams within the organization. Project or company sensitive data is encrypted.

10 Ethical aspects

TERA6G partners must comply with the ethical principles as set out in the GA Article 14 and GA Annex 5, which states that all activities must be carried out in compliance with:

a) Ethical principles, including the highest standards of research integrity—as set out, for instance, in the European Code of Conduct for Research Integrity (European Science Foundation, 2011)—including but not limited to, avoiding the fabrication of data, the falsification of data, plagiarism or any another research misconduct and,

b) applicable international, EU and national law.

Individuals invited to participate in the field trials or other lab tests are the ones employed by the project participants to carry out TERA6G project activities and, more specifically, the ones whose role in the project is pertinent to field trials and lab tests. No volunteers are to be recruited for the needs of project activities including field trials or any other lab tests. All TERA6G project participants and project personnel that will participate in the field trials will be managed according with the EU and national ethics laws and principles. In the unforeseen case of personal data collection during the field trials, every effort will be taken to not process and retain them, or to anonymizing them[5] if it is assessed so.

11 Other issues

No other national, funder, sectorial, or departmental procedures for data management are being used by the TERRA6G partners.

12 Conclusion

This document has provided the first version of the Data Management Plan defined by the TERA6G consortium in the first six months of the project. The DMP has identified several data sets which will be generated within the project and, for each of them, has defined the strategy for the documentation, open access sharing and maintenance of the associated data, according to the guidelines provided by the EC in the Horizon Europe (HORIZON) Framework Programme.

This document will be progressively updated during the project lifespan, to reflect any possible changes and additions in the datasets as well as any refinement in the strategy to maximize the sharing and re-use of the project outcome. These possible changes should be also reflected in the periodic reports of the TERA6G project. Particularly, at least one further updated version of the DMP will be released in M24 (December 2024).

[5] Data anonymization according to WP 29 Opinion 05/2014 on Anonymisation Techniques